IFIP AICT 590

David Kreps
Taro Komukai
T. V. Gopal
Kaori Ishii
(Eds.)

Human-Centric Computing in a Data-Driven Society

14th IFIP TC 9 International Conference
on Human Choice and Computers, HCC14 2020
Tokyo, Japan, September 9–11, 2020, Proceedings

Springer

IFIP Advances in Information and Communication Technology 590

IFIP – The International Federation for Information Processing

IFIP was founded in 1960 under the auspices of UNESCO, following the first World Computer Congress held in Paris the previous year. A federation for societies working in information processing, IFIP's aim is two-fold: to support information processing in the countries of its members and to encourage technology transfer to developing nations. As its mission statement clearly states:

> *IFIP is the global non-profit federation of societies of ICT professionals that aims at achieving a worldwide professional and socially responsible development and application of information and communication technologies.*

IFIP is a non-profit-making organization, run almost solely by 2500 volunteers. It operates through a number of technical committees and working groups, which organize events and publications. IFIP's events range from large international open conferences to working conferences and local seminars.

The flagship event is the IFIP World Computer Congress, at which both invited and contributed papers are presented. Contributed papers are rigorously refereed and the rejection rate is high.

As with the Congress, participation in the open conferences is open to all and papers may be invited or submitted. Again, submitted papers are stringently refereed.

The working conferences are structured differently. They are usually run by a working group and attendance is generally smaller and occasionally by invitation only. Their purpose is to create an atmosphere conducive to innovation and development. Refereeing is also rigorous and papers are subjected to extensive group discussion.

Publications arising from IFIP events vary. The papers presented at the IFIP World Computer Congress and at open conferences are published as conference proceedings, while the results of the working conferences are often published as collections of selected and edited papers.

IFIP distinguishes three types of institutional membership: Country Representative Members, Members at Large, and Associate Members. The type of organization that can apply for membership is a wide variety and includes national or international societies of individual computer scientists/ICT professionals, associations or federations of such societies, government institutions/government related organizations, national or international research institutes or consortia, universities, academies of sciences, companies, national or international associations or federations of companies.

More information about this series at http://www.springer.com/series/6102

David Kreps · Taro Komukai ·
T. V. Gopal · Kaori Ishii (Eds.)

Human-Centric Computing in a Data-Driven Society

14th IFIP TC 9 International Conference on Human Choice and Computers, HCC14 2020 Tokyo, Japan, September 9–11, 2020 Proceedings

Springer

Editors
David Kreps
University of Salford
Salford, UK

T. V. Gopal
Anna University
Chennai, India

Taro Komukai
Chuo University
Tokyo, Japan

Kaori Ishii
Chuo University
Tokyo, Japan

ISSN 1868-4238 ISSN 1868-422X (electronic)
IFIP Advances in Information and Communication Technology
ISBN 978-3-030-62802-4 ISBN 978-3-030-62803-1 (eBook)
https://doi.org/10.1007/978-3-030-62803-1

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book contains the proceedings of the 14th International Conference on Human Choice and Computers (HCC 2020), which was planned to be held at the Faculty of Global Informatics (iTL), Chuo University, Japan, during September 9–11, 2020. The conference was organized by the International Federation for Information Processing (IFIP) Technical Committee 9 (TC9): Information and Communication Technology (ICT) and Society.

As with so many events of 2020, however, the conference was cancelled on March 28, 2020, owing to the global COVID-19 pandemic, with the resolution that, nonetheless, the proceedings of the conference – already written and in review by that time – would still be published.

The conference co-chairs, Taro Komukai (Japan representative), David Kreps, (chair of WG 9.3 and chair of TC9), Gopal TV (guest program chair), and Kaori Ishii (guest program chair), chose the theme for this year's conference: "Human-Centric Computing in a Data-Driven Society." Tracks were advertised in the call for papers addressing a range of concerns across the working groups of TC9, and the accepted papers coalesced into six groups: Ethical and Legal Considerations; Data; Peace and War; Our Digital Lives; Social Accountability; and Gender.

The papers selected for this book are based on both academic research and the professional experience of information systems practitioners working in the field. It is the continued intention of TC9 that academics, practitioners, governments, and international organizations alike will benefit from the contributions of these proceedings.

Details of the activities of IFIP TC9's activities are posted at http://www.ifiptc9.org/.

August 2020 David Kreps

Organization

Program Chairs

Taro Komukai	Chuo University, Japan
David Kreps	University of Salford, UK
Gopal T. V.	Anna University, India
Kaori Ishii	Chuo University, Japan

Track Chairs

Taro Komukai	Chuo University, Japan
Kaori Ishii	Chuo University, Japan
Gopal T. V.	Anna University, India
Jani Koskinen	University of Turku, Finland
Petros Chamakiotis	ESCP Business School, Spain
Brad McKenna	University of East Anglia, UK
Sisse Finken	IT University of Copenhagen, Denmark
Johanna Sefyrin	Linköping University, Sweden
Charles Ess	University of Oslo, Norway
Brett van Niekerk	University of KwaZulu-Natal, South Africa

Program Committee

Arzoo Atiq	Massey University, New Zealand
Kathrin Bednar	University of Vienna, Austria
Arnab Bhattacharya	IIT Kanpur, India
Oliver Burmeister	Charles Sturt University, Australia
Wenjie Cai	University of Greenwich, UK
Julie Cameron	InfoTech Solutions, Australia
Petros Chamakiotis	ESCP Business School, Spain
Liezel Cilliers	University of Fort Hare, South Africa
Anna Croon	Umeå University, Sweden
Richard Dron	University of Salford, UK
Penny Duquenoy	Middlesex University, UK
Pirjo Elovaara	Blekinge Institute of Technology, Sweden
Sisse Finken	IT University of Copenhagen, Denmark
Gordon Fletcher	University of Salford, UK
Per Fors	Uppsala University, Sweden
Gopal T. V.	Anna University, India
Don Gotterbarn	East Tennessee State University, USA
Marie Griffiths	University of Salford, UK
Marthie Grobler	CSIRO's Data61, Australia

Satoshi Hashimoto	Tokushima Bunri University, Japan
Olli I. Heimo	University of Turku, BID Technology, Finland
Magda Hercheui	University College London, UK
Mitsuyoshi Hiratsuka	Tokyo University of Science, Japan
Jun Iio	Chuo University, Japan
Kaori Ishii	Chuo University, Japan
Michihiro Iwakuma	Chuo University, Japan
Iordanis Kavathatzopoulos	Uppsala University, Sweden
Kai Kimppa	University of Turku, Finland
Jay Kishigami	Muroran Institute of Technology, Japan
Taro Komukai	Chuo University, Japan
Jani Koskinen	University of Turku, Finland
David Kreps	University of Salford, UK
Mikael Laaksoharju	Uppsala University, Sweden
Chris Leslie	South China University of Technology, China
Takayuki Matsuo	Momo-o, Matsuo & Namba Law Office, Japan
Brad McKenna	University of East Anglia, UK
Christina Mörtberg	Linnaeus University, Sweden
Yosuke Murakami	KDDI Research, Inc., Japan
Kiyoshi Murata	Meiji University, Japan
Mika Nakashima	Chuo University, Japan
Juhani Naskali	University of Turku, Finland
Yoshiaki Nishigai	Chiba University, Japan
Norberto Patrignani	Politecnico di Torino, Italy
Jackie Phahlamohlaka	CSIR, South Africa
Trishana Ramluckan	University of KwaZulu-Natal, South Africa
Minna Rantanen	University of Turku, Finland
Machiko Sakai	University of Tokyo, Japan
Johanna Sefyrin	Linköping University, Sweden
Srinath Srinivasa	IIIT Bangalore, India
Riana Steyn	University of Pretoria, South Africa
Sam Takavarasha Jr.	Women's University in Africa, Zimbabwe
Kuninobu Takeda	Osaka University, Japan
Richard Taylor	International Baccalaureate, UK
Anne-Marie Tuikka	University of Turku, Finland
Jean-Paul Van Belle	University of Cape Town, South Africa
Maja Van Der Velden	IFI, Uiniversity of Oslo, Norway
Brett van Niekerk	University of KwaZulu-Natal, South Africa
Christine Van Toorn	University of New South Wales, Australia
Will Venters	London School of Economics, UK
Ruth Wario	University of the Free State, South Africa
Martin Warnke	Leuphana University of Lüneburg, Germany
Chris Zielinski	University of Winchester, UK

Contents

Ethical and Legal Considerations in a Data-Driven Society

A Study on Abusing Superior Bargaining Position in the Anti-Monopoly Act and Its Relation to the Act on the Protection of Personal Information in Japan 5
Kaori Ishii

The Laws of Big Data: How Data Protection Law, Competition Law and Contract Law Deal with the Challenges of a Data-Driven World. 16
Lena Mischau

The Legal Frameworks of the Right to Request the Deletion of Personal Data in the EU, the U.S. and Japan and the Right to Be Forgotten: A Study Focusing on Search Businesses 29
Mika Nakashima

Examining the Possibility of Expanding the Use of Digital Images Recorded in the "Audio-Visual Recording of Custodial Interrogation" System 41
Mariko Nakamura

Governance Framework for Facial Recognition Systems in Japan 52
Aimi Ozaki

Advanced Software and Patents: A Patentability Balance for Fostering Technology 64
Michele Baccelli, Kazuto Kobayashi, Steven C. Sereboff, and Mitsuyoshi Hiratsuka

IPR Risk Assessment of Companies Implementing Standard Essential Patents and Validation 77
Kazuto Kobayashi, Mitsuyoshi Hiratsuka, and Michele Baccelli

Qualitative Analysis of Interviews with Municipal Officers Toward the Human-Centered Improvement of the eLen Regulation Database System 86
Aki Shima and Tokuyasu Kakuta

Japan-EU Passenger Name Record Negotiations and Their Implications. 100
Toru Maruhashi

The Data-Driven Society

Heaven and Earth – Cloud and Territory in the Internet 117
Martin Warnke

Self-Sovereign Identity and Blockchain-Based Content Management 130
Yusuke Kurihara

Assessing the Japanese Turn in AI and Robot Ethics: Extracting Meaningful Principles Between Exoticism and Empiricism in the Case of AIBO 141
Vassilis Galanos and Mary Reisel

Intelligent Monitoring of Chronic Illness for the Ageing Rural Population: Opportunities and Cautions 158
Jenni Greig, Anwaar Ul-Haq, Greg Dresser, Oliver K. Burmeister, and Sabih-Ur Rehman

The Relevance of Humans and Structure: Managerial and Organizational Challenges in Smart Factories. 171
Dennis Grenda and Anne-Marie Tuikka

Automated Testing of Refreshable Braille Display. 181
Shivam Kumar Singh, Sujit Kumar Chakrabarti, and Dinesh Babu Jayagopi

Data Processing Automation for Bulk Water Supply Monitoring 193
Arno de Coning and Francois Mouton

Peace and War

Using Cyber Applications Towards Positive Psychology Interventions in Africa 207
Carien Van ‘t Wout, Joey Jansen van Vuuren, and Anna-Marie Jansen van Vuuren

States’ Capacity Building for Cybersecurity: An IR Approach 222
Seiko Watanabe

An Analysis of Twitter During the 2017 Zimbabwean Military Intervention 233
Brett van Niekerk, Martina Jennifer Zucule De Barros, and Trishana Ramluckan

Cybersecurity and Cybercrime Combatting Culture for African Police Services 248
Louise Leenen, Joey Jansen van Vuuren, and Anna-Marie Jansen van Vuuren

A Consideration of the Case Study of Disinformation and Its Legal Problems 262
Tomoko Nagasako

Our Digital Lives

Exploring Human Nature in a Technology-Driven Society 281
Kathrin Bednar

Towards a Typology of Intentionally Inaccurate Representations of Reality in Media Content 291
Matthew J. Davis and Per Fors

The Influence of Personal and Professional Commitments on Digitally Disconnected Experiences 305
Brad McKenna, Lena Waizenegger, and Wenjie Cai

Social Materiality of Smartphone Game Apps: Case Analysis of *Pokémon GO* 315
Hiroshi Koga

Individuals in Data-Driven Society

Humans of the European Data Economy Ecosystem - What Do They Demand from a Fair Data Economy?. 327
Minna M. Rantanen and Jani Koskinen

Public Fora and Freedom of Expression. 340
Olli I. Heimo and Kai K. Kimppa

Self-tracking, Power, and the Transition from Discipline to Control. 351
Jukka Vuorinen and Harley Bergroth

Towards Social Enterprise with Internet of Office Desks 361
Elena Vildjiounaite, Johanna Kallio, Julia Kantorovitch, Vesa Kyllönen, Pauli Räsänen, and Jussi Ronkainen

Gender, Diversity and ICT

What Can Statistics Tell About the Gender Gap in ICT? Tracing Men and Women’s Participation in the ICT Sector Through Numbers. 379
Morten Simonsen and Hilde G. Corneliussen

Silent Data, Active Patients 398
Cæcilie Sloth Laursen and Sisse Finken

Author Index 409

Ethical and Legal Considerations in a Data-Driven Society

Ethical and Legal Considerations in a Data-Driven Society

Taro Komukai[1] and Kaori Ishii[2]

[1] Faculty of Global Informatics, Chuo University, Tokyo, Japan
komukai@tamacc.chuo-u.ac.jp
[2] Faculty of Global Informatics, Chuo University, Tokyo, Japan
kaoriish@tamacc.chuo-u.ac.jp

Legal issues are absolutely essential when we discuss "human-centric computing in a data-driven society" because the development of a data-driven society creates new concerns, including threats to privacy and data protection, unforeseen trouble in using new technologies, and difficulties in protecting intellectual property rights (IPR).

We present nine papers that focus on the legal issues associated with the use of ICTs. Two papers explore practical issues in the field of IPR and ICTs. Kobayashi, Hiratsuka, and Baccelli (2020) focus on the risk assessment associated with the implementation of standard-essential patents (SEPs). When thinking about implementing SEPs, companies often face the risk of high royalties because specific royalties are usually not defined in a patent policy. This paper provides a formula for calculating the IPR of the implementer of an SEP by modeling patentees for SEPs. Baccelli, Kobayashi, Sereboff, and Hiratsuka (2020) discuss the issue of software and patents. Although patents have also been conceived as tools for rewarding and motivating the inventor's endeavors, many people believe that inventions of software that are purely non-technical should be excluded from patent protection. However, it is often difficult to determine whether an invention is purely non-technical. This paper includes a discussion regarding the fair balance between the need for patent protection for intangible inventions based on technology and the avoidance of non-technical arrangements; accordingly, the legal situation at the European Patent Office (EPO) is analyzed.

Nakamura (2020) examines the possibility of using digital images recorded by new technology in a criminal court. Although Japan has recently introduced the "Audio-Visual Recording of Custodial Interrogation" system, the digital images recorded in the system are supposed to be used as supplementary evidence. In this paper, the possibility of using these images to prove guilt or innocence is discussed.

Shima and Kakuta (2020) present the results of their research on the actions of municipal officers regarding the human-centered improvement of the eLen regulation database system. They clarify the problems remaining with the e-legislation support system.

Ozaki's paper (2020) explores the appropriate governance framework for facial recognition systems used by private companies. The paper provides a practical framework for business planning divisions when providing personal identification services using facial recognition technology.

The papers by Nakashima (2020) and Maruhashi (2020) concern data protection and related issues. Nakashima (2020) analyzes the comparative legal frameworks of the

EU, the USA, and Japan on “the right to be forgotten” by focusing on search engine providers. She elaborates on the possibility of a legal right whereby a natural person can request the deletion of search results on him/her from search engine providers. She also points out the relevant issues to be discussed. Responding to the recent legal approach taken in the EU and California, as well as the Supreme Court’s decision in Japan, she recognizes the need to study the ideal state of responsibility of new media known as search businesses. The work by Maruhashi (2020) delves into the international data flow system by looking at the EU-Japan PNR (passenger name record) Negotiation. Maruhashi (2020) provides an overview of PNR and the nature of its processing and the current legal and practical framework in Japan, and then compares the Japanese PNR system with the original draft negotiation directive of the Japan-EU PNR agreement. He insists on the need to standardize the way of controlling the PNR’s algorithmic pattern-based search to combat terrorism and internationally organized crime.

Recent legal issues regarding data protection are intercorrelated with other legal areas, particularly competition law, consumer protection law, and contract law. Mischau (2020) and Ishii (2020) draw attention to this aspect. Mischau (2020) focuses on big data’s role in competition law, data protection law, and contract law. Her approach mostly reflects the European perspective. Her paper concludes with the following: “law will have to balance very different interests with regard to Big Data and data analytics and balance the chances and risks Big Data comes with in particular.” Professor Ishii’s discussion of the intersection between the Act on the Protection of Personal Information and the Anti-Monopoly Act in Japan focuses on abuses of superior bargaining positions by digital platformers. She argues that while privacy and personal data must be the first priority in laws designed to protect personal information, competition law and other adjacent laws are increasingly significant, and studying them can offer a different perspective on the protection of personal information.

A Study on Abusing Superior Bargaining Position in the Anti-Monopoly Act and Its Relation to the Act on the Protection of Personal Information in Japan

Kaori Ishii(✉)

Faculty of Global Informatics, Chuo University, Tokyo, Japan
kaoriish@tamacc.chuo-u.ac.jp

Abstract. This study discusses the intersection between the Act on the Protection of Personal Information (APPI) and the Anti-Monopoly Act (AMA) in Japan, by focusing on abusing superior bargaining position from platform operators. My analysis is based on examinations of the provisions and related guidelines of AMA, the relevant provisions of APPI, and comparisons between the two regulations. Based on these findings: (1) most of the types of abuse which the Guidelines on Abusing Superior Bargaining Position (ASBP Guidelines) presented by Japan Fair Trade Commission (JFTC) overlap with APPI provisions; (2) restrictions on abusing superior bargaining positions could play a specific role by applying itself to profiling activities which APPI might not effectively regulate. However, the possibility for indefinitely expanding the scope of "superior bargaining position" and the scarce experiences of administrative fines would be challenges for AMA. In addition, clarifying the theoretical reason to incorporate privacy and personal data protection into AMA would be a fundamental issue. Other than abusing superior bargaining positions, cooperation or conflict between anti-monopoly law and protection of personal information law need to be carefully examined depending on the situation, such as refusal of deal, and merger. While privacy and personal data must be the first priority in laws designed to protect personal information, competition law and other adjacent laws are increasingly significant. Studying them can offer a different perspective on the protection of personal information.

Keywords: Abusing superior bargaining position · Privacy · Personal information · Competition

1 Introduction

Big data analytics, Internet of Things (IoT), and Artificial Intelligence (AI) have created tremendous amounts of global data flow which is rapidly changing the online world. For instance, CISCO's survey estimates that annual global IP traffic will reach 4.8 ZB per year by 2022, or 396 exabytes (EB) per month [1].

Borderless data flow has gradually broken-down jurisprudence barriers. The more complicated the online world becomes, the more likely it is that information-related intersecting legal issues will increase. This tendency is applicable among data

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 5–15, 2020.
https://doi.org/10.1007/978-3-030-62803-1_1

protection laws, competition laws, and consumer protection laws in relation to regulations on "platform operators." The definition of platform operators encompasses a wide range of service providers: online shopping malls, internet auctions, online flea markets, apps markets, search services, contents distribution services (image, video, music, e-book, etc.), booking services, sharing economy platforms, social networking services (SNS), video sharing services, electronic payment services, and so forth [2, p. 2]. Deceptive data practices by these services or their users would simultaneously provoke infringements of consumer contract law, personal information protection laws, and competition laws. Cabinet Office, JFTC, and Consumer Affairs Agency in Japan, and other similar agencies have held expert meetings to launch new policy strategies designed to address legal issues raised by platform operators. One policy strategy, for example, was the enactment of the Act on Improving Transparency and Fairness of Transactions of Specified Digital Platform Operators, which passed the National Diet on May 27, 2020 [3].

Among a series of legal challenges, this paper focuses on one of the intersection between APPI and AMA. On December 19, 2019, JFTC released the "Guidelines Concerning Abuse of a Superior Bargaining Position in Transactions between Digital Platform Operators and Consumers that Provide Personal Information, etc." [2] They are the first guidelines which shed light on how infringements of personal information can be regulated by AMA. In principle, AMA is an economic law which aims to ensure sound competitive surroundings, not protect privacy and personal information. While APPI is better suited to protecting personal information, these types of data have started to affect legitimate competition as many kinds of personal data have been traded in digital markets.

Based on the above information, this paper overviews provisions of abusing bargaining position in AMA, JFTC Guidelines, and APPI provisions, then discusses the division of roles between AMA and APPI.

2 JFTC ASBP Guidelines

2.1 Abusing Superior Bargaining Position

On December 17, 2019, JFTC published the Abusing Superior Bargaining Position (ASPB) Guidelines. They are the first guidelines in Japan covering practices that relate to the processing of personal information.

Article 19 of AMA restricts "unfair trade practices." This term includes abusing a superior bargaining position defined in Article 2(9)(v) [4].

- (v) engaging in any act specified in one of the following by making use of one's superior bargaining position over the counterparty unjustly, in light of normal business practices:
  - (a) causing the counterparty in continuous transactions (including a party with whom one newly intends to engage in continuous transactions; the same applies in (b) below) to purchase goods or services other than those to which the relevant transactions pertain;

- (b) causing the counterparty in continuous transactions to provide money, services, or other economic benefits;
- (c) refusing to receive goods in transactions with the counterparty, causing the counterparty to take back such goods after receiving them from the counterparty, delaying payment to the counterparty or reducing the amount of payment, or otherwise establishing or changing trade terms or executing transactions in a way disadvantageous to the counterparty.

If a platformer abuses their bargaining position toward a consumer by processing their personal data, that practice could be deemed an unfair trade practice based on "otherwise establishing or changing trade terms or executing transactions in a way disadvantageous to the counterparty" under Article 2(9)(v)(c) of AMA.

In order to regulate a platform operator's ability to abuse bargaining power relating to the processing of personal information against a consumer, the scope of definitions must be clarified. First, the ASBP Guidelines state that a "'digital platform' has the characteristics of providing third parties with online platforms for various services by using information and communication technologies and data in a way which creates multi-sided markets with multiple user segments and a so-called indirect network effect." [2, p. 2] A "digital platform operator" encompasses a broad range of businesses, such as online shopping markets, sharing economy platforms, and social networking services (SNS). Second, ASBP Guidelines cover "personal information, etc." While "personal information" in these guidelines are identical to Article 2(1) of APPI, which means "information relating to a living individual," ASBP Guidelines also cover "etc." which refers to "information relating to an individual except for personal information [2, p. 3]." This is intended to cover a broader scope of information than APPI.[1] Third, "consumer" refers to an individual, but not one who use the service provided as a business or for business purposes [2, p. 3].

2.2 Exploitative Abuse Against a Consumer

Interpretive challenges exist when addressing exploitative abuse against Japanese consumers. Regulations on superior bargaining power have been primarily designed to protect small and medium-sized enterprises. Therefore, the scope of "counterparty" has not necessarily consumer-protection into consideration [6, pp. 190–191]. AMA does not explicitly restrict some abuses, such as those where consumers are directly harmed by unfair terms or conditions being imposed on them; this is present in other regulations, such as Article 102 of the Treaty on the Functioning of the European Union.

Since the term counterparty itself does not theoretically exclude consumers, exploitative abuse against consumers is increasingly problematic in platform operator services, and protection of personal information is often compromised by abusing a superior bargaining position.

[1] The term "personal data" in ASBP Guidelines covers the same scope of APPI. Under APPI, "personal information" and "personal data" have different definitions. The latter means personal information constituting a personal information database etc. under Article 2(6) of APPI, which is narrower than the former [5].

ASBP Guidelines may be informed by this background information when they allow that "counterparty (in continuous transactions)" includes consumers. ASBP Guidelines say "The personal information, etc. includes all information related to the individual consumer, such as the consumer's personal attributes and activities. Such information is used in the digital platform operator's businesses and thus has economic value. "Therefore, when it is found that consumers provide personal information, etc. in exchange for the use of the services provided by a digital platform operator, then such consumers obviously fall within the definition of a 'counterparty (in continuous transactions)' of the digital platform operator" [2, p. 4].

2.3 Types of Abuse of a Superior Bargaining Position

Regarding what would constitute "unjustly in light of normal business practices," ASBP Guidelines explain that "abuse of a superior bargaining position is determined on a case-by-case basis from the viewpoint of the maintenance and promotion of fair competitive order. "Normal business practices" here are acceptable in terms of the maintenance and promotion of fair competitive order." [2, p. 6]

In order to clarify illicit activities, ASBP Guidelines show several examples of abuse of a superior bargaining position [2, p. 6–11].

The first example is unjustifiable acquisition of personal information, defined by breaking down three patterns: (1) acquiring personal information without stating the purpose of its use to consumers on its webpage or in any other ways; (2) acquiring personal information against consumers' intention beyond the scope necessary to achieve the purpose of use[2]; (3) acquiring personal data without taking the precautions necessary and appropriate for safe management of personal information; (4) causing consumers in continuous use of services to provide other economic information, such as personal information, beyond that required for the use of services.

Second, the guidelines address unjustifiable use of personal information. According to ASBP Guidelines, the issue will arise "if a digital platform operator provides "information relating to an individual except for personal information" from consumers in order to make a third party collate "information relating to an individual except for personal information" acquired from consumers with other information and used for the purpose of causing a disadvantage for consumers" [2, p. 10]. This example would cover profiling, which is discussed in the next section.

The second type is divided into two patterns: (1) using personal information against consumer intention beyond the scope necessary to achieve the purpose of use; (2) using personal data without taking the precautions necessary and appropriate for the safe management of personal information. Pattern (1) includes not only processing personal data beyond its original purpose, but also providing personal data to a third party without obtaining the consent of the consumer concerned.

[2] For instance, this is exemplified by a case that a digital platform operator acquires gender and occupation information from consumers beyond the scope necessary for the sale of goods without obtaining the consumers' consent.

2.4 Sanctions

Abusing a superior bargaining position in transactions constitutes a violation of AMA, which is subject to a cease and desist order under Article 20 and administrative fines that must be paid to the national treasury. The surcharge is an amount equivalent to one percent of the enterprise's sales to the counterparty to the violated act under Article 20-6. Administrative fines were introduced in the 2009 amendment of AMA.

3 APPI

3.1 Overview of the APPI

APPI is one of the Japanese Acts on the Protection of Personal Information enacted in 2003.[3] APPI obligates businesses handling personal information[4] to comply a set of duties as below [5] [5]

- Specifying a purpose for use (Article 15): Specifying the purpose for the use of personal information as explicitly as possible when handling personal information is required;
- Restriction to handle personal data beyond the original purpose (Article 16): Handling personal information beyond the originally specified purpose is prohibited without obtaining an individual's consent in advance;
- Appropriate collection (Article 17): Collecting personal information by deceit or other improper means is prohibited;
- Notification of a purpose or purposes for use when collecting the personal information (Article 18): Promptly notifying the individual of a purpose or purposes for use when collecting personal information is required;
- Accuracy of personal data (Article 19): Striving to keep personal data accurate and up to date is required and immediately deleting personal data when its use has become unnecessary.
- Security of personal data (Articles 20–22): Taking necessary and appropriate measures for the security control of personal data such as preventing the leakage, loss, or damage of handled personal data is required. Measures include supervision over both employees and trustees;
- Restriction on providing personal data to a third party (Article 23): Providing personal data to a third party without obtaining an individual's consent in advance is prohibited.

[3] The other acts include Act on the Protection of Personal Information Held by Administrative Organs, Act on the Protection of Personal Information Held by Independent Administrative Institutions.

[4] This refers to someone handling a personal information database etc. for business use (Article 2(5)). A "personal information database etc." roughly means a systematically organized collective body of information comprising personal information (Article 2(4)).

[5] APPI tentative translation was partially altered to make the provisions clearer.

- Restriction on providing personal data to a third party in a foreign country (Article 24): Providing personal data to a third party in a third country is prohibited except for a case fulfilling the stipulated requirements.

Other than these duties, businesses handling personal information are required to keep a record on a third party provision (Article 25), to confirm specified matters when receiving personal data from a third party (Article 26), and to make identified items of retained personal data public (Article 27). An individual has the right to access, correct, and to cease handling of their retained personal data (Articles 28–30).

One characteristic which defines APPI is its definition of personal information. APPI differentiates "personal information," "personal data," and "retained personal data," depending on the duty concerned. This is to prevent excessive extension of duties. The most fundamental definition is "personal information," which is defined as information relating to a living individual (Article 2(1))[6]. It includes information which can be easily collated with other information and thereby identify a specific individual (Article 2(1)(i)).

APPI established the Personal Information Protection Commission (PPC) by amendment in 2015, and it underwent a subsequent amendment in 2020. The amendment made in 2020 includes clear prohibitions regarding the inappropriate use of personal data, strengthened restriction of conditions that allow personal data to be provided to a third party, and the creation of requirements regarding notifications after a personal data breach [7].

PPC has supervising powers including requiring a report, conducting an onsite inspection (Article 40), issuing guidance and advice (Article 41), and issuing recommendations and orders (Article 42). A business handling personal information which violated an order by PPC would be punished by imprisonment with labor for not more than six months or a fine of not more than 300,000 yen (Article 84).[7] However, enforcement activities by PPC are moderate; it has not yet issued an order,[8] so penal sanctions under Article 84 have not been imposed. The amount of fine as per Article 84 was doubled by the 2020 amendment, but the date of enforcement would be within two years of the promulgation.

3.2 Rikunabi Scandal

Since July 2019, the Rikunabi scandal has shaken public trust. This scandal involved Recruit Career wrongfully handling job seekers' information [8, 9]. Recruit Career is a large platformer service provider for both job seekers and client companies. It had been

[6] More detailed definition is provided in Article 2(1)(i)–(ii).

[7] Other than Article 84, if an operator handling personal information, its employee, or a person who used to be such a business operator or employee has provided or exploited personal information database etc., for the purpose of seeking their own or a third party's illegal profits, they would be punished by imprisonment with labor for not more than one year or a fine of not more than 500,000 yen (Article 83).

[8] An order would be made when a business handling personal information has ignored a recommendation issued by PPC (Article 42(2)). If there is a need to take urgent action due to an event that seriously harms an individual's rights and interests, PPC is authorized to make an imminent order to a business handling personal information (Article 42(3)).

operating a platform called Rikunabi which provided a wide range of employment information. Recruit Career admitted that it made predictions about job-seeking students' odds of declining job offers and sold the data to 38 companies without obtaining proper permission from the candidates. This was done through the Rikunabi DMP Follow service, which was terminated on August 5th, 2019.

Recruit Career stated that it started selling students' data after March 2018, but only to clients who agreed not to use it to make a hiring decision. It explained in its privacy policy that it provided the information to client companies to support hiring activities, but also denied that such information would be used for a hiring decision. If client companies had used the purchased scores for hiring decisions, students' opportunities to obtain formal job offers would have been seriously distorted. The usages of data by client companies are still unknown.

PPC issued administrative recommendations and advice regarding this case in August 2019. The statements to Recruit Career indicated that it had lacked necessary security measures and fulfillment of requirements to provide personal data to third parties [10]. In December 2019, PCC advised companies which purchased data from Recruit Career that they needed to appropriately inform the involved individuals of the purpose of use for their personal data and also properly take control over trustees [11]. The data of around 26,000 individuals was subject to PCC supervision.

On another note, this case has provoked profiling issues. Profiling is regulated under the European Union's General Data Protection Regulation (GDPR) [12] and is defined as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements" (Article 4(4) of GDPR). According to GDPR, a data subject has the right to object to processing of personal data concerning them, including profiling (Article 21 of GDPR). A data subject also has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects them (Article 22 of GDPR). It seems plausible that these provisions could be applied to the Rikunabi case if a similar case took place in Europe. If companies used the scores for their selection processes, that represented a serious infringement of job seekers' right to make decisions about their lives. Calculating the possibility of declining a job offer using AI technology represents a typical case of processing of personal data to evaluate certain personal aspects.

The Rikunabi scandal only affected those in Japan and APPI does not explicitly stipulate provisions on profiling. However, the APPI amendment in 2020 has introduced restrictions on inappropriate use of personal information, and to strengthen requirements regarding providing personal data to a third party. Administrative fines were not introduced in APPI in this amendment.

4 Discussion

The following knowledge can be gained from comparing the APPI and ASBP Guidelines.

First, most of the types of abuse covered by the ASBP Guidelines overlap with APPI provisions. Acquiring personal information without stating the purpose of use in ASBP Guidelines constitutes a violation of Article 15, which requires that the purpose of use be specified. Acquiring personal information beyond the scope necessary to achieve the purpose of use could constitute a violation of the same provision, and acquiring personal data without safe management constitutes violation of security of personal data (Articles 20–22). Only practices causing consumers to provide other economic interests, such as personal information, in addition to the data already provided in exchange for the use of services would be a specific example of abusing a superior bargaining position. If the purpose of use is properly specified, collecting personal information is not legally restricted under APPI. Therefore, if collecting additional personal information is inappropriate in the context of service provision concerned, ASBP Guidelines would be effective. The issue is that concrete examples have not been clarified.

Similarly, unjustifiable uses of personal information under ASBP are covered by APPI. Using personal information against the intention of consumers beyond the scope necessary to achieve the purpose of use constitutes violation of the restriction against handling personal data beyond the original purpose (Article 16 of APPI) as well as restriction regarding providing personal data to a third party (Article 23 of APPI), while using personal data without the safe management constitutes violation of personal data security (Articles 20–22). Regarding this, the 2020 APPI amendment has introduced restrictions on the inappropriate use of personal data and strengthen conditions on providing personal data to a third party, which will broaden restrictions of using personal information. The overlap between ASBP Guidelines and APPI would therefore increase.

In the context of restricting the inappropriate processing of personal information, both APPI and AMA have similar perspectives. The purpose of the ASBP provisions under AMA are to restrict the infringement of freedom of the trading party, which includes the consumer as an individual. APPI aims to protect the personal information of an individual. Both laws could overlap concerning the protection of individuals in terms of handling personal information.

APPI is surely better suited for protecting personal data, but if it proves difficult for APPI to handle the problem, then AMA may address the new challenge.

Attention should also be drawn to the issue of profiling. ASBP Guidelines mentioned a case, for example, where a digital platform operator provides data other than personal information to a third party in order to produce personal information; this is done by collating information with other data, and can disadvantage consumers. This could be considered a profiling issue. In addition, the Rikunabi scandal could have been addressed by ASBP Guidelines, which interpret Article 2(9)(v) of AMA. In this case, Recruit Career was a platform operator, job applicants were counterparties, and the former party had a dominant position over the latter party. Recruit Career inappropriately combined, analyzed, and provided the applicants' data to client companies, which may have caused unfair hiring decisions for applicants.

Although PPC issued formal recommendations and advice to Recruit Career and also issued advices to other involved companies, APPI does not explicitly stipulate profiling provisions, and PPC does not have the authorization to levy administrative fines. Under AMA, an act falling within Article 2(9)(v) is subject to not only cease and desist order under Article 20, but also administrative fines under Article 20-6. This is just one example of how AMA could cover the shortages of APPI.

While AMA would be effective in some cases, there are also challenges. One is that the scope of ASBP which covers inappropriate processing of personal information could expand indefinitely. In the context of a relationship between a digital platform operator and a consumer, the former is deemed to have superior bargaining power against the latter in most cases as a consumer is compelled to accept the terms of use to become a user of the platform service.[9] "One's superior bargaining position over the counterparty" under Article 2(9)(v) of AMA would not be an effective limitation to define the scope. Another complication would be the limited experience of applying administrative fines. Administrative fines due to having abused superior bargaining positions have only been imposed on enterprises in five cases during 2011–2014; all of them are still in dispute [6, p. 123]. No administrative fine based on ASBP has been imposed since 2014, meaning that experience in applying administrative fines due to ASBP is also limited.

Moreover, as a fundamental issue, the theoretical reasons behind why legal rights falling within the sphere of human rights, such as privacy and personal data protection, could be incorporated into economic law needs to be clarified.[10] In this regard, the ASBP Guidelines state, "If services are provided to consumers in the manner which violates APPI, they are interpreted as not having the minimum quality of service level, thereby such services provided for profit are harmful to consumers" [2, p. 6, 10].

"Quality of service" could encompass many categories of value if a lack of such value compromises consumer interests. Privacy and personal data protection are surely included in the categories. However, it should be noted that AMA clearly stipulates its purposes as "to promote fair and free competition, stimulate the creative initiative of enterprise, encourage business activity, [and] heighten the level of employment and actual national income" (Article 1 of AMA). In contrast, privacy and personal data protection aim to ensure the peace of mind of each individual, which is not listed as the purposes of AMA. While AMA has the potential to shield consumers from various kinds of harm, it has an inherent limitation due to the AMA purposes. In order to avoid blurring the scope of law, AMA should carefully ensure that privacy and personal data protection are incorporated into its application to the extent possible.

[9] ASBP Guidelines state "A digital platform operator has a superior bargaining position over consumers who provide personal information, etc. when the consumer, even though suffering detrimental treatment from the digital platform operator, is compelled to accept this treatment in order to use the services provided by the digital platform operator." [2, p. 4–5] This condition is applied to most cases between a digital platformer and a consumer transaction when personal information is a subject of trade.

[10] *See* Ohlhausen, M.K., Okuliar, A.P.: Competition, consumer protection, and the right [approach] to privacy. Antitrust Law J. **80**(1), 121–156 (2015). *See also*, Averitt, N.W., Lande, R.H.: Using the 'consumer choice' approach to antitrust law. Antitrust Law J. **74**(1), 175–264 (2007); Costa-Cabral, F., Lynskey, O.: Family ties: The intersection of data protection and competition law in EU law. Common Mkt. Law Rev. **54**(1), 11–50 (2017).

5 Conclusion

This paper dealt with the division of roles between APPI and AMA, in the context of ASBP, relating to digital platform businesses.

The above discussions indicate that: (1) most types of abuse under ASBP Guidelines presented overlap with APPI provisions; (2) restrictions on ASBP could play a specific role by applying itself to profiling activities that might not be effectively regulated by APPI. However, the possibility for indefinitely expanding the scope of superior bargaining positions and the small amount of experience regarding administrative fines would be challenges to AMA regulation. In addition, clarifying the theoretical reason to incorporate privacy and personal data protection into AMA would be a fundamental issue.

While privacy and personal data must be protected primarily by laws regarding the protection of personal information, competition law and other adjacent legal fields are increasingly significant. These fields can offer a different perspective on how laws can protect personal information.

Other than ASBP, interplay between laws on protecting privacy, personal information, and competition laws arise in cases of refusal to deal, and mergers. The former would involve personal data portability, and the latter raises questions regarding whether privacy could be incorporated into the competition parameter.

Unlike from ASBP, the requirements of AMA and APPI conflict in cases of refusal to deal. If JFTC orders an enterprise to allow competitors to access their data, APPI could be violated since it prohibits a business from providing personal data to a third party without an individual's consent. Interests protected by both laws should be adjusted in this case. As another example, APPI is not applicable in cases of business succession. Article 23(5)(ii) of APPI allows providing personal data to a third party as a result of the succession of business in a merger or otherwise. APPI cannot restrict providing personal data even if personal data might be compromised by a merger. AMA would be expected to protect personal data by incorporating its value into competition parameters[11].

[11] The JFTC Guidelines to Application of the Antimonopoly Act concerning review of Business Combination (latest revision Dec. 17, 2019) shows how to assess the importance of data, by exemplifying a case when Company A has material input goods, such as data, and enters into a conglomerate business combination with Company B. The factors taken into consideration are: (1) what kind of data are held or collected by Company B; (2) how many data are held and how many data are collected daily by Company B from how wide an area; (3) how frequently does Company B collect data; (4) how many data are held or collected by Company B related to the improvement of the service provided by Company A in the product market. It is also considered how advantageous are the data held or collected by Company B as compared with the data available to the competitor (Company X) in the product market of Company A from the above perspectives (1) to (4). *See* Japan Federal Trade Commission. Guidelines to Application of the Antimonopoly Act concerning review of Business Combination (2019), https://www.jftc.go.jp/en/legislation_gls/imonopoly_guidelines_files/191217GL.pdf, p. 57. Though these factors do not directly affect the abuse of the superior bargaining position of a company to a consumer, they could be applied to a merger of companies profiting by handling personal information.

Cooperation or conflict between AMA and APPI needs to be carefully examined depending on the situation.

References

1. Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper (2019). https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html#_Toc484813971
2. Japan Federal Trade Commission. Guidelines Concerning Abuse of a Superior Bargaining Position in Transactions between Digital Platform Operators and Consumers that Provide Personal Information, etc. https://www.jftc.go.jp/en/pressreleases/yearly-2019/December/191217DPconsumerGL.pdf (2019)
3. House of Councillors, the National Diet of Japan. An Act on Improving Transparency and Fairness of Transactions of Specified Digital Platform Operators, (in Japanese). https://www.sangiin.go.jp/japanese/joho1/kousei/gian/201/meisai/m201080201023.htm (2020)
4. Act on Prohibition of Private Monopolization and Maintenance of Fair Trade (Act No. 54 of April 14, 1947). https://www.jftc.go.jp/en/legislation_gls/amended_ama09/index.html (1947)
5. Amended Act on the Protection of Personal Information (Tentative Translation). https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.pdf (2017)
6. Shiraishi, T.: Textbook on Anti-Monopoly Act, 8th edition. (2018). (in Japanese)
7. Personal Information Protection Commission. Amendment of the Personal Information Protection Act. (2020). (in Japanese). https://www.ppc.go.jp/news/press/2020/200612/
8. Asahi Shimbun. Recruit Career sold student data to firms without explicit consent (2019). http://www.asahi.com/ajw/articles/AJ201908020059.html
9. Asahi Shimbun. Rikunabi scandal highlights risks of exploitation of personal data (2019). http://www.asahi.com/ajw/articles/AJ201908130019.html
10. Personal Information Protection Commission. Recommendation, etc., under Article 42(1) of APPI (2019), (in Japanese). https://www.ppc.go.jp/files/pdf/190826_houdou.pdf
11. Personal Information Protection Commission. Corrective Measures under APPI (2019), (in Japanese). https://www.ppc.go.jp/files/pdf/191204_houdou.pdf
12. Union, European: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official J. L **119**, 1–88 (2016)

The Laws of Big Data

How Data Protection Law, Competition Law and Contract Law Deal with the Challenges of a Data-Driven World

Lena Mischau[1,2](✉)

[1] Weizenbaum Institute for the Networked Society, Hardenbergstraße 32,
10623 Berlin, Germany
[2] Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
lena.mischau@rewi.hu-berlin.de

Abstract. This paper presents a selection of legal topics in the context of data analytics and Big Data from a lawyer's perspective. After introducing the reader to the role of law, both in the analogue and the digital world (1), the paper gives a systematic overview of some of the currently most relevant data-related legal topics (2). While digitalisation and data processing poses new questions to all areas of law, this paper focusses on the role Big Data plays in competition, data protection and contract law, as those are closely interlinked and address similar data-related phenomena. The paper was written from a mainly European perspective and presents some specific approaches European law takes to address the challenges we face with the advent of Big Data.

Keywords: Competition law · Data protection law · Contract law

1 Introduction–the Role of Law in a Data-Driven World

Law serves multiple functions in society. Contract law, for instance, aims to provide a clear legal framework for diverse parties to conclude contracts and exchange goods and services in a fair manner. It aims to rebalance interests and asymmetries of information or negotiation power between businesses and consumers, but also between businesses themselves. For instance, European consumer law imposes information and transparency obligations on sellers [1], prohibits particularly unfair and problematic provisions in General Terms and Conditions [2], and imposes certain rights and duties on the parties in cases where the product is not in conformity with the contract [3]. Competition law, in turn, wants to protect and foster competition with all its positive effects by prohibiting cartels and any abuse of dominant market positions, and by controlling mergers with regards to the impact they will likely have on competition [4]. Data protection law first and foremost aims at protecting privacy and self-determination of data subjects, that is, of individuals [5].

This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant No. 16DII111 ("Deutsches Internet-Institut").

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 16–28, 2020.
https://doi.org/10.1007/978-3-030-62803-1_2

All of these areas of law existed long before the rise and success of the online platform economy and data ecosystems involving the Internet of Things ("IoT"). Nevertheless, they apply to the digital world just as they do to the analogue one. It used to be a widespread misperception that the digital world and the Internet in particular were "free" from the law. However, no space – neither analogue nor virtual – exists that is not subject to any legal rules at all. Law continues to pursue its respective societal objectives in digital environments. Power asymmetries should be rebalanced, consumers protected and self-determination and privacy ensured online too. The legal imperatives are all the more vital given the increasing blurring of lines between the online and offline world, for instance, with regards to smart products and IoT. However, it is true indeed that *enforcing* the law in digital contexts is often much more difficult than in analogue ones. Irrespective of the question of enforcement, the legal rules as such often do not even need to be adapted to digital cases and it does not make any difference from a legal point of view whether a certain event takes place online or offline.

Example: Under European law, for instance, the buyer of a laptop has the same legal rights should the laptop prove to be defective a few months after purchase, regardless of whether he bought it in a physical shop next door or via an online shop [6]. Similarly, competition law works according to the same principles regardless of whether an undertaking does business in a traditional market or in a digital market, such as for online search engines. In both cases, the same essential rules apply. Finally, data protection law treats data processing in mainly the same way, regardless of whether the data processor is, for instance, a company that asks you to fill in a paper questionnaire that will be integrated into a filing system, or whether the company sends you an email with a link to an online form.

Nevertheless, digitalisation is confronting our societies with a number of novel phenomena that the law in its current state is not always able to address adequately. The unprecedented availability of enormous quantities of data – often of very high quality – that can be processed at high speed ("Big Data") [7] is one of these new phenomena. Big Data paves the way for progress that the whole of society can benefit from [8]. For instance, Big Data can help us make processes more efficient, increase knowledge by demonstrating correlations we would otherwise not be able to see, and improve forecasts [9]. Businesses and public bodies can benefit from using Big Data applications in very similar ways [10]. As a result, Big Data can have very positive impacts for individuals and for society. For instance, patients might enjoy better healthcare services thanks to optimised medical treatment or early identification of individual health risks [11]. Technology based on Big Data such as intelligent traffic systems and data use in urban systems could increase security and the mobility of citizens [12]. In addition, the environment could also benefit from Big Data, where more efficient processes help avoid pollution and unnecessary waste of resources, or where Big Data applications help stabilise energy grids to be able to cope with the volatility of renewable energy sources, such as solar or wind power [13]. At the same time, we need to be aware of and fight the dangers that come with Big Data, such as potential discrimination, erroneous decision-making based on inaccurate data, and privacy breaches [14]. It is up to the law to establish a legal framework that enables us to benefit from the opportunities while protecting us from the risks of Big Data.

2 A Selection of Legal Topics in the Context of Big Data

Big Data poses specific and complex challenges for data protection law, competition law and contract law. It makes sense to shed light on these three areas of law together, as they interact very closely, pursuing objectives that are sometimes very similar, other times very different [15].

2.1 Big Data and Data Protection Law

Data protection law is the main legal field dealing with Big Data. Throughout the last years, we have witnessed a large number of data protection scandals [16], such as Cambridge Analytica [17], that illustrate both the various dangers of collecting personal data on a large scale and the importance of effective privacy protection.

According to the European General Data Protection Regulation ("GDPR") [18], any processing of personal data is forbidden unless the processing is justified by a legal ground listed in Art. 6(1) GDPR [19]. Such a justification may be given where the data subject – in other words, the natural person whose data are being processed – has freely given her explicit consent for the specific purpose of data processing (Art. 6(1)(a) GDPR). In addition to the receipt of explicit consent, data processing can also be legitimate if it is necessary to perform a contractual obligation (Art. 6(1)(b) GDPR), if it is necessary to comply with a legal obligation of the data controller (Art. 6(1)(c) GDPR), if it is necessary to protect vital interests of the data subject or another natural person (Art. 6(1)(d) GDPR) or if it is necessary to perform a task carried out in the public interest (Art. 6(1)(e) GDPR). However, the processing may also be justified due to legitimate interests of the controller or third parties, provided that these interests are not overridden by the data subject's interests (Art. 6(1)(f) GDPR). In any case of data processing, basic principles need to be respected, for instance, the principles of lawfulness, fairness and transparency (Art. 5(1)(a) GDPR), and the principles of integrity and confidentiality (Art. 5(1)(f) GDPR). Moreover, the GDPR provides specific rules for certain aspects of Big Data, such as data processing on a large scale, systematic monitoring of publicly available areas and profiling, as those are considered as particularly risky [20].

Example: A consumer would like to make use of an online food delivery service. Before he can make an order, he is asked to fill in his name, address and bank account details in an online form. In addition, he is asked to answer certain questions, such as how he became aware of that particular food delivery service, what his favourite dishes are, whether he likes cooking as a hobby, and whether he has any specific allergies or illnesses that require certain diets.

Regarding his name, address and bank account details, no consent would be needed, because the food delivery service could not provide the service without the information (Art. 6(1)(b) GDPR). In contrast, any processing of the additional information mentioned above would only be justified if the consumer grants his consent (Art. 6(1)(a) GDPR). The information about potential allergies falls within the scope of health data and merits special protection (Art. 9 GDPR). In all cases, the service provider needs to inform the consumer about certain aspects of the data processing, including, for example, its purpose, the storage period and the consumer's rights, such as the right to access his data (Art. 13(1)(c), (2)(a)(b) GDPR).

Although these rules seem restrictive at first glance, effective data protection faces a large number of difficulties in practice. To start with, data protection policies are usually very long and complex, and consumers usually do not read them [21]. Consumers therefore lack sufficient information as to what happens to their data and are not able to give consent in an informed and self-determined way. Several initiatives and projects try to tackle this issue. For instance, so called "privacy icons" aim to visualise certain data processing aspects and to help consumers better understand the way their data will be processed [22]. Personal Information Management Systems (PIMS) promise to support consumers in enforcing their privacy preferences by, for example, automatically allowing or forbidding data processing according to these preferences [23]. Others, in contrast, consider moving away from consent and instead determining on a societal – rather than individual – level what kinds of data processing should be allowed or forbidden [24].

Another issue in this regard is closely connected to both contract law and competition law. Many digital businesses claim to provide digital goods or services to the consumer "for free", that is they do not ask for any monetary compensation to be paid by the consumer. However, they then finance their business models by other means, which may include offering advertisement services to third parties [25], for example. Nowadays, such types of online advertisement often are personalised using data available about an individual consumer or about groups of persons similar to the individual consumer in question. In such a context, the businesses with access to the biggest and most detailed consumer profile and preference datasets will benefit the most. One might therefore say that consumers do "pay" for these services, but with their personal data and/or with the attention they dedicate to the advertisements shown them based on their profiles [26].

At the same time, Art. 7(4) GDPR is very critical towards practices where the "performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract". The provision does not prohibit such business models in general, but emphasises that consent must be voluntary [27]. If this is not the case, the data processing in question is not legitimate [28]. The reason for this rejection lies in the nature of Big Data and possible long-term consequences both for individuals and societies as a whole. Perceiving a product as "free" can have great influence on the consumer's behaviour [29] and impede him from taking a fully self-determined decision regarding his privacy. In this context, one also needs to be aware of the potential danger of the emergence of a "two-class society" regarding data protection – that is, a society consisting of a consumer group that is aware of the risks of disclosing data and that has the financial means to use paying alternatives, and a – potentially much bigger – group of consumers who are not aware or who are aware but cannot afford more privacy-friendly and therefore more expensive products [30]. In practice, many online services continue to tie the use of their services to user consent. It is up to the data protection authorities and courts to increase legal certainty and enforcement.

Example: A consumer would like to read an article of an online newspaper. Before he can access it, the newspaper asks the consumer whether he would like to pay a monetary price of 0,99 EUR for the single article (option A) or of 5,99 EUR for a one-month subscription to the online newspaper (option B) or whether he prefers not to pay any monetary price, but to give his consent to the processing of his personal data for advertisement purposes (option C). He chooses option C.

In this scenario, the consumer's consent can be considered as "freely given", because he had a real choice of whether or not he wanted to give consent in order to read the article by having at least one realistic alternative (option A, option B).

However, even where data processing takes place in perfect compliance with data protection law and where consumers actually are able to make a free and informed choice of whether or not they want to give consent, self-determination regarding personal information is fragile in light of the analytical potential of Big Data. Even if individuals behave in very privacy-protective ways and withhold consent, this does not hinder Big Data applications to reveal probable information about them as soon as a sufficient amount of data is available about other persons who share similar traits or behave in similar ways [31].

2.2 Big Data and Competition Law

After data protection law, it is competition law that is the most concerned with Big Data. On the one hand, the availability of large datasets in almost real-time can stimulate competition by fostering innovative business models and boost the performance of existing businesses [32]. On the other hand, Big Data also poses several risks to competition.

Competition law comprises various tools to protect and foster competition. Long before other areas of private law, competition law scholars already discussed the particular characteristics of multi-sided markets and the platform economy in depth [33]. For example, the question was asked whether a market in the meaning of competition law is established in cases where business models are data-driven and digital products advertised as "free" [34]. Today, the answer to this question is clearly in the affirmative [35]. One of the European member states (Germany) even changed its law to explicitly clarify that such constellations may also be subject to the control of abusive practices [36]. This development is very welcome, as the alternative would be that such businesses might escape certain competition law restrictions despite their real impact on the economy. Considering that businesses offering "free" services to consumers, such as online search engines and social networks, have some of the highest market capitalisations in the world [37], such a result would be absurd.

More importantly, competition law scholars and authorities have examined the impact of data power on the concept of market power, that is, whether access to certain kinds of data helps businesses gain or defend a dominant position in a specific market. While the answer varies from case to case, some general observations can be made. The ability to access large amounts of very granular data in almost real-time clearly has a particularly positive impact on a company's market position in cases where the access to that kind of data is important to be successful in the specific market, where competitors cannot access these data and where network effects and economies of scale exist and are reinforced by the availability of data [38]. Network effects are particularly important in this context. The term refers to positive feedback effects within networks and multi-sided markets that appear either between users of the same user group or between two or more different user groups [39]. For example, social networks collect vast amounts of data about their users. The more users take part in a social network, the more attractive it becomes for other users (direct network effects). The more members a social network has and the more those make use of it, the more data it creates. And the more exclusive data a social network has at its disposal, the more it can improve its services, for instance, by showing more interesting information to individual users or offer more elaborate product features such as calendar functions. As a result, users are even more attracted by the social network in question and even more data is created [40]. In that case, markets tend to "tip" and only one "winner" takes the whole market, meaning that competition is effectively restricted in that market [41].

Example: An undertaking has a dominant market position in a specific market for search engine services. Because of its dominant market position, the undertaking is subject to particular competition rules to make sure that it does not abuse its dominant market position (Art. 102 AEUV). For example, the undertaking would not be allowed to systematically rank its own products or services (other than the search engine itself) in a higher position than those of its competitors [42].

The importance of access to data has been widely acknowledged over the years. More recently, the European Commission has re-affirmed its intention to foster data access as part of its "European Data Strategy" [43]. Germany, for instance, already adapted its Act against Restraints of Competition (ARC) in order to stress the importance of exclusive access to data relevant for competition [44]. In rare cases, data might even become so important that parallels can be drawn with infrastructure facilities. Refusing access to these kinds of data might be considered as an abuse of market dominance and might lead to access claims granted to other companies [45]. Access to data is an important topic for competition law as a whole, but also for certain sectors in particular – for the automobile [46] and financial services sectors [47], notably.

Example: An independent garage offers repair and maintenance services for cars outside the car manufacturer's own distribution system. Nowadays such a garage can only perform those services, if, among others, it is able to have access to certain vehicle on-board diagnostic information, vehicle repair and maintenance information, as well as the necessary software. In this respect, the access to non-real-time data may not be sufficient in the future anymore. New services, such as predictive maintenance, may require access to full real-time data.

Access to data for competition and economic reasons can never be granted without limits. On the contrary, among others, data sharing always needs to take place within the limits of data protection law in particular [48]. At the same time, data sharing also raises competition law concerns in cases where data are made available that reveal sensitive information that is important for competition [49]. The increasing interaction of competition law, data protection law and contract law becomes particularly evident in the current Facebook proceedings in Germany. The German competition authority (Federal Cartel Office) had considered Facebook's practice to collect and process data from third party sources as abusive [50], while the Düsseldorf Court of Appeal disagreed and suspended the decision [51].

2.3 Big Data and Contract Law

The "free" goods and services business model has also raised several questions for contract law. In the European Union, the new Digital Content and Digital Services Directive ("DCSD") grants certain rights to consumers in the context of contracts for digital content and digital services [52]. One of the main questions throughout the legislative procedure was whether or not (personal) data could be considered as counter-performance. On the one hand, one needs to be aware of the fundamental rights dimension of personal data and be reluctant towards any commercialisation [53]. On the other hand, we need to make sure that consumers who provide data instead of a monetary price also get to benefit from consumer protection, for example, in cases where a digital service is defective [54]. Contract law and data protection law should thus go hand in hand to provide consumers with protection within their respective scope of application [55].

Other important contract law issues related to data and Big Data in particular concern the questions of security, compatibility, interoperability, updates and potential changes of data-driven products. In the European Union, the DCSD grants consumers and businesses certain rights in this regard for the very first time [56]. For IoT products similar rules exist [57].

Example: A consumer enters into a contract with a provider of music streaming services. To do so, he can choose between a paying option (A) and a non-paying option where the provider processes the consumer's personal data in order to offer personalised advertisement for third parties and create revenues (B). The consumer chooses the non-paying option (B). One year later, the consumer is not happy with the service anymore, for example, because the service does not function properly, is not compatible with the consumer's devices anymore, or because the provider does not offer any necessary security updates (cf. Art. 7(a), (d), Art. 8(1)(b), (2) DCSD).

Under certain circumstances, the consumer has now the right to have the streaming service brought into conformity or to terminate the contract, although he did not pay any monetary price but provided personal data (Art. 14(1)-(4) DCSD). In addition to his contractual rights under the DCSD, the consumer can still exercise his data protection rights under the GDPR, for example, by revoking the consent he gave for processing his data (Art. 3(8), recital 39 DCSD, Art. 7(3) GDPR).

Furthermore, the topic of data sharing that we have already seen in the context of competition law plays an important role in contract law too. Regarding data from the private sector, many data sharing models can be observed [58]. However, no harmonised horizontal rules exist in this regard. So far, the European Commission has elaborated basic principles and a check-list for the sharing of IoT data between businesses, and between businesses and the public sector [59]. While these are not legally binding, a new "Data Act" is expected to be proposed in 2021 [60].

Example: Company A and company B would like to exchange non-personal data that was created by IoT objects, such as smart machines in a factory. According to the European Commission's guidance, both parties need to comply with certain principles. For example, they should make sure to respect each other's commercial interests and secrets, ensure undistorted competition and minimise data lock-ins.

3 Conclusion

Big Data already falls under the scope of several fields of European law today and more initiatives to adapt existing law or establish new rules are expected within the next few years. Some of the most important areas of law one needs to consider in this context are data protection, competition and contract law. Several data-related topics are addressed by all of these – albeit from different perspectives and in different ways – such as the phenomenon of "free" services and the importance of access to or the protection of certain kinds of data. In many of these data-related cases, data protection law, competition and contract law interact closely and in very complex ways. It is therefore necessary to examine a specific Big Data scenario carefully from all three perspectives and assess how those impact each other.

Of course, other fields of law may also be concerned with Big Data. Public law, for example, has to determine those cases where the processing of data is considered justified under the GDPR because of public interests (Art. 6(1)(e), (3) GDPR) [61]. Criminal law, in turn, will have to deal with an increasing number of data-related crimes, such as data espionage, phishing and providing access to stolen data [62]. IP law, in particular, faces a broad range of challenges concerning the question of whether and to what extent data should be protected by copyright law, as part of databases or trade secrets, and how text and data mining should be treated [63]. Moreover, there has been an intense debate about whether or not new absolute rights are needed, in particular regarding non personal data, such as a "data producer's right" [64].

To fully address Big Data topics, law therefore requires a truly holistic approach. This may include reconsidering carefully the scope of application, the purposes and the means to achieve those purposes of the individual legislative acts.

References

1. For instance, Directive (EU) 2011/83 of 22 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council OJ L 304/64 (2011)
2. For instance, Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts OJ L 95/29 (1993)
3. For instance, the buyer of a product may have the right to receive a proportionate reduction in the price where the acquired product is not in conformity with the contract. See Directive (EU) 2019/771 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the sale of goods, amending Regulation (EU) 2017/2394 and Directive 2009/22/EC, and repealing Directive 1999/44/EC OJ L 136/28 (2019)

[4] Treaty on the Functioning of the European Union (TFEU) [2012] OJ C 326/47, arts 101, 102; Council Regulation (EC) No 139/2004 of 20 January 2004 on the control of concentrations between undertakings (the EC Merger Regulation) OJ L 24/1 (2004)

[5] Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJ L 119/1 ("GDPR") (2016)

[6] Cf Directive (EU) 2019/771 (n 3)

[7] Regarding the term "Big Data", see, for instance, Laney, D.: 3-D Data Management: Controlling data volume, velocity, and variety. Application Delivery Strategies, META Group Inc. (2001). http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Accessed 24 June 2020; Monopolkommission: Competition policy: The challenge of digital markets. Special Report No 68, (2015), paras 67ff, http://www.monopolkommission.de/images/PDF/SG/s68_fulltext_eng.pdf. Accessed 24 June 2020; Stucke, M., Grunes A.: Big Data and Competition Policy. OUP, Oxford, paras 2.04ff (2016)

[8] For an in-depth overview of the diverse risks and opportunities of Big Data, see, for instance, Tene, O., Polonetsky, J.: Big Data for All: Privacy and User Control in the Age of Analytics. Nw. J. Tech. Intell. Prop. **11**(5) pp. 239, 243ff, 251ff (2013)

[9] Bollier, D.: The Promise and Perils of Big Data. The Aspen Institute, Washington, DC, pp. 20ff (2010)

[10] For some examples in practice, see Bartel, J., et al.: Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte. BITKOM e. V., Berlin, pp. 51ff (2012), https://www.bitkom.org/sites/default/files/file/import/BITKOM-LF-big-data-2012-online1.pdf. Accessed 24 June 2020

[11] Minssen, T., Schovsbo J.: Big Data in the Health and Life Sciences: What Are the Challenges for European Competition Law and Where Can They Be Found? In: Seuba, X., Geiger, C., Penin, J. (eds.), Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data. Global Perspectives for the Intellectual Property System, CEIPI and ICTSD, Geneva and Strasbourg, pp. 121, 123 (2018); Federal Trade Commission: Big Data – A Tool for Inclusion or Exclusion? p. 5 (2016). https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf. Accessed 24 June 2020; Bartel, J., et al. (n 10) pp. 70, 76

[12] European Commission: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European strategy on Cooperative Intelligent Transport Systems, a milestone towards cooperative, connected and automated mobility. COM(2016) 766 final; OECD: Data-Driven Innovation: Big Data for Growth and Well-Being. OECD Publishing, Paris, pp. 379ff (2015). https://dx.doi.org/10.1787/9789264229358-en; Bartel, J., et al. (n 10) p. 69

[13] OECD (n 12) p. 383; Zhang,Y., Huang, T., Bompard, E.F.: Big data analytics in smart grids: a review. Energy Inform. **1**(8) pp. 2, 9f, 12, 15f. (2018). https://doi.org/10.1186/s42162-018-0007-5

[14] See, for instance, Federal Trade Commission (n 11) pp. 8ff; Bollier, D. (n 9) pp. 23, 31

[15] Cf European Data Protection Supervisor: Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy. Preliminary Opinion (2014). https://edps.europa.eu/sites/edp/files/publication/14-03-26_competitition_law_big_data_en.pdf. Accessed 24 June 2020; cf Crémer, J., de Montjoye, Y.-A., Schweitzer, H.: Competition Policy for the Digital Age: Final Report. Publications Office of the European Union, Luxembourg, pp. 76ff (2019). https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf. Accessed 24 June 2020

16. For an overview see, for instance, Bonneau, V., et al.: Digital Transformation Monitor, Big Data: a complex and evolving regulatory framework. IDATE et al. (eds.), European Union, p. 3 (2017). https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_Big%20Data%20v1_0.pdf. Accessed 24 June 2020. For privacy breaches in the context of statistical data, see Nissim, K., et al.: Bridging the gap between computer science and legal approaches to privacy. Harvard J. Law Technol. **31**(2), pp. 687, 700ff (2018)
17. Cadwalladr, C., Graham-Harrison, E.: Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian (17 March 2018). https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. Accessed 24 June 2020
18. See (n 5). For an introduction to US privacy laws, see, for instance, Nissim, K., et al. (n 16) pp. 706ff
19. For sensitive data, see specific rules in GDPR, art 9(2)
20. GDPR, art 35(3); Article 29 Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. WP 248 rev.01 (2017) pp. 8ff. http://ec.europa.eu/newsroom/document.cfm?doc_id=47711. Accessed 24 June 2020; cf Efroni, Z., et al.: Privacy icons: a risk-based approach to visualisation of data processing. EDPL **5**(3) pp. 352, 361ff (2019)
21. See, for instance, Efroni, Z., et al. (n 20) pp. 355f with further references
22. Efroni, Z., et al. (n 20) pp. 357ff. Cf Hansen, M.: Putting Privacy Pictograms into Practice - a European Perspective. In: Fischer, S., Maehle, E., Reischuk, R. (eds.), Informatik 2009: Im Focus das Leben, Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 28.9.-2.10.2009, Lübeck. Gesellschaft für Informatik e.V., Bonn (2009)
23. European Data Protection Supervisor, EDPS Opinion on Personal Information Management Systems: Towards More User Empowerment in Managing and Processing Personal Data. Opinion 9/2016, paras 5ff (2016). https://edps.europa.eu/sites/edp/files/publication/16-10-20_pims_opinion_en.pdf. Accessed 24 June 2020; Horn, N., Riechert, A., Müller, C.: Neue Wege bei der Einwilligung im Datenschutz – technische, rechtliche und ökonomische Herausforderungen. Stiftung Datenschutz, pp. 10ff (2017). https://stiftungdatenschutz.org/fileadmin/Redaktion/Bilder/Abschluss_Studie_30032017/stiftungdatenschutz_broschuere_20170611_01.pdf. Accessed 24 June 2020; Betkier, M.: Privacy Online, Law and the Effective Regulation of Online Services. Intersentia, Cambridge, pp. 79ff (2019)
24. See, for instance, Radlanski, P.: Das Konzept der Einwilligung in der datenschutzrechtlichen Realität. Mohr Siebeck, Tübingen, pp. 97, 204ff, 232f (2016)
25. Monopolkommission (n 7) paras 39f
26. Weber, R.H.: Information at the crossroads of competition and data protection law. Zeitschrift für Wettbewerbsrecht **12**(2), pp. 169, 175 (2014). For the contract law perspective on this topic, see ch 2.3
27. Metzger, A.: Data as Counter-Performance: What Rights and Duties do Parties Have? JIPITEC **8**(1) p. 2, para 12 (2017)
28. Article 29 Working Party: Guidelines on Consent under Regulation 2016/679. WP259 rev.01, pp. 5ff, (2017). https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51030. Accessed 24 June 2020
29. Friedman, D.A.: Free Offers: A New Look. New Mexico L Rev **38**(1) pp. 49ff (2008)
30. Cf Krohm, N., Müller-Peltzer, P.: Auswirkungen des Kopplungsverbots auf die Praxistauglichkeit der Einwilligung: Das Aus für das Modell „Service gegen Daten"? ZD **7**(12) pp. 551, 553 (2017); Härting, N.: „Dateneigentum" – Schutz durch Immaterialgüterrecht? Was sich aus dem Verständnis von Software für den zivilrechtlichen Umgang mit Daten gewinnen lässt. CR **36**(10) pp. 646, 648 (2016)

31. Barocos, S., Nissenbaum, H.: Big Data's End Run around Anonymity and Consent. In: Julia Lane et al. (eds.), Privacy, Big Data, and the Public Good: Frameworks for Engagement. CUP, Cambridge pp. 44, 61ff (2004) call this phenomenon "the tyranny of the minority"; Hermstrüwer, Y.: Contracting Around Privacy: The (Behavioral) Law and Economics of Consent and Big Data. JIPITEC **8**(1) pp. 9, 12ff, paras 11ff (2017). In the context of statistical data, see Nissim, K., et al. (n 16) pp. 700ff
32. Federal Cartel Office: Big Data und Wettbewerb. In: Federal Cartel Office (ed.), Schriftenreihe Wettbewerb und Verbraucherschutz in der digitalen Wirtschaft, Bonn, p. 9, (2017). http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Schriftenreihe_Digitales/Schriftenreihe_Digitales_1.pdf?__blob=publicationFile&hx0026;v=3. Accessed 24 June 2020
33. For instance, Evans, D.S.: The Antitrust Economics of Multi-Sided Platform Markets. Yale Journal on Regulation **20**(2) pp. 324ff (2003); Evans, D.S., Noel, M.: Defining Antitrust Markets When Firms Operate Two-Sided Platforms. Colum. Bus. L. Rev. **2005**(3) pp. 101ff (2005)
34. Cf text to n 25 in ch 2.1 of this paper
35. For Germany see, for instance, Federal Cartel Office: Clearance of Merger of Online Real Estate Platforms. Case summary B6-39/15, p. 3 (2015); Federal Cartel Office: Acquisition of the online comparison platform Verivox by ProSiebenSat.1. approved. Case summary B8-76/15, p. 3 (2015); Podszun, R., Franz, B.: Was ist ein Markt? – Unentgeltliche Leistungsbeziehungen im Kartellrecht. NZKart **3**(3) pp. 121ff (2005)
36. Act against Restraints of Competition in the version published on 26 June 2013 (Federal Law Gazette 2013 I, 1750, 3245), as last amended by art 10 of the law of 12 July 2018 (Federal Law Gazette 2018 I, 1151) ("ARC"), § 18(2a). A translation is provided by the Language Service of the Bundeskartellamt and Renate Tietjen. http://www.gesetze-im-internet.de/englisch_gwb/englisch_gwb.html. Accessed 24 June 2020
37. Forbes: GLOBAL 2000: The World's Largest Public Companies. (2020). www.forbes.com/global2000/list/. Accessed 24 June 2020
38. Federal Cartel Office (n 32) pp. 7f
39. OECD: Rethinking Antitrust Tools for Multi-Sided Platforms pp. 189f (2018). www.oecd.org/competition/rethinking-antitrust-tools-for-multi-sided-platforms.htm. Accessed 24 June 2020
40. These feedback loops can be referred to as "snowball effects"; Federal Cartel Office (n 32) pp. 7f
41. See, for instance, Barwise, T., Watkins, L.: The Evolution of Digital Dominance: how and why we got to GAFA. In: Moore, M., Tambini, D. (eds.), Digital dominance: The power of Google, Amazon, Facebook, and Apple. OUP, New York (2018), pp. 21, 22ff; Federal Cartel Office: Working Paper – The Market Power of Platforms and Networks, Executive Summary, B6-113/15. p. 9 (2016). www.bundeskartellamt.de/SharedDocs/Publikation/EN/Berichte/Think-Tank-Bericht-Zusammenfassung.pdf?__blob=publicationFile&Cv=4. Accessed 24 June 2020
42. European Commission: AT.39740 Google Search (Shopping); European Commission: Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service – Factsheet. (2017). https://ec.europa.eu/commission/presscorner/detail/en/MEMO_17_1785. Accessed 24 June 2020
43. European Commission: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European strategy for data. COM(2020) 66 final, pp. 12ff

44. ARC, § 18(3a). A new draft proposal for further amendments to the ACR puts even more emphasis on data access; see Federal Ministry for Economic Affairs and Energy: Referentenentwurf des Bundesministeriums für Wirtschaft und Energie, Entwurf eines Zehnten Gesetzes zur Änderung des Gesetzes gegen Wettbewerbsbeschränkungen für ein fokussiertes, proaktives und digitales Wettbewerbsrecht 4.0, GWB-Digitalisierungsgesetz. (2020). https://www.bmwi.de/Redaktion/DE/Downloads/G/gwb-digitalisierungsgesetz-referentenentwurf.pdf?__blob=publicationFile&hx0026;v=10. Accessed 24 June 2020. For more details, see Mischau, L.: Market Power Assessment in Digital Markets – a German Perspective. GRUR Int. **69**(3), pp. 233, 245ff (2020)
45. Regarding data and the "essential facilities doctrine", see, for instance, Graef, I.: EU competition law, data protection and online platforms: Data as Essential Facility. Kluwer Law International, Alphen aan den Rijn, pp. 249ff (2016); Drexl, J.: Designing Competitive Markets for Industrial Data – Between Propertisation and Access. JIPITEC **8**(4) pp. 257, 278ff (2017); Schweitzer, H. et al.: Modernisierung der Missbrauchsaufsicht für marktmächtige Unternehmen: Endbericht. DICE Consult, pp. 131ff (2018). https://www.bmwi.de/Redaktion/DE/Publikationen/Wirtschaft/modernisierung-der-missbrauchsaufsicht-fuer-marktmaechtige-unternehmen.pdf?__blob=publicationFile&hx0026;v=15. Accessed 24 June 2020
46. Cf Regulation (EU) 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC OJ L 151/1, art 61(1) (2018); Kerber, W., Gill, D.: Access to Data in Connected Cars and the Recent Reform of the Motor Vehicle Type Approval Regulation. JIPITEC **10**(2), pp. 244ff (2019)
47. Cf Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC OJ L 337/35, art 66(1) (2015), (4)(b), art 67(1), (3)(b)
48. Crémer, J., de Montjoye, Y.-A., Schweitzer, H. (n 15) pp. 77ff; Bourreau, M., de Streel, A., Graef, I.: Big Data and Competition Policy: Market Power, Personalised Pricing and Advertising: Project Report. CERRE, pp. 15ff (2017). https://www.cerre.eu/sites/cerre/files/170216_CERRE_CompData_FinalReport.pdf. Accessed 24 June 2020
49. Crémer, J., de Montjoye, Y.-A., Schweitzer, H. (n 15) pp. 77ff; Federal Ministry for Economic Affairs and Energy: A new competition framework for the digital economy: Report by the Commission Competition Law 4.0. BMWi, Berlin, pp. 56ff (2019). https://www.bmwi.de/Redaktion/EN/Publikationen/Wirtschaft/a-new-competition-framework-for-the-digital-economy.pdf?__blob=publicationFile&v=3. Accessed 24 June 2020
50. Federal Cartel Office: B6–22/16 Facebook, paras 136ff; Federal Cartel Office: Bundeskartellamt prohibits Facebook from combining user data from different sources: Background information on the Bundeskartellamt's Facebook proceeding, p. 5 (2019). www.bundeskartellamt.de/SharedDocs/Publikation/EN/Pressemitteilungen/2019/07_02_2019_Facebook_FAQs.pdf?__blob¼publicationFile&v¼6. Accessed 24 June 2020
51. Düsseldorf CA, 26 August 2019, Kart 1/19 (V) Facebook
52. Directive 2019/770/EU of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services OJ L 136/1 (2019) ("DCSD")
53. European Data Protection Supervisor: Opinion 4/2017 on the Proposal for a Directive on certain aspects concerning contracts for the supply of digital content. paras 14ff (2017). https://edps.europa.eu/sites/edp/files/publication/17–03-14_opinion_digital_content_en.pdf. Accessed 24 June 2020

54. DCSD, art 3(1), recital 24. Metzger, A. et al.: Data-Related Aspects of the Digital Content Directive. JIPITEC **9**(1), pp. 90, 93ff, paras 12ff (2018)
55. Langhanke, C., Schmidt-Kessel, M.: Consumer Data as Consideration. EuCML **4**(6), pp. 218, 219f (2015). Regarding the complex interplay between the DCSD and the GDPR, see Metzger, A.: A Market Model for Personal Data: State of the Play under the New Directive on Digital Content and Digital Services. In Lohsse, S., Schulze, R., Staudenmayer, D. (eds.), Data as Counter-Performance – Contract Law 2.0? Münster Colloquia on EU Law and the Digital Economy V, Nomos, Baden-Baden (2020)
56. DCSD, art 7(a), (d), art 8(1)(b), (2), art 19, recitals 42, 47
57. Directive (EU) 2019/771 (n 3), arts 6(a), (d), 7(1)(d), (3), recitals 28, 30f. In contrast to the DCSD, these rules for IoT products only apply in cases where the consumer has paid a monetary price in exchange for the product
58. European Commission: Commission Staff Working Document: Guidance on sharing private sector data in the European data economy. SWD(2018) 125 final, pp. 5, 8ff
59. European Commission (n 58) pp. 3, 6ff
60. European Commission (n 43) pp. 13, 15
61. One example is the European eCall regulation, which allows the processing of certain personal data in cases of car accidents in order to accelerate the rescue work of the police and the fire brigarde; see Regulation (EU) 2015/758 of the European Parliament and of the Council of 29 April 2015 concerning type-approval requirements for the deployment of the eCall in-vehicle system based on the 112 service and amending Directive 2007/46/EC OJ L 123/77 (2015); Klink-Straub, J., Straub, T.: Nächste Ausfahrt DS-GVO – Datenschutzrechtliche Herausforderungen beim automatisierten Fahren. NJW **71**(44), pp. 3201, 3203 (2018)
62. Cf, for instance, the German Criminal Code in the version published on 13 November 1998 (Federal Law Gazette I, p. 3322), as last amended by Article 2 of the Act of 19 June 2019 (Federal Law Gazette I, p. 844), §§ 202a-202d. A translation is provided by Prof. Dr Michael Bohlander that is completely revised and regularly updated by Ute Reusch. http://www.gesetze-im-internet.de/englisch_stgb/index.html. Accessed 24 June 2020
63. These topics are very complex. As a starting point, see Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC OJ L 130/92 (2019); see also Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases OJ L 77/20 (1996); see also Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure OJ L 157/1 (2016). Cf, for instance, Wiebe, A.: Protection of industrial data – a new property right for the digital economy? GRUR Int. **65** (10), pp. 877, 879ff (2016)
64. European Commission: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: “Building a European Data Economy”. COM(2017) 9 final, p. 13. Cf, for instance, Zech, H.: Information as Property. JIPITEC **6**(3), pp. 192, 196f, paras 31ff (2015); Wiebe, A. (n 63) pp. 881ff. However, the focus of the debate has in the meantime shifted towards the question of data access, and the introduction of new data property rights has become rather unlikely; see Mischau, L. (n44) pp. 237f with further references

The Legal Frameworks of the Right to Request the Deletion of Personal Data in the EU, the U.S. and Japan and the Right to Be Forgotten: A Study Focusing on Search Businesses

Mika Nakashima(✉)

Faculty of Global Informatics, Chuo University, Tokyo, Japan
nakashima@tamacc.chuo-u.ac.jp

Abstract. The issue of the "right to be forgotten" presents a modern problem with regard to a person's right to request search engine providers for the deletion of search results generated by entry on his/her name. In recent years, legislation introducing the right to request the deletion of personal data has been taking place in the EU and the U.S. This paper reviews the legal frameworks with regard to the right to request the deletion of personal data in the EU, the U.S. and Japan and studies whether there is a right for a natural person to request the deletion of search results on him/her from search businesses (in other words, the "right to be forgotten") in each of these jurisdictions. In addition, the author examines the challenges of the Japanese legal system.

Keywords: GDPR · CCPA · APPI · Right of deletion · Right to be forgotten

1 Introduction

Through the development of information technology, it has become possible to easily reproduce, preserve, and spread digitized information. On the other hand, digitized information is not expected to fade into obscurity; hence, once information related to an individual's privacy is made open to the public on the internet, it may not only create serious damage at the time of publication but also be harmful, being preserved in the internet space, for many years or possibly for good—a problem of "digital tattoo," so to speak [34].

The issue of the "right to be forgotten" presents a modern problem in that it is centered around a person's right to request search engine providers for the deletion of search results generated by entry on his/her name. This is a novel issue, occurring as a result of the dramatic rise in access to information on the internet because of the spread of search services. Even if information exists on the internet, it would be highly difficult if not impossible practically to access such information without the assistance of a search engine.

In recent years, legislation introducing a person's right to request the deletion of personal data has been taking place in the EU and the U.S. This paper reviews the legal frameworks in The EU, the U.S. and Japan with regard to the right to request the

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 29–40, 2020.
https://doi.org/10.1007/978-3-030-62803-1_3

deletion of data and studies whether there is a right for a natural person to request the deletion of search results based on his/her name searches from search businesses (in other words, the "right to be forgotten") in each of these jurisdictions. In addition, the author would like to outline the present situation and issues concerning the topic in Japan.

Please note that while the General Data Protection Regulation (hereinafter referred to as "GDPR") uses the term "erasure," the California Consumer Privacy Act of 2018 (hereinafter referred to as "CCPA") and the Act on the Protection of Personal Information (hereinafter referred to as "APPI") uses the term "deletion" to mean the erasure of data. This paper basically uses the term "deletion" like the CCPA and the APPI, but, where it considers the GDPR, conforms to its terminology. Similarly, while the GDPR uses the term "personal data," the CCPA and the APPI use the term "personal information" to mean the information related to an individual. The APPI also uses the term "personal data" when the information is recorded in a "personal information database, etc." so as to be searchable. This paper basically uses "personal information" like the CCPA and the APPI, but where it considers the GDPR, conforms to its terminology.

2 The Legal Framework of Each Jurisdiction

2.1 The EU (GDPR)

On April 27, 2016, the EU adopted the "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and the free movement of such data (General Data Protection Regulation)". The GDPR came into force in member countries on May 25, 2018. Article 7 of the Charter of Fundamental Rights of the European Union, which was signed in 2000, establishes the "respect for private life and family life," and Article 8 thereof establishes the "protection of personal data[1]." Based on the Charter, the GDPR, in its preface, states that it respects the protection of personal data as a fundamental right of natural persons. It should be noted that the GDPR superseded the pre-existing the "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data" (hereinafter referred to as "Directive").

The GDPR establishes the rights of data subjects and provides the "right to erasure (right to be forgotten)" in Article 17. Paragraph 1 of the Article states that the data subject shall have the right to request the controller the erasure of personal data where, e.g., the personal data is no longer necessary for the original purposes, the data subject

[1] For example, in the case of the European Court of Human Rights, Rotaru v. Romania, 4 May 2000, the applicant alleged a violation of his right to respect for his private life on account of the holding and use by the Romanian Intelligence Service of a file containing personal information and an infringement of his right of access to a court and his right to a remedy before a national authority that could rule on his application to have the file amended or destroyed. The ECHR concluded that both the storing of that information and the use of it, which were coupled with a refusal to allow the applicant an opportunity to refute it, amounted to interference with his right to respect for his private life as guaranteed by Article 8, Paragraph 1.

withdraws consent to or objects to the processing of personal data, or the personal data has been unlawfully processed. Paragraph 2 of the Article provides that where the controller has made the personal data public and is obliged to erase them under the Paragraph 1, the controller must take reasonable steps, including technical measures, to inform (other) controllers that are processing them that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data. Paragraph 3 of the Article also provides that Paragraphs 1 and 2 shall not apply to the extent that processing is necessary for exercising the right of freedom of expression and information, for compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller, for reasons of public interest in the area of public health, for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, for the establishment, exercise or defence of legal claims.

Considering that the European Court of Justice (hereinafter referred to as “ECJ”), in its preliminary ruling of May 13, 2014 (mentioned below), acknowledged that under the Directive, a request for deletion could be made for search results based on name searches, that Article 17, Paragraph 2 of the GDPR specifically prescribes the erasure *“of any links to*, or copy or replication of, those personal data” (italicized by the author), and that the Directive was the predecessor and basis of the GDPR, it may safely be concluded that the GDPR contemplates the cases where under Article 17 search engine providers shall be obliged on the request of the data subject to erase the search results [6, p. 99] [18, p. 156].

However, if we check the description of the GDPR’s preface on the “right to be forgotten”, we find that the main focus of the right to erasure is rather in cases where the data subject seeks the erasure of information posted on SNS that he/she used in his/her childhood, and not explicitly search services are mentioned.

Article 82, Paragraph 1 states that when a person has suffered material or non-material damage as a result of an infringement, he/she shall have the right to receive compensation from the controller or processor for the damage suffered. The liability for an infringement will not stop there, however. The following Article 83, Paragraph 5 prescribes administrative fines as a direct sanction, whereas Article 24 of the Directive would entrust the member states what sanctions to adopt. The fines imposed may run up to €20 million or up to 4% of the total worldwide annual turnover of the preceding financial year, whichever is higher.

2.2 The State of California, the United States (CCPA)

In the U.S., federal law in relation to privacy and personal information protection is based on self-regulations in the private sector, and individual laws exist in specific fields such as finance, medical treatment, and communication. However, there is no encompassing Data Protection Law at the federal level[2].

[2] On the other hand, the Federal Trade Commission (FTC) of the U.S. has operating authority based on the Federal Trade Commission Act of 1914 from the position of consumer protection [11, p. 408].

The State of California enacted "California Consumer Privacy Act of 2018," and it went into effect on January 1, 2020. CCPA is the first comprehensive Personal Information Protection Law in the U.S. While the State of California is a global leader in the development of new information technologies and related industries, the California Constitution guarantees the right of privacy, and the State has enacted several privacy-related laws as concrete endeavors to protect privacy (for example, the "Online Erasure Law," which was enacted in 2013, grants minors the right of deletion of posts made by themselves on SNS and related platforms). The CCPA is also one of these concrete endeavors. It is incorporated to Part 4 of Division 3 of the Civil Code as "Title 1.81.5 California Consumer Privacy Act of 2018 [1798.100—1798.199]". Global companies must comply with the CCPA. While they have already been preoccupied with responding to the GDPR, they are pressed to respond to the differences between the GDPR and the CCPA.

The CCPA is thought to have been influenced by the preceding GDPR; there are several similarities in the two acts. For example, the Section 798.100(d) of the CCPA which defines the business's obligation to consumers to disclose and provide personal information establishes "if provided electronically, the information shall be in a portable and, to the extent technically feasible, readily useable format that allows the consumer to transmit this information to another entity". This is similar to the right to data portability of Article 20, Paragraph 1 of the GDPR which establishes "the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided". Section 1798.105 states that "(a) The section provides that a consumer shall have the right to request that a business delete any personal information about the consumer which the business has collected from the consumer," and, "(c) a business that receives a verifiable consumer request from a consumer to delete the consumer's personal information pursuant to subdivision (a) of this section shall delete the consumer's personal information from its records and direct any 'service providers' to delete the consumer's personal information from their records." However, it is not clear as to the kind of situations this section anticipates, and it is also not clear whether deletions must be made by search businesses.

The CCPA provides statutory damages in the case of lawsuits brought by individual consumers and class action lawsuits (Section 1798.150). Furthermore, the CCPA establishes a civil penalty system enforced by the California Attorney General (Section 1798.155). Statutory damages are available, and in those cases, either damages of between $100 and $750 are available for each consumer per incident of infringement, or the consumer may recover actual damages, whichever amount is larger. Hence, it is possible for the consumer to institute a civil action (however, the framework of the statutory damages system seems to fundamentally assume cases of data breach). Regarding the civil penalty, violators may be liable for a civil penalty of not more than $2,500 for each violation or $7,500 for each intentional violation.

2.3 Japan (Civil Code and APPI)

The legal framework of personal information protection in Japan is based on the right of privacy and the APPI.

The right to privacy has been formulated by case law over the past half century under Article 709 of the Civil Code. The Article states "A person who has intentionally or negligently violated any right of others… shall be liable to compensate any damages resulting in consequence" (the omitted part indicated with "…" is a phrase inserted in a recent amendment of the Code). The right of privacy came to be counted as a "right" protected under the Article. In 1964 the Tokyo District Court explicated the right of privacy as the legal basis of its decision in the case of the novelist Yukio Mishima's roman-à-clef, *After the Banquet* [1]. In 1994 the Supreme Court acknowledged the "legal interest as not to have facts related to criminal records, etc. made public" in the case of a non-fiction book, *Reversal* [21]. Since then, more Supreme Court cases have followed, some with explicitly mentioning "privacy," and it can now be said that the right of privacy is part of law in Japan.

Although Article 709 the Civil Code only prescribes monetary compensation as its remedy, an order of injunction has now been established in lieu of or in addition to damages through case law. In a case of 2002, the Supreme Court upheld the decision of the High Court which ordered an injunction against publication of a roman-à-clef, which is regarded as the precedent for the availability of injunction based on the right of privacy[3].

The APPI was promulgated on May 30, 2003, and enforced on April 1, 2005. It is an administrative law that establishes rules related to the proper handling of personal information. On the basis of the Act, the Personal Information Protection Commission (hereinafter referred to as "PPC"), was established as an administrative organization, independently enforcing its authority. The APPI fulfills preventive functions against the improper handlings of a person's personal information but it does not directly provide him/her private remedies [9, p. 30] [30, p. 42].

A decade passed and, given the need to prepare for an environment wherein proper utilization and application of big data—including personal data—are to be made more readily possible and in view of responses to the globalization of business activities, the amendment to the Act was promulgated on September 9, 2015, and enforced on May 30, 2017 [28]. Furthermore, in accordance with the "revision in every 3 years" provision of the amended law (Article 12 of the supplementary provisions), the Act for Partial Amendment of the APPI (Act No. 44 of 2020) was enacted on June 5, 2020. It was passed and promulgated on June 12 of the same year. The following is based on the latest amended APPI, 2020 (not yet in force at the time of writing).

The APPI establishes the data subject's rights to request with regard to the disclosure of information about him/herself (Article 28) and the correction, etc. (Article 29) and the utilization cease, etc. (Article 30) of the retained personal data. A framework that allows the involvement of the data subject in certain cases is established. In relation to the deletion of personal data, apart from Article 19, which requires the personal information handling business operator to make reasonable efforts to delete the personal data that has

[3] "*Fish that swim in rocks*" Case, Decision of the Supreme Court on September 24, 2002, Hanreijiho, No. 1802, p. 60.

become unnecessary, Article 29 establishes the right to request a deletion. It provides that the principal may, when the contents of retained personal data that can identify the principal are not true to the facts, request the business operator to make a correction, addition or *deletion* (hereinafter referred to as "correction, etc.") in regard to the contents of the retained personal data. With this regulation, however, it should be noted that even though the word "deletion" is used here, its nature is not that of the right to request the complete erasure of data by the business but is no more than being a method of "correction." [9, pp. 312–316] [10, pp. 238–244] [14, pp. 214–218].

Article 30, Paragraph 1 stipulates that a principal may, when the retained personal data that can identify the principal is being handled in violation of the provisions of Article 16 or has been acquired or used in violation of the provisions of Articles 17 and 16-2, demand of a personal information handling business operator a utilization cease or deletion of the retained personal data. In addition, a new clause has been added to the Act to prescribe that a principal may do the same where there is a possibility that his/her rights or legitimate interests are harmed (Article 30, Paragraph 5). Prior to the Amendment, utilization cease, etc. was a remedy limited to cases where personal information was used for purposes other than those for which it was intended or acquired inappropriately (inappropriate use has been added in the Amendment 2020) [13], but it is extended to cases where there is a possibility that his/her rights or legitimate interests are harmed, which is noteworthy as it eases the requirements for requesting the erasure of personal data [33].

The "personal information handling business operator" defined in Article 2, Paragraph 5 of the Act are those who use a "personal information database etc." for business. Thus whether or not a search businesses is a personal information handling business operator as defined, depends on the meaning of Paragraph 4 of the same Article which states the definition of a "personal information database etc." Looking into the discussion of the Bill for the Act in 2003, the opinion of the government introducing the Bill was such that the databases of search businesses would not fall under "personal information database etc." [31, 32] Presumably it is what has been accepted in academia [10, pp. 79–80] [14, p. 72]. The main reasons for it are that the databases of search businesses are mixed with information other than personal information, that searching for information other than personal information, such as place names, is possible, and that no attached index is available as personal information. Opposing views have also been asserted, with their reasoning being the following: that even when information other than personal information can be searched for, it will not be an obstruction for it to fall under "those systematically organized so as to be able to search for particular personal information using a computer" of Item 1 of Paragraph 4 of Article 2, that a keyword search performed on a specific person's name on the search service is based on the index that the search business has created, and that an expansive use of search services in order to get a person's personal information is actually made. It may also be noted that at the time the Bill was introduced in 2003, search services were not as widespread as they are today [9, pp. 312–316][4].

[4] The search service of Google Inc., "Google," appeared at around 1998, and its Japanese version appeared at around 2000.

The APPI establishes criminal penalties. Imprisonment for not more than six months or a fine of not more than ¥ 300,000 is the penalty when the orders under the current APPI related to utilization cease, etc. have been infringed (Article 84 of the current Act). The amended Act, 2020 raises the statutory penalties for violations of the PPC's orders, false reports to the PPC, and other offenses. For violations of the PPC's order, the penalty is increased from the said above to imprisonment for not more than one year or a fine of not more than ¥ 1 million" (Article 83). As for false reports, the penalty is raised from a fine of not more than ¥ 300,000" to "a fine of not more than ¥ 500,000" (Article 85). As for the illegal provision of databases and the violations of the PPC's orders, the violating corporation (or natural person running the business) may be punished with a fine. In case of violations of the PPC's orders the maximum fine for a corporation is increased hugely to not more than ¥ 1 billion (from ¥ 300,000 under the current Act), taking into account the disparity in financial resources between a corporation and an actor (Article 87).

3 Judicial Precedents in Each Jurisdiction

3.1 The EU - the Preliminary Ruling of the ECJ in the González Case of May 13, 2014

On May 13, 2014 the ECJ delivered the preliminary ruling in Google Spain v. González, which is now credited as the first judicial precedent acknowledging a person's "right to be forgotten" as meaning the right to request the deletion of search results on his/her name searches. The ECJ held in summary as follows:

> The operator is, in certain circumstances, obliged to remove links to web pages that are published by third parties and contain information relating to a person from the list of results displayed following a search made on the basis of that person's name. When the data subject requests that links to web pages be removed from such a list of results on the grounds that he wishes the information appearing on those pages relating to him personally to be 'forgotten' after a certain time, and if it is found that the inclusion of those links in the list is, at this point in time, incompatible with the Directive, the links and information in the list of results must be erased.
> In this case, if one were to consider the sensitivity of information as far as the private life of the data subject was concerned and the fact that that information was first made public 16 years ago, then the information in question should no longer be linked to the name of the data subject through that list of search results, and it is allowed for the data subject to make such a request directly to the search business [5, 15].

The ruling was delivered in response to a request from Audiencia Nacional Spain concerning the interpretations of Articles of the Directive. Although the Directive was repealed and the GDPR superseded it, the latter provides "Right to erasure ('right to be forgotten')" as Article 17. Because it can be said that the data subject's right to request the deletion of search results based on his/her name searches is, practically speaking, the core concept of "right to be forgotten", the González case will be taken to be a leading precedent pertinent to interpreting Articles of the GDPR.

3.2 The U.S.

Presently, no direct judicial precedents in the U.S. acknowledge the "right to be forgotten" as a right to request deletion from search services. In fact, the discussions in the U.S. are centered around the issue of whether to limit the liability of search businesses as providers [19].[5] Article 230 (c) (1) of the Communications Decency Act provides immunity from tortious liability for providers and users of an interactive computer service who publish information provided by third-party users. Reportedly the courts have in years been stretching the meaning of "interactive computer service" to immunize web hosts, websites, search engines, and content creators although no cases of search engines are cited therewith [22, p. 371].

3.3 Japan - the Supreme Court Ruling on January 31, 2017

In and around when the EUJ ruling in González case was reported, applications for provisional injunction or actions on the merits, seeking the deletion of search results, began to be brought before the lower courts. Some of those cases were reported by news media with referring to the EUJ's ruling. One of those application cases went up in judicial ladder of appeals and, on January 31, 2017, The Supreme Court delivered its first-ever judgment on the topic matter. The Court decided, in summary, as follows: [3, 16].

> The interests of not having facts of one's privacy made public unless one gives permission" should be protected by law as this Court has held repeatedly. As search programs are created in such a way that search engines collect information on the internet which aligns itself with their programing policies, providing search results generated by the search engine is an "act of expression" by the search engine provider itself. Also as the provision of search results generated by search engine fulfills a major role as the foundation of information distribution on the internet in modern society, facilitating the public to publish information on the internet or get information they need from the vast quantities of information on the internet, to hold a search engine provider liable for providing a certain search result and obliged to delete it would constitute a restriction of this role, as well as a restriction of the act of expression.
>
> In the light of the nature and the functions of the search engine providers said above, whether a provider should be made liable for providing search results, in response to a search request on one person, with URLs to the articles on the internet which contain facts of that person's privacy should be decided on balancing of interests, taking into consideration the circumstances for the legal interest of not having the said facts made public, such as the nature and content of the said facts, the range of transmission enabled and the level of actual damage that person has suffered by way of the said URLs provided, the social status and influence of that person, the content and purpose of the said article, the social situations of the time and afterward it was published on the internet, and the necessity of publishing the said facts in it, and also the circumstances in respect of reasons for providing the URLs in the search results. One may request the search engine provider to delete the said URLs from the search result if on balancing the legal interest of not having the said facts made public is *clearly* superior to the other interest (italicized by the author).

[5] See [8, 23, 26] for details on the state of debate in the U.S.

> In this case, the fact that the appellant was arrested for child prostitution is a fact of privacy that he would not want to be known by others without permission, but in light of child prostitution being a subject of strong social reproach and banned with penalties, it still remains a matter related to the interests of the public. Even upon considering the circumstances, such as the appellant not having committed a crime for a certain period of time since then, it cannot be said that the legal interest of not having the facts of his arrest made public is superior. The ruling of the High Court (the appellate court) to turn down the appellant's application is correct.

In the first instance of this case the court referred to "the right to be forgotten" when it approved the order of deleting search result. The High Court (in the appellate instance) repudiated the introduction of the concept, annulling the order. Expectations were raised that the Supreme Court would possibly give some words of approval or disapproval on it but nothing was referred to. It may be safe for now, therefore, to say that the concept of "the right to be forgotten" has not been settled judicially in Japan.

4 The Present Situation of the Legal Framework in Japan and the Future

In relation to the "right to be forgotten" as a right to request deletion from search services, assessing the present situation of the legal system in Japan leads to the following.

Whereas the right of privacy and the injunction order as a remedy for its infringement have well been established through case law under the Article 709 of the Civil Code, the right to be forgotten has yet to be acknowledged judicially.

Article 30, Section 5 of the amended APPI, 2020 provides that "a person may request the utilization cease or deletion of the retained personal data where there is a possibility that his/her rights or legitimate interests are harmed". If a search business operator is to be construed as a "personal information handling business operator" under the Act (Article 2, Paragraph 5), this newly established provision may serve as a ground for requesting the deletion of search results when the amended Act comes into force.

Since the government opinion at the time of the introduction of the Bill, as shown above (p. 6), it has been a common understanding that the APPI does not assume the databases of search businesses to be the "personal information database etc." and, accordingly, that the search businesses do not fall under the "personal information handling business operator." Considering the rapid prevalence of search services since then,[6] however, we face an untraversed situation in which information of a diverse nature regarding any person, whether most famous or unknown, can be searched by

[6] On July 27, 2010, Yahoo Japan Corporation gathered attention when it used the search engine of Google Inc., as the back engine of "Yahoo! JAPAN." See [25] for the market share of search engines in Japan.

merely making a name search of the individual on the internet. There is room for Japan to reconsider the way in which she think about the legal nature of search businesses[7].

The APPI does not only establish the obligations of businesses as of administrative level but also acknowledges the rights of a data subject to request utilization cease, etc. in certain cases. However, it merely acknowledges the involvement of data subjects within the prescribed aims. The GDPR, grounded on fundamental human rights, gives strong "personal data protection" and guarantees the "right to erasure (the right to be forgotten)," whereas the CCPA, based on the "right of privacy" of the Constitution, guarantees a consumer the "right to request that a business delete any personal information about the consumer which the business has collected from the consumer". In contrast with them, Japan's APPI does not directly establish the rights of the individual and an actor committing an infringement of an obligation under the Act is punished with imprisonment and/or fines (a committing corporation also suffers the latter sanction in certain cases) but is not rendered liable to an aggrieved person for damages, unlike the GDPR and the CCPA, which provide for a data subject's or consumer's right to claim damages. Admitting in the APPI the right to request deletion of search results (whether through further amendments to the Act or through interpretation of the Act) would be incompatible with the legislative intent of the Act. The legal framework of the right to request deletion has no choice but to depend on the right of privacy as the right to request under private law; this is the situation at present.

Regarding whether the right to request deletion from search businesses is acknowledged as the right of privacy, according to the aforementioned Supreme Court decision, the possibility of acknowledging the request of deletion is left open in the case where search results should "clearly" be deleted upon comparison with the "act of expression" of the search business (n.b., which reasoning is an obiter dictum because the application for an order of deletion in the case was not approved). In relation to the requirement of "clear"-ness, it is hoped that further judicial cases will help clarify its meaning [8] as well as academic discussion should be made on it, in which the peculiar characteristics of search businesses—such as only providing a list of search results, being unable to know details with regard to the information on the original website that

[7] On December 13, 2019, the PPC made public an outline of system amendment in reviewing the APPI every 3 years [33]. This outline holds up the easing of the requirements of "deletion," but it is unclear whether debates have been carried out in relation to the right of deletion and the "right to be forgotten" that has search businesses as the subject. Furthermore, according to the written report gathered together by workshop under the Ministry of Internal Affairs and Communications after the first decision of the ECJ mentioned above was given, the issue of the deletion of search results carried out by search businesses has the premise of fundamentally entrusting to the self-regulation of the businesses and carrying out inspections within the legal framework related to the existing the right of privacy [17].

[8] On December 12, 2019, in a case requesting deletion of search results (decision on the merits), where a man requested the deletion of search results from Google LLC, the Sapporo District Court stated that the interests of not making it public are superior to that of maintaining of the display and gave a decision of ordering deletion (Westlaw. JAPAN, reference number: 2019WLJPCA 12126001).

Furthermore, for the tendencies inside and outside the country before the aforementioned decision of the Supreme Court, see [12, 35]. The summary of domestic developments since the Supreme Court decision, see [24].

the links displayed in the search results lead to (they are not in a position to determine the veracity of the information published on the original website), and being able to actively continue to display a person's past privacy information through search results based on his/her names—should be considered to determine when and how the new media of search business should be obliged to delete a person's privacy information on his/her request.

References

1. After the Banquet Case, Decision of the Tokyo District Court on September 28, 1964, Hanreijiho, No. 385, p. 12. (1964)
2. California Consumer Privacy Act of 2018, 1.81.5. CIVIL CODE §§1798.100 - 1798.199 (2018). http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article=
3. Deletion of Posted Articles Case: Decision of the Supreme Court on January 31, 2017, 2016 (Appeal by Permission Case) No. 45, Order of Provisional Injunction of Deletion of Posted Articles and Appeal Rejected, Minshu, 71(1), p. 63; Hanreijiho, No. 2328, p. 10; Hanrei Times, No. 1434, p. 48 (2017)
4. DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (1995). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en
5. González Case: 131/12, Google Spain SL, v. Agencia Española de Protección de Datos and Mario Costeja González, Google Inc. (2014). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62012CJ0131&form=EN
6. Miyashita, H.: The General Data Protection Regulation, Keiso Shobo (2018)
7. Miyashita, H.: The Restoration of The Right of Privacy – The Clash of Freedom and Dignity, Chuo University Press (2015)
8. Miyashita, H.: The right to be forgotten and the legal liability of search engines. Comp. Law J. **50**(1), 35 (2016)
9. Okamura, H.: The Act on the Protection of Personal Information. 3rd Edn. Shojihomu (2017)
10. Sonobe, I., Fujiwara, S. (eds.): Explanation of the Personal Information Protection Law <Second Revised Edition>, Gyosei (2018)
11. Ishii, K.: New Edition: The Present and Future of the Personal Information Protection Law – Global Trends and the Future Image of Japan, Keiso Shobo (2017)
12. Ishii, K.: Supreme court decision in Google search results removal request case, No. 2353, p. 148. Hanreijiho (2018)
13. Ishii, K.: Legislation of the so-called right to be forgotten': outline of interim arrangements for revisions to the personal information protection law. Bus. Law **19**(8), 82 (2019)
14. Uga, K.: Article by Article Explanation of the Act on the Protection of Personal Information – Act on the Protection of Personal Information, Act on the Protection of Personal Information Held by Administrative Organs, Act on the Protection of Personal Information Retained by Independent Administrative Institutions 6th Edition, Yuhikaku (2018)
15. Nakashima, M.: Search Services and the Right to be Forgotten – On the Preliminary Ruling of the European Court of Justice in the Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González Incident (May 13, 2014), Legal Practices of Information Networks, Daiichihoki (2016)

16. Nakashima, M.: The Deletion of Search Results and the Right to be Forgotten: Regarding the State of the Debate of Theories, Starting from the Supreme Court decision of January 31, 2017, Tokai Law Review, No. 56, p. 117 (2019). https://www.u-tokai.ac.jp/academics/undergraduate/law/kiyou/pdf/2019_56/07.pdf
17. Ministry of Internal Affairs and Communications: Regarding Responses Towards the Distribution of Personal Information, User Information, Etc., on the internet: Written Report by ICT Service Safety and Security Research Society (2015). http://www.soumu.go.jp/main_content/000369245.pdf
18. Voigt, P., von dem Bussche, A.: The EU General Data Protection Regulation (GDPR): A Practical Guide. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57959-7
19. Parker v. Google, Inc.: 422 F. Supp. 2D 492 (2006)
20. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504
21. "Reversal" Case, Decision of the Supreme Court on February 8, 1994, Minshu, 48(2), p. 149 (1994)
22. Rustad, M.L., Koenig, T.H.: Rebooting Cybertort Law. Wash. Law Rev. **80**, 355 (2005)
23. Narihara, S.: The state of debate in Japan, the U.S. and the EU surrounding the "right to be forgotten". Admin. Inform. Syst. **51**(6), 54 (2015)
24. Narihara, S.: Freedom of expression and moral rights regarding search engines - a review of the supreme court decision in 2017 and case studies on the deletion of search results since that decision. J. Law Inf. Syst. (7), 47 (2020)
25. StatCounter GlobalStats, Search Engine Market Share Japan, January 2009–November 2019. https://gs.statcounter.com/search-engine-market-share/all/japan/#monthly-200901-201911
26. Komukai, T.: The Right to be Forgotten and the U.S. Communications Decency Act, Information Processing Society of Japan Research Report vol. 2015-EIP-69 No. 15 (2015). https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=144945&file_id=1&file_no=1
27. Komukai, T., Ishii, K.: Outline: The GDPR, NTT Publishing (2019)
28. The amended Act on the Protection of Personal Information. Enforced May 30, 2017. https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.pdf
29. The Communications Decency Act (CDA), 47 U.S.C. Article 230 (1996)
30. The General Affairs Agency Administrative Management Bureau (Supervision): Article by Article Explanation of the Act on the Protection of Personal Information, Newly-Revised Edition, Daiichihoki (1991)
31. The House of Councillors: Special Committee Related to the Protection of Personal Information, Record of Proceedings on May 13, 2003 (No. 3), Answers by Akio Fujii as Government Witness (2003). http://kokkai.ndl.go.jp/SENTAKU/sangiin/156/0071/15605130071003.pdf
32. The House of Representatives: Special Committee Related to the Protection of Personal Information, Record of Proceedings on April 18, 2003 (No. 6), Answers by Minister of State, Hiroyuki Hosoda (2003). http://kokkai.ndl.go.jp/SENTAKU/syugiin/156/0017/15604180017006.pdf
33. The Personal Information Protection Commission: The Act on the Protection of Personal Information: Revision in Every 3 Years – Outline of Framework Amendment (2019). https://www.ppc.go.jp/files/pdf/seidokaiseitaiko.pdf
34. Mayer-Schönberger, V.: Delete: The Virtue of Forgetting in the Digital Age, Princeton University Press, Revised (2011)
35. Okuda, Y. (ed.): Internet Society and the Right to be Forgotten – Court Cases of Deleting Personal Data and their Legal Principles, Gendaijinbunsha (2015)

Examining the Possibility of Expanding the Use of Digital Images Recorded in the "Audio-Visual Recording of Custodial Interrogation" System

Mariko Nakamura(✉)

Chuo University, Tokyo, Japan
m_nkmr@tamacc.chuo-u.ac.jp

Abstract. Japan has recently introduced the "Audio-Visual Recording of Custodial Interrogation" system to verify the voluntariness of confessions or other disadvantageous statements of the accused that are contained in their depositions. Although digital images recorded in the new system are supposed to be used as supplementary evidence, it appears that they can be used to directly prove the truth of the matter asserted, which is the question to be left open. This article considers whether the digital images can be used to prove guilt or innocence, and if so, whether they need to meet more strict requirements than those for depositions. First, the article discusses the legal grounds pertaining to the application of a conventional hearsay exception to the digital images. Second, it explains how the use of the digital images, which could be concerning, can meet the legal demands of the adversary system underlying the Hearsay Rule. The article concludes that these recordings may be used to prove guilt or innocence if the accused or defense counsels have the meaningful opportunity for impeachment.

Keywords: Admissibility · Evidence · Digital data · Interrogation · Hearsay

1 Introduction

The Japanese criminal justice system made a significant shift in 2016 with "Act to Amend Parts of the Code of Criminal Procedure and Other Acts (Act No. 54 of 2016)". The core revision of this act was the introduction of the "Audio-Visual Recording of Custodial Interrogation" system, which was officially implemented in 2019. In a mail fraud case widely covered by the media as a false accusation, the Special Investigation Department of the Osaka District Public Prosecutors Office allegedly engaged in suggestive interrogations and evidence tampering, which substantially led to the amendment that aimed to reduce the excessive dependence on interrogations and depositions.

Digital images recorded in the system are supposed to be used as supplementary evidence to prove whether custodial suspects in certain cases have voluntarily made confessions or other disadvantageous statements included in depositions, upon objections from the accused or defense counsels when public prosecutors make requests for

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 41–51, 2020.
https://doi.org/10.1007/978-3-030-62803-1_4

examinations of the depositions. This is based on the premise that police officers or public prosecutors have recorded interrogations of custodial suspects and retained those images.

It can be expected that public prosecutors or the accused or defense counsels will eventually make requests for conducting examinations of the digital images themselves to directly prove whether the accused have committed the charged crimes. In this situation, whether such images are admissible under the same requirements as depositions of the accused may become an issue, which is left open. Because digital images constitute "electronic or magnetic records" under Article 7-2 of the Penal Code, they should be carefully authenticated to ensure that they have been recorded without falsification or other manipulations. Although the accused are presumed to be innocent during trials, it is possible that digital images of the accused being interrogated or confessing may give fact-finders, especially Saiban-in (lay judges), the impression that the accused are guilty before such guilt is established.

Part 2 of this article explains the "Audio-Visual Recording of Custodial Interrogation" system and the reasons for its introduction. Part 3 reviews certain courts' reluctance to use digital images of custodial interrogations as evidence for the truth beyond the prescribed supplementary use. Part 4 considers whether digital images recorded in the new system can be used to prove guilt or innocence, and if so, whether they need to meet more strict requirements than those for depositions. The article concludes that the recordings may be used to prove guilt or innocence if the accused or defense counsels have the meaningful opportunity for impeachment.

2 Audio-Visual Recording of Custodial Interrogation

This section describes the types of cases for which the "Audio-Visual Recording of Custodial Interrogation" system is used and how it affects the notion of voluntariness. Article 301–2[1] was added to the Code of Criminal Procedure by Act No. 54 of 2016 as

[1] Article 301–2 provides as below (The author translated this new provision with reference to [1], which does not currently include the new provision. Translations of Japanese codes in this article also refer to [1]).

"(1) With regard to the cases prescribed in the following items, when a public prosecutor requests examination of a written statement made during an interrogation about the case pursuant to the provision of paragraph (1) of Article 198 (limited to the interrogation of an arrested or detained suspect; the same shall apply hereinafter in paragraph (3)) or an opportunity for explanation pursuant to the provision of paragraph (1) of Article 203, paragraph (1) of Article 204 or paragraph (1) of Article 205 (including cases to which these provisions apply *mutatis mutandis* pursuant to the provision of Article 211 and Article 216; the same shall apply hereinafter in paragraph (3)) that may be admissible pursuant to the provision of paragraph (1) of article 322 and contains an admission of a disadvantageous fact, and the accused or defense counsel raises an objection to the request of examination on the grounds that there is a doubt about the admission being voluntary, the public prosecutor shall request examination of the recording medium on which the statement of the accused and the circumstances during the interrogation or opportunity for explanation are recorded from beginning to end in accordance with the provision of paragraph (4) of this Article to prove that the admission has been made voluntarily; provided, however, that this shall not apply when there is no such recording medium because of the fact that the statements of the accused and the a provision related to the examination of evidence. It requires (1) police officers or public prosecutors to perform the audio-visual recording of custodial interrogations and (2) public prosecutors to present the digital images to prove the voluntariness of

(Footnote 1 continued)
circumstances have not been recorded in accordance with the provision of paragraph (4) due to any of the items of the paragraph, or other unavoidable reasons.

(i) cases involving offenses punishable by death penalty or life imprisonment with or without work;

(ii) cases involving offenses that are punishable by imprisonment with or without work for a minimum period not less than one year and that have caused a victim to die by intentional criminal acts; and

(iii) cases other than cases that a judicial police officer has sent to a public prosecutor (excluding those falling under the preceding two items)

(2) When a public prosecutor does not request examination of the recording medium prescribed in the preceding paragraph in violation of the preceding paragraph, a court shall rule to dismiss the request of examination of the written statement prescribed in the preceding paragraph.

(3) The preceding two paragraphs shall apply when, with regard to the cases prescribed in any of the items of the paragraph (1) of this Article, the accused or defense counsel raises an objection to the use of the statement made by a person other than the accused that may be admissible pursuant to the provision of paragraph (1) of Article 322 applied *mutatis mutandis* by the provision of paragraph (1) of Article 324 and that contains the statement of the accused (limited to the statement containing an admission of a disadvantageous fact) made during an interrogation about the case pursuant to the provision of paragraph (1) of Article 198 or an opportunity for explanation pursuant to the provision of paragraph (1) of Article 203, paragraph (1) of Article 204 or paragraph (1) of Article 205 on the grounds that there is doubt about the admission being voluntary.

(4) When a public prosecutor or public prosecutor's assistant officer, with regard to cases prescribed in any of the items of the paragraph (1) of this Article (excluding those falling the item (iii) of the paragraph whose related cases have been sent to a public prosecutor and that are expected to be sent to a public prosecutor by a judicial police officer due to the fact that a judicial police officer is investigating the cases or other circumstances), interrogates an arrested or detained suspect pursuant to the provision of paragraph (1) of Article 198 or gives a suspect an opportunity for explanation pursuant to the provision of paragraph (1) of Article 204 or paragraph (1) of Article 205 (including cases to which these provisions apply *mutatis mutandis* pursuant to the provision of Article 211 and Article 216), the public prosecutor or public prosecutor's assistant officer shall, except when any of the following items applies, record the statement of the suspect and the circumstances on the recording medium by means of the audio-visual recording. The same shall apply when a judicial police official, with regard to cases prescribed in the item (i) or (ii) of the paragraph (1) of this Article, interrogates an arrested or detained suspect pursuant to the provision of paragraph (1) of Article 198 or gives a suspect an opportunity for explanation pursuant to the provision of paragraph (1) of Article 203 (including cases to which these provisions apply *mutatis mutandis* pursuant to the provision of Article 211 and Article 216).

(i) when the malfunction of equipment necessary for recording or other unavoidable reasons prevent the recording;

(ii) when deeming that, judging from his/her refusal of the recording or other behaviors, the suspect will be unable to make enough statements if recorded;

(iii) when deeming that the case involves an offense that has been committed by a member of an organized crime group that has been designated by the Prefectural Pubic Safety Commission in accordance with the provision of Article 3 of the Act on Prevention of Unjust Acts by Organized Crime Groups; and

(iv) in addition to those provided in the preceding two items, when deeming that, judging from that there is a risk of physical or property harm, or threat or confusion of the suspect or his/her relatives through disclosing of his/her statements and the circumstances according to the nature of the crime, the behaviors of those concerned in the case, the character of a group to which the suspect belongs or other circumstances, the suspect will be unable to make enough statements if recorded".

confessions or other disadvantageous statements in depositions. Because suspects in custody practically assume that they have the legal obligation to undergo interrogations, the rate of objections from them against the voluntariness is much higher than from suspects not in custody.[2]

The following cases are subject to this system: (1) ones subject to the Saiban-in system, the Japanese version of the jury system, which is applicable for certain serious crimes such as Homicide, Robbery Causing Death or Injury, Arson of Inhabited Buildings, Kidnapping for Ransom and Dangerous Driving Causing Death; and (2) ones that public prosecutors exceptionally investigate independently of police officers such as corruption cases or corporate crimes, which require more sophisticated knowledge. Both types exhibit relatively higher rate of objections against the voluntariness; and the former is related to serious crimes as noted above and the latter lacks opportunities for separate scrutiny by two independent institutions, namely, the police and public prosecution.

Depositions of the accused are admissible as an exception to the Hearsay Rule[3] when the accused have made confessions or other disadvantageous statements voluntarily, or other statements under special circumstances with credibility. In relation to the former statements, when the accused or defense counsels raise objections against the voluntariness, public prosecutors who have requested for the examination of depositions are required to also request for the examination of digital images of the given custodial interrogations in principle. If public prosecutors do not present those images without any reasonable ground, the courts shall dismiss the requests for examination of the depositions, which is premised on the fact that police officers or public prosecutors have the legal obligation to perform the audio-visual recording of custodial interrogations.

3 Courts' Reluctance to Use Digital Images of Custodial Interrogations as Evidence for the Truth

Because police officers or public prosecutors have the legal obligation to perform the audio-visual recording of custodial interrogations in certain cases, it is natural for public prosecutors or the accused or defense counsels to want to use the digital images as evidence to prove guilt or innocence.

However, some courts have expressed concerns about using digital images recorded during custodial interrogations to prove the truth of the matter asserted, while others have admitted them into evidence. In 2016, the Tokyo High Court affirmed the court below to have dismissed a request by a public prosecutor to examine digital

[2] When there is a doubt that given confessions or other disadvantageous statements have been made voluntarily, they may not be admitted into evidence.

[3] No "document" nor "statement of another person made on a day other than the trial date" that is produced to prove the truth of the matter asserted may be admitted into evidence in principle primarily because an out-of-court statement contained in said document or statement cannot be subjected to a cross-examination.

images for the truth that were recorded during the experimental period of the system, taking a cautious stance as follows.[4]

> "If generally allowing the recording medium into evidence for the truth, beyond its use prescribed by the Act, because it makes it possible to judge the credibility of statements through the demeanor of a suspect as well as contents of the statements, aside from the difficulty or risk to judge the credibility through the demeanor during interrogations, there is not only a fear that, given the system or practice of interrogations in Japan, a trial would be a procedure to watch the recording medium of prolonged interrogations under the control of investigators and review the appropriateness, but also a doubt that such a procedure is found a proper trial, given that it significantly departs from the rule of fact-finding through evidences directly examined by courts and that it is not deemed to be independent of the investigation. Also, if a lot of time and energy is spent during a procedure for examining evidences to watch the recording medium to judge the credibility of statements through the demeanor of a suspect during interrogations, a trial might substantively and substantially lose balance compared to examinations for objective or otherwise critical evidences and would not meet the social requirement underlying the Act to break over-dependence on interrogations and written statements.
> Therefore, the admissibility of the recording medium of interrogations for the truth, or if allowing it into evidence, its condition, etc. should be discussed carefully in view of a proper trial".

Another judgment of the Tokyo High Court, whose chief justice was the same as the one who presided over the above-cited case, addressed a notorious murder case named "Imaichi-Jiken",[5] where the court below examined digital images of interrogations during the experimental period for about seven hours out of recorded 80 h or so and was reversed because it illegally found that the accused was the perpetrator directly from his demeanor in the digital images that were allowed into evidence only to prove the voluntariness or credibility of his confessions (although the Tokyo High Court found the accused guilty considering other circumstantial evidences and rendered the same punishment as the court below pronounced, now pending appeal). The relevant parts are quoted below.

> "There is a fine line between whether the accused can be found a murderer and whether statements of the accused admitting that he/she is a murderer can be considered trustworthy".

> "Even if the recording medium of interrogations reveals the objective circumstances of the interrogations, there is a strong doubt about trying to judge the credibility of confessions of a suspect through his/her demeanor during the interrogations replayed by the recording medium, despite its inability to reflect his/her mind audio-visually".

> "Although our system and practice of interrogations may produce false confessions as past experiences show, there will still be a risk that confessions during investigations are evaluated to have higher probative value than in reality. Therefore, in judging the credibility of confessions, needless to say that the statements must not be coerced, it is deemed to be important to weigh the verifiable factors heavily, such as with or without the revelation of a secret, the consistency with objective facts or other evidences, etc.; and consider multi-directionally along with reasonableness or naturalness, etc. of the contents and calmly at a proper distance from the confessions. However, evaluating the credibility of statements by watching the recording medium of interrogations… may result in intuitive judgments about confessions under the

[4] Case cited in [2].

[5] Case cited in [3].

> unique circumstances of custodial interrogations based on impressions influenced by subjective views of fact-finders in a way that is closely associated to the confessions in excess, such as whether the accused seems to tell the truth or not through watching his/her demeanor during the interrogations replayed audio-visually, and may have a negative impact by hindering the careful consideration explained above".

Although the latter precedent was concerned about using the digital images as supplementary evidence for the credibility of statements by the accused, its mention may be the case with using them to prove guilt or innocence.

4 Discussions by Scholars

4.1 The Hearsay Exception of Article 322 (1)

Although the main focus of Article 301–2 of the Code of Criminal Procedure is depositions of the accused, it also applies to statements by people other than the accused that contain the accused's out-of-court confessions or other disadvantageous statements. Some argue that digital images of custodial interrogations may not be used to prove guilt or innocence because the new provision assumes to use them only as supplementary evidence even for statements other than those by the accused, which appear to replay the accused's interrogations less faithfully than digital images.[6] However, most do not deny that they may be allowed into evidence as hearsay exceptions.

The Hearsay Rule applies to "document[s]" or "statement[s] of [others] made on a day other than the trial date". Article 322 (1), which is one of the exceptions to the Hearsay Rule, provides that "written statement[s] made by the accused" or "written statement[s] recording the statement[s] of the accused which ha[ve] the accused's signature[s] or seal[s] affixed by [the accused]" are admissible when the accused have made confessions or other disadvantageous statements voluntarily, or other statements under special circumstances with credibility. Although some argue that digital images of custodial interrogations do not literally constitute "written statement[s]",[7] equally electronically recorded audiotapes that include out-of-court statements have been conventionally addressed in analogy with "written statement[s]" on the premise that audiotapes constitute "hearsay" requiring scrutiny.

The remaining requirements for the hearsay exception are signatures or seals. Some argue that they are procedural guarantees for suspects to personally decide whether their own statements may be used as evidence in future trials, and digital images of custodial interrogations may not be used for the truth unless they have similar guarantees as alternatives to signatures or seals.[8] However, with regard to documents with pictures in which the accused in a molestation case explained and demonstrated what he thought had happened before police officers, the Supreme Court took them to have

[6] [4] pp.134-135.

[7] [5] p.167.

[8] [6] p.16, [7] pp.73-74, [8] pp.154-156, [9] pp.84-85, [10] p.54, [11] pp.336-337, [12] p.165, [13] pp.31-32.

been produced as “hearsay” to prove the truth of the matter asserted and demonstrated in the documents and their attached pictures.[9] Although it required them to meet the same conditions as usual depositions accordingly, it exempted parts of the pictures from signatures or seals because the recording process such as shooting and developing is implemented mechanically. Signatures or seals are supposed to be placed on usual depositions to assure the accuracy of the recording process. If digital images of custodial interrogations may be addressed similarly, they are not required to have signatures or seals either.

4.2 The Rule of Fact-Finding Through Evidences Directly Examined by Courts

The most concerning factor indicated in the precedents above is clearly the rule of fact-finding through evidences directly examined by courts. Regarding this point, some scholars counter that because playing digital images is equivalent to reading out documents, it is not against the rule in the sense that courts may personally examine digital images similar to how they examine documents.[10] Additionally, although the rule may be interpreted as requiring fact-finders to examine original evidences as well, digital images seem to be more original than depositions that are allowed into evidence as exceptions to the Hearsay Rule because they have mechanically recorded not only the contents of statements but also the circumstances before and during the statements and the demeanor of the speakers.[11] Even compared to statements in court, one argues that those in digital images are deemed to be original as well because both are instances of information produced by those who have personally experienced the facts.[12]

The rule of fact-finding through evidences directly examined by courts is typically explained in relation to the Hearsay Rule. However, the Hearsay Rule is originally based on the adversary system, which guarantees the opportunity to control fact-findings,[13] and it also applies to out-of-court statements of the accused for both parties to have this opportunity. Although the accused are generally present in court during trials, their statements during the investigation stage are nevertheless necessary, particularly when they have changed their attitudes into silence or denial about the crimes, and depositions of their former statements may not be used as evidence because they have not given signatures or seals, or depositions have not been created for some reasons.[14] When the accused (or at least their defense counsels) are in court, they can rebut the public prosecutors’ cases using their statements during the investigation stage by personally explaining to fact-finders why they have made the statements in the past. This seems to be the case regardless of whether the statements are contained in depositions or digital images.

[9] The case [14].

[10] [15] p.17, [16] p.17.

[11] [17] pp.7–8.

[12] [16] p.18.

[13] [18] p.426.

[14] [19] p.15, [20] p.66, [21] pp.111–112.

It is known that digital data generally have “black box dangers” in the sense that “a… machine conveyance… might be false or misleading because the machine is programmed to render false information…, is inarticulate, or has engaged in analytical missteps”.[15] However, because digital images of custodial interrogations are deemed to be simply “conduits” for the assertions of people, the assertions themselves have only to be “subject to… the usual safeguards that apply to human testimon[ies]”,[16] which include the Hearsay Rule.

Another noteworthy point about digital data is that they are “inherently mutable”.[17] Although digital data can be manipulated, “the admission of such evidence[s] may turn on authenticating whether… human declarant[s] [have] actually made the statement [s]”.[18] With regard to digital images of custodial interrogations, verification is required about whether their edited versions wrongfully provide fact-finders too much disadvantageous impressions against the accused on the precondition that the digital images themselves have not been falsified.

Public prosecutors must promptly disclose evidences that they have requested to be examined, thus providing the accused or defense counsels the opportunity to inspect “documentary evidence or articles of evidence” (Article 316-14 (i)). Although digital images of custodial interrogations do not literally constitute given evidences, they are also deemed to be subject to the pretrial disclosure.[19] Even when public prosecutors have not made requests for examinations of and disclose full versions, the accused or defense counsels may request the disclosure of “documents of recorded statements of the accused” (Article 316-15 (1) (vii)), which also include the “recording medium which is able to record images or sound and on which the [accused’s] statement[s] ha [ve] been recorded” (Article 316-14 (ii)).

The accused or defense counsels, on the other hand, must clearly indicate their opinions on whether they consent or have no objection relating to the public prosecutors’ requests of examination (Article 316-16 (1)). Additionally, they may request disclosure of evidences other than ones that public prosecutors have already disclosed in certain situations (Article 316-20 (1)). Thus, they have the opportunity before trials to watch digital images of custodial interrogations, and if necessary, to ask for making the given parts to be added to or deleted from the edited versions.

Although digital images might be affected by “lens, angle, speed, placement, cameraperson bias, or other variables”,[20] some of which are indicated as “Camera Perspective Bias”,[21] the important aspect of fact-finding is deemed to involve considering not only the demeanor of the accused when making the given statements but

[15] [22] pp.1977, 1989-2000.

[16] [22] pp.1977, 2002–2003.

[17] [23] pp.214-215, 217.

[18] [22] p.2002.

[19] They are practically addressed in the same manner ([24] pp.28-29).

[20] [22] p.2038.

[21] [25] pp.144-145.

also the whole circumstances of interrogations leading to the statements.[22] Although digital images appear to be better evidences to watch the accused's demeanor than depositions,[23] the accused or defense counsels may ask judges to instruct Saiban-in toward the appropriate manner of fact-finding or call expert witnesses of psychology in cases where Saiban-in, who are not supposed to be involved in the pretrial disclosure, might be at risk of being excessively reliant on the impressions from digital images.[24]

When the accused or defense counsels eventually have the appropriate and sufficient opportunity to defend,[25] implying the meaningful opportunity for impeachment through a set of guarantees explained above, the demand of the adversary system underlying the Hearsay Rule will be met.

5 Conclusion

This article has discussed the use of digital images recorded during custodial interrogations. They may be used to prove guilt or innocence extending beyond supplementary evidence for the voluntariness or credibility of the accused's statements if the accused or defense counsels have the meaningful opportunity for impeachment. When statements contained in digital images overlap with former statements of the accused made in response to questions in court, presiding judges shall deny public prosecutors' requests of examination for the digital images. Therefore, it must be carefully considered whether digital images themselves or the whole edited versions of which public prosecutors have made requests for examinations are substantially necessary to be examined,[26] i.e. whether they are deemed to be the best evidences.[27] Above all, we need to be aware that confessions themselves are supposed to be complementary evidences to avoid wrong judgments.[28]

Acknowledgment. This article was supported by JSPS KAKENHI Grant Number JP18H05656 and JP19K20860.

References

1. Ministry of Justice: Japanese Law Translation. http://www.japaneselawtranslation.go.jp/?re=02. Accessed 5 Feb 2020
2. Tokyo High Court, 8/10/2016, Law Times Report (1429) 132 (in Japanese)
3. Tokyo High Court, 8/3/2018, Law Times Report (1456) 75 (in Japanese)

[22] [21] pp.110–111, [26] p.18, [27] p.47.

[23] [15] p.19, [28] p.188.

[24] Examples of studies that approach the "Audio-Visual Recording of Custodial Interrogation" system from a psychological perspective are [29], [30] and [31].

[25] [32] p.6.

[26] [33] p.16, [34] p.57.

[27] [35] p.365.

[28] [36] pp.356-357.

4. Kyoji, M.: Torishirabe Rokuon Rokuga No Shiko Ni Muketa Bengo Katsudo No Tenbo—Un-yo Kakudai Ni Yoru Zenken Zenkatei Kiroku No Jitsugen Wo Mezashite, Osaka Bar Association Torishirabe No Kashika Osaka Honbu (ed): Kommentar Kashikaho, pp. 110–138, Gendaijinbun-sha, Tokyo, Japan (2017) (in Japanese)
5. Takahiro, N.: A commentary of [2]. Horitsu Jiho **89**(5), 164 (2017) (in Japanese)
6. Yushi, M.: Higisha Torishirabe No "Kashika"—Rokuga DVD No Shoko Riyo No Zehi. Horitsu Jiho **84**(9), 10 (2012) (in Japanese)
7. Mutsumi, I.: Torishirabe Kashika To Shokoho. Horitsu Jiho **85**(9), 69 (2013) (in Japanese)
8. Shota, A.: Electronic recording of interrogation and the admissibility of the recording media in evidence law. Aoyama Law Forum **3**(1), 125 (2014) (in Japanese)
9. Takayuki, A.: The use of the medium of recorded interrogation as a substantial evidence. Keio Law J. (31), 61 (2015) (in Japanese)
10. Kazuhiro, M.: Torishirabe DVD No Jisshitsu Shokoka. Quarterly Keiji-Bengo (82), 50 (2015) (in Japanese)
11. Hisao, T.: Kohan Shinri Kara Mita Sousa—Yoshinteki Shiten No Saihyoka. Seibundoh, Tokyo, Japan (2016) (in Japanese)
12. Hiroyuki, K.: Keiji Shiho Kaikaku To Keiji Bengo. Gendaijinbun-sha, Tokyo, Japan (2016) (in Japanese)
13. Takao, F.: Rokuon Rokuga Kiroku Baitai No Jisshitsu Shokoka Wo Meguru Mondaiten. Quarterly Keiji-Bengo (91), 26 (2015) (in Japanese)
14. The Supreme Court, 9/27/2005, Keishu 59(7), 753 (in Japanese)
15. Yuichiro, T.: A commentary of [2]. Sousa Kenkyu (805), 5 (2018) (in Japanese)
16. Yuichiro, T.: A commentary of [3]. Sousa Kenkyu (819), 11 (2019) (in Japanese)
17. Hiromi, M.: Higisha Torishirabe No Rokuon Rokuga Kiroku Baitai Katsuyo Wo Megutte. Kenshu (842), 3 (2018) (in Japanese)
18. Toyo, A.: Criminal Procedure, second ed. Yuhikaku, Tokyo, Japan (2006) (in Japanese)
19. Takuichi, K.: A commentary of [3]. Kenshu. (845), 3 (2018) (in Japanese)
20. Tatsuya, I.: Higisha Torishirabe Oyobi Kyojutsu Chosho No Arikata. Houritsu No Hiroba **66**(6), 56 (2013) (in Japanese)
21. Toshihiro, K.: The point at issues on criminal procedure act (9). J. Police Sci. **72**(2), 98 (2019) (in Japanese)
22. Andrea, R.: Machine Testimony, 126 Yale L.J. 1972 (2017)
23. Steven, W.T.: Teppler: Testable reliability—a modernized approach to ESI admissibility. 12 Ave Maria L. Rev. 213 (2014)
24. Supreme Public Prosecutors Office: Saiban-in Saiban Taisho Jiken Ni Okeru Higisha Torishirabe No Rokuon Rokuga No Shiko Kakudai Ni Tsuite. http://www.kensatsu.go.jp/content/000127631.pdf. Accessed 5 Feb 2020 (in Japanese)
25. Makoto, I.: Higisha Torishirabe Rokuga Eizo No Impact—Jisshitsu Shokoka No Kikensei Wo Megutte, Keiichi A., et al. (eds) The Legal Process in Contemporary Japan, vol. 2, Shinzansha, Tokyo, Japan (2017) (in Japanese)
26. Ken-ichi, K.: Sousa Dankai No Kyojutsu Rissho Ni Kansuru Mondai Kaiketsu Ni Muketa Ichi Kosatsu. Hanreijiho (2312), 14 (2017) (in Japanese)
27. Takashi, U.: Torishirabe No Rokuon Rokuga Kiroku Baitai No Shoko Riyo Ni Tsuite. Criminal Law J. (60), 44 (2019) (in Japanese)
28. Masayuki, T.: A commentary of [3]. J. Police Sci. **71**(11), 176 (2018) (in Japanese)
29. Makoto, I.: Cutting edge of the suspect interview recording—approach from the law and empirical science. Horitsu Bunka Sha, Kyoto, Japan (2016) (in Japanese)
30. Tatsuya S., Kosuke W.: Torishirabe Kashikaron No Shinrigakuteki Kento. Horitsu Jiho **83**(2), 54 (2011) (in Japanese)

31. Kosuke, W.: Shinrigaku Ni Okeru Torishirabe Rokuon Rokuga No Riyo No Kongo. Quarterly Keiji-Bengo (89), 138 (2017) (in Japanese)
32. Shugo, H.: Torishirabe No Rokuon Rokuga Kiroku No Shoko Riyo—Toriwake Jisshitsu Shoko Riyo No Kanousei Ni Tsuite. Sousa Kenkyu (785), 2 (2016) (in Japanese)
33. Akira, G.: Keisoho Kaisei To Torishiarbe No Rokuon Rokuga Seido. Horitsu Jiho **88**(1), 12 (2016) (in Japanese)
34. Toshiharu, K.: Torishirabe No Rokuon Rokuga Kiroku Baitai No Shoko Riyo No Arikata—Kensatsu No Tachiba Kara. Criminal Law J. (60), 50 (2019) (in Japanese)
35. Yuki, T.: A commentary of [3]. Hogaku Shimpo **126**(11,12), 341 (2020) (in Japanese)
36. Takao, N.: The criminal justice system in a new era. J. Criminal Law **56**(3), 346 (2017) (in Japanese)

Governance Framework for Facial Recognition Systems in Japan

Aimi Ozaki

Faculty of Social Sciences, Kyorin University, Tokyo, Japan
ai-ozaki@ks.kyorin-u.ac.jp

Abstract. With the evolution of machine learning, the facial recognition technology has already been used in different kinds of situations. In a field where such technological developments have a significant impact on society and the people who live in it, rulemaking is carried out by a variety of actors. Therefore, the METI suggests that what is most important is how quickly rules can be changed in line with technological innovations and the changes in business settings resulting from such innovations and how we can build a mechanism for such changes. This paper introduces how the Japanese government and the private sector analyze the risks posed by using the face recognition technology. Then, the issues of these analyses are discussed. Generally, businesses are at the heart of using personal data to drive innovation and create value for society. As a result, they are expected to play a central role in the design of governance. Also, given the nature of face recognition, which fosters discrimination and surveillance, this paper introduces some ideas that can be used as a reference point for businesses to consider when establishing governance in the use of the face recognition technology.

Keywords: Facial recognition · Privacy · Personal data · Algorithmic bias · Fairness

1 Governance Design in the Facial Recognition Technology

1.1 The Development of the Facial Recognition Technology

With the evolution of machine learning technologies and algorithms, the facial recognition technology has been used in many kinds of situations [1]. The facial recognition technology is a type of biometric authentication technology that can automatically identify a person based on his or her physical and behavioral characteristics. The biometric technology is a technology for identification, and it is classified into two types: those that use physical characteristics and those that use behavioral characteristics. The facial recognition technology is categorized as the former type [2]. In general, face recognition is defined as an authentication method in which a person is identified by comparing his or her face information, which is captured using a camera, with other face information in a database. There are two types of face recognition: "active authentication," where the person is authenticated with his or her consent, and "non-active authentication," where the person is authenticated without being aware of

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 52–63, 2020.
https://doi.org/10.1007/978-3-030-62803-1_5

it. An example of the former is the identity verification system, which is used at events to determine if a person who purchased a ticket is the same one who enters the venue. An example of using the face recognition technology for investigative purposes is to search for wanted criminals and suspects on the run. In such cases, it would be difficult or almost impossible to obtain the consent of the wanted criminals or suspects for using face recognition. Therefore, the use of the face recognition technology for investigative purposes is considered a "non-proactive authentication [3]."

1.2 The Need for Governance Design in the Facial Recognition Technology

In a field where such technological developments have a significant impact on society and the people who live in it, rulemaking is carried out by a variety of actors. The rules of each country influence each other and are increasingly cross-referenced in many situations. Furthermore, international harmonization is increasingly necessary for such situations. The actors in this context include the state, citizens, companies, and platforms. Therefore, the METI (Ministry of Economy, Trade and Industry) suggests that what is most important is how quickly rules can be changed in line with technological innovations and the changes in business settings resulting from such innovations and how we can build a mechanism for such changes [4, 5].

1.3 Direction of the Governance Design

The METI states that innovative technologies and services can also rapidly change the risk landscape. In this case, the risks include the following. In data collection and management, digital service providers collect enormous amounts of accurate personal data, such as a person's behavior history, health status, economic activities, thoughts, beliefs, hobbies, and preferences, thus raising many privacy risks. For example, the use of such personal data in targeted political advertisements can potentially cause harm to democratic systems. In areas of data analysis, since the autonomous decision-making by algorithms, which involves no human intervention, now takes over important positions in society, the necessity to discuss the safety and adequacy of such algorithms is increasing. Since machine learning is not based on static rules and its outputs are dynamically produced by adjusting the weights of variables through statistical processes, it is difficult to fully predict its behavior. Additionally, in the case of deep learning algorithms, even if a decision made by a specific algorithm is found to be inadequate, we are currently faced with the problem that the cause of such behavior can sometimes be difficult to explain. When these outputs of the algorithms are fed back into the physical world, there are risks of unpredictable accidents, magnifying discriminations, and unfair bias driven from datasets [6].

Unlike the DNA, which contains genetic information related to diseases, etc., face information does not have a high degree of privacy on its own. However, the facial recognition technology integrates (links) the face information with other personal information, which transforms the nature of the face information into a high degree of privacy. In the use of the facial recognition technology, an individual's face information can be collected by businesses and other entities. Businesses are at the heart of using personal data for driving innovations and creating social values. As a result, they are expected to play a central role in the design of governance. Thus, it is necessary for society to properly control the risks posed by such innovations and to design governance that realizes various social values, such as privacy, democracy, and anti-discrimination.

Governments are also supposed to make flexible and appropriate decisions depending on the risks posed by different innovations and the situation concerning the corporate self-regulations and guidelines of various organizations. In addition, they need to be equipped with the necessary expertise so that they can be able to make decisions on whether regulations would be necessary or not.

Accordingly, in the next chapter, this paper introduces how the Japanese government and the private sector analyze the risks posed by the use of the face recognition technology. Then, the issues of these analyses are discussed.

2 Guidelines Provided by the Government and Organizations in Japan

2.1 Report for the Large Face Recognition Study in the Osaka Station

The NICT (National Institute of Information and Communications Technology) planned to begin a large-scale face recognition experiment in April 2014 with the aim of improving the emergency response in the events of disasters. The NICT planned to install 92 digital video cameras in the Osaka station to film passers-by and test whether or not it is possible to create human traffic information. The purpose of this experiment was to perform a human flow analysis based on the face information acquired from the surveillance cameras and the face recognition technology. Also, the NICT spokesman emphasized that the data cannot be used to identify people and that it will abide by Japan’s Personal Information Protection Law when it is handled. However, the experiment was met with so much criticism that it was eventually postponed.

A report on this experiment was submitted by an independent committee, which made the following observations. “In today’s society, where the advanced digital technology and the Internet are widely used, extracting the individual’s unique information from images of the whole body and face, etc. makes it possible to collect and record behavioral history without consent. This has the potential risk of harming the privacies of life. Consequently, the interest in not having such information extracted without due diligence constitutes a right to privacy. Then, the interest deserves legal protection. In addition, the average facial data required to generate feature information from facial images may be publicly available. Algorithms for extracting the feature information may be made available to the public as well. If so, third parties can use these to track individuals and obtain their behavioral histories. Moreover, the feature information generated from facial images and gait patterns cannot be changed unlike passwords. They require the same level of legal protection as biometric information such as fingerprints, iris and DNA information. Even if the risk of re-identification of a particular individual is minimized by the implementation of thorough security measures, it is assumed that there are those who wish to refuse to be photographed. The location for video cameras is a place where people using the Osaka Station have no choice but to pass through. Therefore, some means of refusing to be photographed should be provided."

2.2 Opinion Concerning the Legal Restrictions on Facial Recognition Systems

The Japan Federation of Bar Associations ("JFBA") prepared its "Opinion Concerning the Legal Restrictions on Facial Recognition Systems" on September 2016 [7]. The opinion concerns the system whereby the police collect facial image data from an unspecified number of people nearby a crime scene, generate quantified data of characteristics to specify each individual, and conduct searches in a pre-generated database of facial recognition data of specific people, which is used to match the identity of suspects or other persons. In regard to this system, it is the opinion of the JFBA that the nation should establish laws that incorporate the items listed below, amend relevant legislation and undertake similar measures to enact appropriate regulations, as well as recognize the guarantee of access rights for suspects, defendants and those in similar circumstances.

(1) Limitations on Usage Conditions

(i) Collection of facial image data recorded via security cameras or the like, for the purpose of criminal investigation by the police, should be conducted by court order (however, this would exclude facial image data from stores or other facilities, where such equipment has been legally installed in an area where the installer has authority).

(ii) Generating facial recognition data from images nearby a crime scene should be limited to those instances required for investigation of organized crime, where the crime infringes upon paramount public interests (hereinafter referred to as "serious organized crime"). In this case, facial recognition data that has been legally generated should be destroyed as soon as it is no longer required for said investigation.

(iii) Where use is permitted for facial recognition data generation from facial image data of suspects, previously convicted persons, or the like, whose facial image data is already in the legal possession of police, such use shall be limited to those with a previous conviction for serious organized crime.

(iv) Facial recognition data registered in the facial recognition database should be limited to those with prior convictions for serious organized crime. Furthermore, registration periods should be established for such data, and this data should be deleted immediately after the end of such registration period.

(v) Facial recognition database matching should be limited to such cases that require specific investigations for serious organized crime, and conditions for permitted methods should be clearly predefistances.

(2) Monitoring by the Personal Information Protection Commission

The Personal Information Protection Commission should be able to check whether facial image data collection, facial recognition data generation, usage and disposal, compiling of facial recognition databases, registration to facial recognition databases, usage status of facial recognition databases, and data deletion or the like from the facial recognition databases, are conducted in an accurate and appropriate manner.

(3) Disclosure of Basic Information
The mechanisms and search accuracies of facial recognition systems should be periodically published.

(4) Rights of Suspects, Defendants and Those in Similar Circumstances
The facial recognition system can provide a means for claiming an alibi for those not connected to the facts of the crime. Requests by suspects, defendants, and those in similar circumstances for matching using the facial recognition system should be recognized. Furthermore, those erroneously registered in the facial recognition system should have recognized rights to request disclosure and deletion.

2.3 Guidebook on the Processing of Personal Data in One ID Service in Airports Utilizing Face Recognition Technology

In March 2020, the Civil Aviation Bureau of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) established Guidebook on the Processing of Personal Data in One ID Service in Airports utilizing Face Recognition Technology. The One ID service uses the facial recognition technology to verify the identities of passengers during international flight departure procedures. The Narita and Haneda airports introduced the One ID service with the aim of making the boarding process faster. This is the first official guideline that specifically addresses the handling of the face recognition technology in Japan, and it serves as a basis for the formation of a norm for the future use of biometric information. Also, many businesses use this guideline as a reference.

Face information is a highly immutable identifier that can easily be obtained without a user’s intention. In addition, there have been some cases in which people have strongly objected to the use of face recognition for tracking purposes. Therefore, in the use of the face recognition technology, not only do organizations have to comply with the Personal Information Protection Law and other relevant laws and regulations, but they also have to make sure that it is necessary to provide passengers with a clear explanation of the purpose of use and of the information management procedures. The guideline focuses on the following three points as ways to deal with privacy risks.

(1) Limitation of the purpose of use
The use of personal data should be limited to the boarding process only. Even if new needs arise in the future, the purpose of use should not be easily expanded or changed. In addition, the terms and conditions for handling personal data to be used should be clearly displayed to passengers.

(2) Secure traditional procedures
Only the passengers who wish to use procedures that utilize facial recognition should be allowed to use it, and some procedures that do not need facial recognition should remain in place.

(3) Retention period of the personal data

The used personal data in the One ID service should be deleted when the purpose of use is achieved and is no longer needed. The specific procedure is to delete the data within 24 h, and periodic audits of the service should be conducted.

2.4 Brief Summary

The risks in these guidelines concentrate on privacy. However, in recent years, other countries have pointed out that facial recognition has the potential to promote mass surveillance and discrimination. As mentioned in the introduction, the METI states that outputs of the algorithms are fed back into the physical world, there are risks of unpredictable accidents, magnifying discriminations, and unfair bias driven from datasets. In fact, privacy is not the only issue in the recent flaming cases in Japan. In the next chapter, this paper focuses on some recent cases and the new issues that are now discussed.

3 Recent Flaming Cases in Japan

3.1 Project Against Bookstore Shoplifting in Shibuya

In July 2019, bookstores in Shibuya started a joint project to share the facial image data of shoplifters and other relevant data with each other, and they used facial recognition cameras to detect the entry of shoplifters into their stores. Regarding the detection procedure, for example, a bookstore uses a facial recognition system to detect a subject's visit with a security camera. Then, the system notifies the clerk via a smartphone notification that a subject has entered the store. Afterward, the clerk calls out to the subject. Since it is not clear whether the subjects intend to shoplift or not, the clerks usually consider them to be criminals.

In order to ensure the protection of personal information and privacy, the project required compliance with the following requirements.

1. Compliance with the Personal Information Protection Law, etc.
2. Prohibition of use for purposes other than the intended use and the maintenance of confidentiality
3. Ensuring the information accuracy
4. Appropriate operation of the system for utilizing security images
5. Maintaining awareness as a member of the participating shops
6. Study of the operation personnel

3.2 JapanTaxi Use of Facial Recognition for Creating Targeted Ads

The JapanTaxi company installed tablets in the backseats of its vehicles, where the tablets have facial recognition systems for scanning the faces of passengers so as to determine their gender, age, and other characteristics. The purpose of the system is to deliver targeted ads to the on-board tablets. The existence and purpose of the cameras were announced on the JapanTaxi website, and the facial recognition tablet has been in use.

After the service was launched, a Google engineer, Rosa Golijan, posted a photo of the warning on this tablet, which says "This taxi tablet is using a face recognition system with an image received by the tablet's front camera. The image data is used to estimate gender in order to deliver the most optimized content. The gender estimation runs once at the beginning of the advertisement program and the image data is discarded immediately after the estimation processing. Neither the tablet nor the server records the data." Her tweet was attached to a face with one eyebrow-raised emoji. The tweet led to a flurry of criticisms of JapanTaxi, and the issues at stake were privacy and gender, where the following criticisms were raised: "There is a problem with changing ads based on gender. It is based on the prejudice that women will like this stuff,' but there is no basis for that prejudice. For some people, gender and sexuality do not match. It encourages a biased view of gender."

In September 2019, The Personal Information Protection Commission (PPC) issued an administrative directive to JapanTaxi, and it was the second directive. As a result, JapanTaxi has revised its privacy policy, and it is now in the process of informing its passengers of the acquisition purpose in the clearest manner possible.

3.3 The Demonstration Experiment of the Osaka Metro Ticket Gate with the Facial Recognition System

In December 2019, the Osaka Metro began a demonstration experiment of a face recognition ticket gate. The target of the experiment was the employees of the Osaka Metro, and the period of the experiment is from December 2019 to September 2020. This face recognition system takes a picture of a user's face with a camera when he or she passes through the ticket gate. Then, the acquired face feature data is sent to the server. This data is checked against a pre-registered face photograph. If the data is verified, the door of the ticket gate is opened. The facial recognition camera is constantly running, but it is not recording. As soon as it detects a subject's face, the system converts the face into feature point data, which is then used for verification against the authentication server. The acquired images are managed on a dedicated network that is not connected to the outside world, and the acquired images are not used for any purposes other than the verification. The period of use of the recorded data shall be less than six months after the data is acquired. The expired data is either destroyed or processed in a manner that prevents the identification of individuals, and the verification data may be provided to cooperating manufacturers; however, the use of the data for other purposes is prohibited.

A spokesman for the Osaka Metro said that it would cooperate with the police if asked to provide information for investigative purposes. The Osaka Metro has also indicated that it intends to extract and use attribute data, such as gender and age data, from the personal information. According to some reports, although face matching with sunglasses and masks is not possible, it is possible to perform face matching with a photograph of the subject's face.

4 Summary of the Issues

The criticism found in the flaming cases can be summarized from three perspectives.

4.1 "Database-Related Issues"

First, there are concerns regarding the expansion of the range of the collected face information in the database and the retention period of this information. This is a concern about the database itself, which is the collection and storage of information. In this paper, this issue will be referred to as the "Database-related issues."

In the initial sense, each face is exposed to the public, and the face itself is not confidential. As mentioned at the outset, the face information, unlike the DNA, which contains genetic information related to diseases, etc., does not have a high degree of privacy on its own. Face images can be acquired without the invasion of the human body. Thus, the impact of the face information acquisition is smaller than that of forced urine or forced blood collections. According to a Japanese precedent, any person has the right not to have his/her face or appearance photographed without consent or good reason, and if a police officer, without good reason, has photographed a citizen's face or appearance, such an act is in violation of the purport of Article 13 of the Constitution and therefore it is unallowable [8]. This precedent has not yet lost its force. Nonetheless, in the present-day society, facial images can be collected through open-source social media without the subject's consent. Also, the facial recognition technology integrates (links) the face information with other personal information, which transforms the nature of face information into a high degree of privacy.

4.2 Concerns Regarding the "Surveillance Society"

The Osaka Metro has stated that it may provide the collected face information to investigative agencies. In this case, the facial recognition system could act as a surveillance device for investigative agencies.

As such, facial recognition can be used for mass surveillance. The fundamental rights considerations in the context of law enforcement according to the European Union Agency for Fundamental Rights (FRA) states that the use of facial recognition technologies can have a negative impact on the freedom of assembly if people fear that the facial recognition technology is used to identify them ("chilling effect") [9]. The City of San Francisco has banned the use of face recognition technologies by San Francisco government agencies because "the propensity for facial recognition technology to endanger civil rights and civil liberties substantially outweighs its purported benefits, and the technology can exacerbate racial injustice and threaten our ability to live free of continuous government monitoring [10]." Consequently, it is necessary to prevent the use of the face recognition technology for promoting the "surveillance society."

4.3 Does Facial Recognition Create Discrimination?

The case of the JapanTaxi is an example that gender discrimination may occur depending on the use of the facial recognition technology. Even in other countries, it is argued that the facial recognition technology is racial. The GDPR provides that processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited. In February of 2018, MIT researcher Joy Buolamwini and Timnit Gebru, then a Microsoft researcher, published a study that facial recognition systems have a harder time identifying women and darker-skinned people [11].

5 Conclusion

5.1 Towards Solving "Database-Related Issues"

Modern information technology has made the boundaries between short-term and temporary storage and the creation of databases extremely blurred. Thus, "the collection of information on the basis of short-term preservation should be clearly grounded and controlled by law [12]." According to this view, on the contrary, with the law and the oversight agencies created by the law to make sure that the information is not stored in the long term, many of the privacy issues can be eliminated. In turn, what powers should be granted to such an oversight body? In light of the present situation where the collected face information by the private sector is used by government agencies without a warrant, it is necessary for an independent agency to oversee a series of steps for biometric information, including face information, so as to support the responsible use of the face information, conduct audits, and apply sanctions [13]. In Japan, the Personal Information Protection Commission (PPC) can be the equivalent of this agency. If it were so, the establishment of a commissioner specializing in biometric information as part of the PPC should be proposed.

5.2 How to Avoid Becoming a "Surveillance Society"?

Digital service providers collect precise and vast amounts of personal data, which enables them to precisely understand an individual's behavioral history, health status, economic activities, thoughts, beliefs, hobbies, and interests, etc. Facial recognition can create a "data stigma." Therefore, businesses that provide such services should not reveal any special care-required personal information. Meanwhile, personal controllability should be ensured for customers.

5.3 Anti-discrimination

Before introducing the facial recognition technology, it is necessary for businesses to consider whether fairness is assured for the face recognition technology. Concretely, businesses (especially engineers) need to make sure that development shall not cause discrimination. Also, businesses should be aware that such discrimination can easily occur [14].

Recently, in response to the growing anti-discrimination movement, a number of companies have stopped providing facial recognition systems to criminal investigations. For example, on June 8, 2020, IBM said that it would stop offering facial recognition software for "mass surveillance or racial profiling" to respond to the death in police custody of George Floyd. Amazon has also banned the police from using its controversial facial recognition software for a year.

However, the facial recognition technology should not be uniformly banned. Rather, it is easier to detect fraud in algorithms than in humans, and it is easier to correct fraud in algorithms than in humans. By choosing the appropriate criteria, it is possible to construct algorithms that comply with fairness. It is also possible to increase fairness by using the facial recognition technology appropriately. It is necessary to carefully collect input and test data, generate models, and perform appropriate audits when using the facial recognition technology. It is also important to categorize the purposes and targets of using face recognition technology in detail [15]. There is a need for businesses to put in place an appropriate audit system for the use of technology to avoid leading to those actions. Such a system would also give an advantage in gaining the trust of consumers.

5.4 Appendix: Principles for Proper Use of the Face Recognition Technology

As mentioned in the introduction, businesses are at the heart of using personal data to drive innovation and create value for society. As a result, they are expected to play a central role in the design of governance. Finally, given the nature of face recognition, which fosters discrimination and surveillance, this paper introduces some ideas that can be used as a reference point for businesses to consider when establishing governance in the use of the face recognition technology [16].

(1) Principle of the Self-Determination of Information

Businesses shall enable users to exercise their rights to the self-determination of information. Businesses should aim to design and implement user interfaces for this purpose.

(2) Principle of Effective Remedies

Businesses should understand the potential of harm to users and provide effective remedies for them.

(3) Principle of Providing Alternatives

Businesses should provide a way for users who do not agree with facial recognition to receive the same services as before.

(4) Principle of the Limitation of the Purpose of Use

Businesses shall use the user's face information only for the purpose of its predetermined use.

(5) Principle of Safety Management

Businesses must take security measures, such as encryption, non-retention, and information security audits, by third parties.

(6) Principle of Appropriate Use

Businesses should make sure that the facial recognition technology is not used in a discriminatory way.

(7) Principle of Transparency

Businesses shall establish a policy on how to respond to requests for the disclosure of information about users from law enforcement agencies.

(8) Principle of Non-Use of Excluded Data

Businesses shall not use data that is not included in the scope of the intended use.

(9) Principle of Preliminary Consideration

Businesses should consider each point of these principles in advance.

(10) Principle for Strengthening Communication

Businesses should ensure appropriateness throughout the supply chain, and they shall provide information relevant to the risks that may arise to users to strengthen communication.

References

1. https://www.nec.com/en/global/solutions/biometrics/face/index.html
2. Research and Legislative Reference Bureau National Diet Library, Current Trends in Biometrics, Research Materials 2018-6 http://dl.ndl.go.jp/view/download/digidepo_11257103_po_20180602.pdf?contentNo=1
3. Suzuki, T.: The biometrics with focus on video face recognition technology. J. Inf. Process. Manage. **60–8**, 564–573 (2017). https://www.jstage.jst.go.jp/article/johokanri/60/8/60_564/_pdf
4. Terada, M.: Public Law on Advanced Technology and Regulation, Keiso Shobo (2020)
5. Hioki, T.: Utilization of “face” information and act on the protection of personal information. Bus. Houmu **17–4**, P87 (2017)
6. The METI (Ministry of Economy, Trade and Industry), Governance Innovation: Redesigning Law and Architecture for Society 5.0 https://www.meti.go.jp/press/2020/07/20200713001/20200713001-2.pdf
7. Japan Federation of Bar Associations, Opinion Concerning the Legal Restrictions on Facial Recognition Systems https://www.nichibenren.or.jp/en/document/opinionpapers/20160915.html
8. SaikŌ Saibansho [Supreme Court] Dec. 24, 1969, Showa 44, SaikŌSaibansho Keiji Hanreishu [Keishu] **23**(12), at 162 (1969) https://www.courts.go.jp/app/hanrei_en/detail?id=34
9. European Union Agency for Fundamental Rights (FRA), Facial recognition technology: fundamental rights considerations in the context of law enforcement https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf
10. Stop Secret Surveillance Ordinance https://sfgov.legistar.com/View.ashx?M=F&ID=7206781&GUID=38D37061-4D87-4A94-9AB3-CB113656159A

11. Raji, I., Buolamwini, J.: Actionable auditing: investigating the impact of publicly naming biased performance results of commercial ai products. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 429–435 (2019) https://dam-prod.media.mit.edu/x/2019/01/24/AIES-19_paper_223.pdf
12. Yamamoto, T.: Consider the Right to Privacy Shinzansha (2017)
13. Mann, M., Smith, M.: Automated facial recognition technology: Recent developments and approaches to oversight. UNSWLJ. **40**, p. 121 (2017)
14. Zuiderveen Borgesius, F.: Discrimination, artificial intelligence, and algorithmic decision-making (2018). https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73
15. Kamishima, T., Akaho, S., Asoh, H., Sakuma, J.: The independence of fairness-aware classifiers. In: 2013 IEEE 13th International Conference on Data Mining Workshops, pp. 849–858. IEEE (2013) http://www.kamishima.net/archive/2013-ws-icdm-print.pdf
16. NEC Corporation, Biometric identification (facial features) data between businesses empirical study on architecture for collaboration https://www8.cao.go.jp/cstp/stmain/b-2-13_200318.pdf
17. Hamann, K., Smith, R.: Facial recognition technology: where will it take us. American Bar Association Available at: https://www.americanbar.org/groups/criminal_justice/publications/criminal-justice-magazine/2019/spring/facial-recognition-technology/. Accessed 6 May 2019

Advanced Software and Patents: A Patentability Balance for Fostering Technology

Michele Baccelli[1](✉), Kazuto Kobayashi[2], Steven C. Sereboff[3], and Mitsuyoshi Hiratsuka[4]

[1] Hoffmann Eitle, Munich, Germany
mbaccelli@hoffmanneitle.com
[2] Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan
kobayashi@sangaku.titech.ac.jp
[3] SoCAL IP Law Group LLP, Santa Barbara, CA, USA
ssereboff@socalip.com
[4] Tokyo University of Science, Shinjuku-ku, Tokyo, Japan
hiratsuk@kb3.so-net.ne.jp

Abstract. Technology advances have usually brought wealth and health to society, while also raising ethical and policy issues. Technology and patents are linked to each other in that patents have been conceived also as a tool for rewarding and motivating the inventor's endeavors. Patents have been subject to critical views, especially in the era of software where the innovation might be based on non-technological improvements, sometimes arising from new business models, social behavior or trends. For patents to support innovation in its beneficial contribution to society and avoid abuse, a high-quality patent system should be ensured. Many people believe that this requires in the first place to exclude from patent protection those inventions that are purely non-technical. In this paper, the authors analyze the legal situation at the European Patent Office (EPO), considered as setting amongst the worldwide strictest standards in terms of software patentability, and wherein patents are granted only to technology advances. Within this context, the case of software simulators is discussed, since this is directly linked to intangible software inventions and since it is the subject of a case on a point of law pending before the EPO highest instance, such that its outcome may impact patentability of modern software technologies. In conclusion, the authors argue in favor of maintaining the present EPO practice on simulators, as this is seen as a fair balance in Europe between patent protection for intangible inventions based on technology while excluding from patent protection those inventions that are non-technical arrangements.

Keywords: Patents · Patentability · New software technologies

D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 64–76, 2020.
https://doi.org/10.1007/978-3-030-62803-1_6

1 Introduction

Software innovations are increasingly decoupled from tangible systems. The social value of these innovations will necessarily increase with more human effort and the leverage of technologies such as artificial intelligence (AI).

Patents have been usually justified as an incentive to inventors to make and publish their innovations. Yet, the foundation of patents has been often challenged, and perhaps even more in the era of software, with many of the arguments in favor or against patenting being unsurprisingly and understandably biased by the economic interests and business models of the parties bringing such arguments forward.

Against the above background, the issue of whether and how certain software inventions can be patented is herein discussed, believing that the patent system must provide fair and balanced opportunities for patent protection. More in detail, focus will be on software inventions that do not necessarily produce a tangible result, like for instance in the field of bioinformatics, artificial intelligence, modeling of systems beyond the physical world.

The focus of the present contribution is thus on whether and under which conditions legal provisions should make patent protection available to software inventions, believing that society may benefit from such patents when a right balance is found. We focus here on patentability as seen at the European Patent Office (EPO) [1]. The main reasons are two-fold: the EPO is generally considered as applying the strictest standards in terms of patent eligibility of software, and it links such eligibility to an invention being technical [2].

2 Patentability of Software at the EPO

The EPC limits patentability of software inventions by stipulating that inventions relating to computer programs “as such” [3] shall not be eligible to patent protection. The meaning of “as such” has been clarified by and developed through case law, the main aspects of which are here summarized.

2.1 General Requirements and Technical Character

The Enlarged Board of Appeal (EBoA) [4] has previously held in G2/07 [5] that using forces of nature belongs to the core of an invention. It follows that using forces of nature is a prerequisite for an invention to be patent eligible.

A computer program executed on a computer relies on electrical signals that certainly represent forces of nature and would thus not to be in contradiction with G2/07. However, in milestone decision T1197/97 [6] the Board found that for such a computer program to be patentable it is necessary to produce a technical effect that goes beyond the effect lying in the mere exchange of signals between computer components [7], and must in fact cause the solution of a technical problem. This reflects a public policy decision.

EPO case law has then established that a software invention is patentable if it has technical character, without having provided a definition for technical [8, 9], on grounds that any definition may become obsolete and then not anymore in line with technological progress [10]. It thus appears that the intent of the legislator was to allow room for patent protection of new technologies that could not be foreseen at the time of drafting the EPC. As a general guidance, it can be fairly said that, under the EPC, an invention is technical if it is based on technical considerations and motivations [11].

2.2 Intangible Solutions

Software inventions may be directly applied to the control of a technical system or parts thereof, e.g., controlling the braking of a vehicle, or controlling of a production line or manufacturing process. In such cases, it is rather straightforward to show technicality, since the software is directly applied on or to a real object wherein both the object and the action on the same are immediately recognized as technical.

Certain software inventions are patentable if they produce a technical effect within a physical object. In addition, a computer invention may be technical and patentable if the technical advantage brought about is on an intangible new functionality. Furthermore, a computer invention which uses forces of nature by virtue of the execution on a computer may be patentable also if relates to an intangible technical solution, and not necessarily to the modification or creation of a physical item, at least as long as considerations and knowledge of the functioning of a computer play a role in the conception of such invention [12].

3 Beyond the Physical World: Patentability of Software Simulations

In T1227/05, in the following also Infineon [13], the BoA sets a general condition under which a computer-based software simulation can be regarded as patentable, which was also reflected in the EPO Guidelines for Examination and to which Examiners are expected to abide. This condition, which we name the “Infineon condition”, was deemed necessary and sufficient for an invention to have technical character. However, the Infineon approach came later under criticism by other Boards, culminating in one Board expressly criticizing its reasoning and conclusions to a point where it referred some questions to the EBoA under case G1/19. Thus, the point of law under scrutiny by the EBoA may be summarized as to whether the Infineon condition is also a sufficient condition, or whether other condition(s) need to be met for a claim to be technical (at the time of writing, the G1/19 is pending). The situation can be summarized as in Fig. 2, and is addressed in the following.

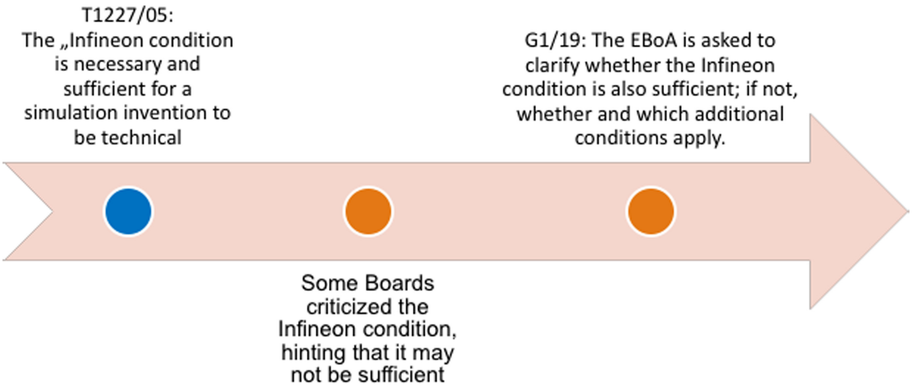

T1227/05:
The „Infineon condition is necessary and sufficient for a simulation invention to be technical
Some Boards criticized the Infineon condition, hinting that it may not be sufficient
G1/19: The EBoA is asked to clarify whether the Infineon condition is also sufficient; if not, whether and which additional conditions apply.

Fig. 1. Case law development on software simulations over time

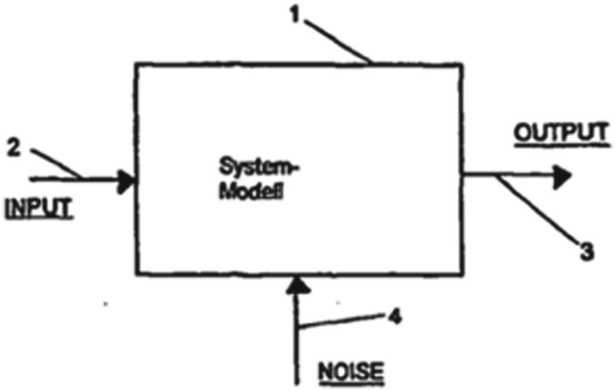


Fig. 2. Illustration of the simulation underlying the Infineon case

3.1 The Infineon Decision

The German company Infineon filed a patent application directed to a method for the numerical simulation of an electronic circuit subject to 1/f noise, wherein the circuit is described by a model featuring input channels, noise input channels and output channels. The simplified claim recites:

Computer-implemented method for the numerical simulation of a circuit with a step size δ which is subject to 1/f noise, wherein:

the circuit is described by a model (1) featuring input channels (2), noise input channels (4) and output channels (3);

[…] input channels (2) and the output channels (3) […] described by […] equations;

an output vector (OUTPUT) is calculated for an input vector (INPUT) […] and for a noise vector (NOISE) […];

[steps for generating the noise vector].

The invention can be better understood by referring to Fig. 1, showing an electronic circuit as a sort of black box (1) having input (2) and output (3) and being subject to noise (4), wherein mathematical equations are used to describe its behavior. The computer performs a simulation in the sense of executing a method wherein the equations are calculated.

Hence, the method allows obtaining an (estimated or simulated) output corresponding to an input and noise. In this way, it is possible anticipating the behavior of a circuit design before the circuit is realized; this may considerably shorten the development cycle of a circuit, since less or virtually no prototype needs to be built.

The Examining Division refused the claim directed to a "method for the numerical simulation of a technical system subject to 1/f noise" on grounds that its steps could also be executed mentally, and that mental activities are not patentable [14]. Further, the Division objected that steps like generating a series of random numbers or defining a noise having a predetermined frequency spectrum are non-technical features since they are based on mathematical models without providing a technical effect. In addition, the first instance decision argued that the terms referring to the technical system (by defining for instance input channels of a model of the system) do not provide the required technical character to the simulation. In summary, the claimed method was considered as merely resulting in mathematical steps without providing technical character.

During appeal, certain claim amendments were made, in particular to specify that the method is a "computer implemented method for the numerical simulation of a circuit… which is subject to 1/f noise", that "the circuit is described by a model featuring input channels, noise input channels and output channels" and that "the performance of the input channels and the output channels is described by a system of differential equations".

The Board found such amended claim to be adequately defining a technical purpose for a computer implemented method, in particular that the simulation of a circuit more precisely specifies the technical ambit of the claim and implicitly the technical considerations involved, and that the simulation by means of a computer make clear that the invention does not anymore encompass a purely mental simulation.

The Board also noted that the mathematical expression mentioned in the claim are not to be regarded as abstract or mathematical formulae as such, but rather as being relevant to the circuit simulation and thus as contributing to the technical character of the simulation. It can be broadly said that, once the technical ambit of the claim is acknowledged, then the mathematical aspects therein mentioned should be also regarded as technical, at least as far as they contribute to achieving the technical effect and object of such claim.

Summarizing, T12275/05 sets out that a simulation claim is technical if it is limited to the simulation of an adequately defined class of technical items. In the following, we will call this the "Infineon condition", representing at least a necessary condition to be fulfilled for a simulation to be technical and patent eligible.

3.2 Criticism to the Infineon Decision

After the decision in T1227/05 Infineon was rendered, also the Guidelines for Examination at the EPO were correspondingly updated, and its approach generally followed.

However, criticism started to emerge as at least indirectly mentioned or hinted in a few decisions by the Boards of Appeal.

For instance, in case T1265/09 dealing with a computer simulation for determining an efficient schedule for call center agents, the Board seemed to outline that the Infineon condition may represent only a necessary condition for acknowledging technical character, but that it may not be sufficient to that effect [15]. However, the Board found that simulating a call center in view of the agents' schedule and skills would not pertain to the simulation of a technical system, such that the Infineon test would already fail; as such, the Board did not investigate further or comment on whether the Infineon condition would also be sufficient to justify technicality of the simulation.

In case T625/11, the Board was called to deal with a computer simulation method for establishing a limit value for an operational parameter of a nuclear reactor, wherein the limit value is based on simulation of the reactor. One concern expressed by the Board related to the fact that the claim is not limited to a use that leads to a technical effect, as the claim would also encompass a simulation having non-technical objectives, like for instance checking compliance with legal requirements which would thus represent an exclusively administrative purpose [16]. Having expressed such concerns, the Board nevertheless decided to follow the earlier approach by Infineon, without further elaborating on possible criticism or hypothetical weaknesses of the Infineon reasoning.

In short, while it was main practice to acknowledge technicality of a claim as long as the Infineon condition was satisfied, some (though minor in number) criticism started to emerge, in particular that some additional conditions may need to be shown by a computer simulation in order to qualify as a technical patentable invention.

3.3 Pending Clarification of the Law: The G1/19 Referral

The G1/19 Referral may be said to further develop and highlight the criticism expressed by other Boards or decisions. Let us look first into what the underlying invention is about.

3.3.1 The Invention Underlying the G1/19 Referral

The invention at issue deals with the computer simulation of pedestrians moving through a building structure, like train stations or airports. The results of the simulation can be used to verify the engineering design of the building, and possibly to assist the same engineers in modifying the design, before the structure is built, in order to meet certain criteria, like for instance the number of passengers the station (or people the building) can handle, or how easily the building structure can be evacuated. The patent application was refused by the patent examiners, and then subject to appeal under case T0489/14 [17]. During the appeal phase, it was acknowledged that the case at issue shares quite several similarities with the Infineon case, since they both relate to the simulation or modelling of an adequately specified class of technical items, a circuit and a building, respectively. However, the Board deciding on the simulation of building structure did not feel comfortable with the Infineon approach, and even stated that it would have decided differently as in that earlier case. Since this could lead to divergence in case law, which is highly undesirable (to say the least), the deciding Board referred the following three questions to the Enlarged Board of Appeal:

1. In the assessment of inventive step, can the computer-implemented simulation of a technical system or process solve a technical problem by producing a technical effect which goes beyond the simulation's implementation on a computer, if the computer-implemented simulation is claimed as such?
2. If the answer to the first question is yes, what are the relevant criteria for assessing whether a computer-implemented simulation claimed as such solves a technical problem? In particular, is it a sufficient condition that the simulation is based, at least in part, on technical principles underlying the simulated system or process?
3. What are the answers to the first and second questions if the computer-implemented simulation is claimed as part of a design process, in particular for verifying a design?

In the following, we attempt to simplify and address some of the issues raised by the referring Board, trusting in the reader's understanding for any inaccuracy inevitably introduced by the simplifications.

3.3.2 The Reasoning of the Referring Board

The referring Board essentiality starts from a premise, namely ignoring that the claimed method recites a computer implementation, with the consequence that the remaining method – deprived of its computer implementation context – can, at least hypothetically, be performed mentally. Within this extrapolated context, the Board considers whether the remaining method requires non-trivial features for being implemented on a computer and finds in the negative on grounds that only knowledge of common data structures and algorithms is needed.

The Board also considers whether the design of the remaining method can be motivated by the internal functioning of the computer; also in this case, the Board finds in the negative on grounds that the implementation on the computer is straightforward.

Then, the Board moves on to discuss whether a further technical effect is provided, and state that it would likely conclude also here in the negative. However, the Board acknowledges similarities with the earlier T1275/07 Infineon decision, wherein the Board found in favor of technicality on grounds that a computer simulation directed to "an adequately defined class of technical items" is technical. The referring Board however disagrees with the earlier decision, hence the above questions to the EBoA. We focus on the following two aspects arising from the referral.

4 Possible Consequences from G1/19 on New Software Technologies

4.1 The "Infineon Condition": Why not a Suitable Solution?

As anticipated, the referring Board apparently disagrees with the Infineon condition, at least in that it is not a sufficient condition for acknowledging technical character of a computer simulation and expresses doubts with regard to that approach [18].

In particular, the referring Board states that a computer-implemented simulation "assists the engineer only in the cognitive process of verifying the design of the circuit or environment, i.e. of studying the behavior of the virtual circuit or environment designed. The circuit or environment, when realized, may be a technical object, but the cognitive process of theoretically verifying its design appears to be fundamentally non-technical", reason 15 (emphasis added).

In other words, the Board seems concerned that a computer simulation – and indeed many other types of computer inventions like in bioinformatics and AI [19] – often produces an intangible solution that can be used at a cognitive or abstract level.

We would like however to make a parallel between a computer simulator of a circuit and an oscilloscope, both available as tools for an electronic circuit designer.

A conventional way for verifying a circuit design is using an oscilloscope, through which the engineer can measure electrical values or voltage waveforms exhibited at certain points of the circuit. What the engineer conceptually and mentally elaborates on the basis of the measured results is not relevant to how the oscilloscope works or how it is internally built to measure the values. In addition, the measurements obtained by the oscilloscope may be used for activities different from design, e.g., confirming whether the product complies with certain legal requirements foreseen for a certain signal.

Nevertheless, there should be no doubts that oscilloscopes are technical, from a patentability point of view. The cognitive activities or administrative purposes that would follow from the usage of a tester or oscilloscope would normally play no role in assessing whether a specific design of the instrument is technical or not.

In the authors' view, there is at least a strong similarity between the computer circuit simulator and common testers and oscilloscopes, in that they are all tools assisting the engineer in verifying the design, wherein the cognitive process – even if itself non-technical – plays no role and comes only after the tool's output is provided.

Thus, the only difference between a simulator and an oscilloscope lies in that a simulator estimates or predicts the behavior of on a non-tangible (or not yet tangible) circuit, while an oscilloscope measures (though with a degree of estimation depending on accuracy) the behavior of a tangible circuit.

There are decisions recognizing that an invention, in the EPC sense of having technical character, need not necessarily result in a physical modification of a tangible part of an object, see e.g. the T423/03 Microsoft discussed above. Leaving aside whether a simulator may be taken to represent a further functionality of a computer in the sense of T423/03 Microsoft (as in fact a simulator may be argued to provide a new functionality to a computer like in Microsoft above), it seems correct stating at least that a simulator creates a new or improved tool available to the engineer for performing his/her tasks. The considerations necessary for programming the simulator appear to the authors as being indeed technical, since at least knowledge of the technical functioning of the simulated system and how this can be modeled to intangibly reproduce its behavior are required in the conception and development of the simulator.

On July 15, 2020, oral proceedings took place before the Enlarged Board of Appeal, with over 1600 registrations for the online streaming of the event having been recorded by the office. During the hearing [20], the parties [21] argued that the practice based on the Infineon decision should be confirmed and in particular that requiring a link to the physical world would not be necessary and not be commensurate especially when having regard of new technologies. It was also argued that it would be desirable lowering the bar, i.e. not mandating that the claim should necessarily specify a "class of technical items". In general, parties also argued in favor of a positive answer to the questions. During the hearing, the Chairman of the Enlarged Board of Appeal indicated – in a preliminary way, though – that the first question and the second part of the second question may be answered in the affirmative; if this is confirmed by the written opinion, and also depending on the actual reasons that will be given, the outcome of G1/19 may indeed be seen as confirming the Infineon practice and as perhaps even allowing to lower that bar, in that the conditions set out by Infineon are sufficient but not strictly necessary to justify technicality of a computer simulation. Which would be the minimum criteria necessary for conferring technical character to a software simulation may however likely remain an open issue subject to further development of case law. Such outcome may allow room for protecting new technologies without necessarily requiring a link to the real world, and thus allowing to possibly reflect the increased intangibility of modern inventions.

4.2 The EPO Practice of Separating Features Within a Claim: Is This Appropriate in View of New Technologies?

As well known to practitioners in the field of computer inventions, in claims including a mix of technical and non-technical features, the EPO considers only the technical features in the assessment of inventive step. The split between technical and non-technical features is done at the very initial stage of examining patentability and without having regard of the prior art and usually; features may thus be determined to be non-technical without resorting to evidence to support such findings. It is thus a crucial point for patent examination of software inventions.

In line with this common EPO practice, the Referral starts with the premise of "ignoring for a moment" the computer implemented limitation from the claim. The referring Board also states that such simulation can be performed purely mentally, however recognizing that this would be possible "at least in principle". In fact, we believe that no one would ever consider running a mental simulation of an electronic circuit, even within a large team of experts, and possibly with the user of pen and paper.

The authors contend the general approach of "ignoring for a moment" certain features related to the computer implementation and asserting that these could be hypothetically be performed mentally, on grounds that such assertion is not based on evidence (and would thus lead to legal uncertainty) and would further overlook that, at least under certain circumstances, certain software claim features are conceived because of the existence of a computer [22]. Furthermore, such an approach becomes untenable when the claim defines the scope of the simulation in terms beyond the scope of human solution, even given billions of people working for thousands of years.

In other words, the existence of computers may be a part of the creative process leading to an invention [23]. In further other words, without a computer, a person dealing with a technical activity would have not conceived those particular features because – without a computer – there would be no prospects of making a useful and practical technical use of the same; rather, that person would have recurred to other solutions, including those really suitable for mental performance when carrying out that technical activity.

Hence, when considering the advanced level of intangibility reached by modern technologies, consistently applying a separation between technical and non-technical features may not always be appropriate, in particular in those circumstances where a mental execution of certain features is undisputedly not feasible and where instead the existence of a computer leads to the conception of an invention [24].

5 Conclusions

As a result of the authors' analysis, especially in relation to the intangible nature of modern technologies, it seems correct still following the Infineon condition, namely that computer implemented simulations, when directed to an adequately specified class of technical items, represent a technical tool available to a skilled person dealing with a technical activity. The features of such a simulation tool should thus be considered technical and examined as for other types of inventions.

Also, separating certain features from their computer implemented context may not always be the correct way for assessing inventions, since it could ignore that the invention conceiving process finds motivations in the existence of the computer, and that the conceived solution makes practical sense and has actual applicability only because there is a computer to execute the same.

The above conclusion is believed not to be in contradiction with the basic requirements that an invention must be based on forces of nature (see G2/07), since the simulator makes use of a computer that functions thanks to such forces. The conclusion is also believed not to be in contradiction with the logic of the further technical effect underlying T1173/97 IBM [6]: In fact, a simulator provides a new computer-based technical functionality, wherein such functionality goes beyond the mere interaction between computer components when running any computer instructions. Still further, whether the simulation tool can be used in mental activities, including those of technical designing, should not be relevant to the present discussion in the same way as the use of a tangible tool like an oscilloscope does not deprive the technical character inherent to the oscilloscope. Last but not least, claiming the "adequately specified class of technical items" would put the simulator into its technical context and highlight the presence of certain technical considerations about the causal relationships underlying the "technical item" to be simulated, since otherwise without such technical considerations the simulator would not be capable of producing the intended output.

The founders of the EPC did not define what is technical or not and thus left somewhat open what should to excluded from patentability, exactly to allow the case law to develop in parallel with technological advances, which are embodied nowadays in the form of intangible software solutions. Putting additional conditions beyond those already set out by the Infineon decision would thus carry the risk of potentially excluding certain modern technologies from patent protection.

In summary, the authors believe that the Infineon condition should be considered as a necessary and also sufficient condition for acknowledging technicality, and that it in fact provides a fair and balance criteria for allowing protection to those intangible inventions that are still based on technology, while excluding from patent protection those software solutions that owe their innovations to non-technical recognitions and insights. Based on the oral presentations made during the recent hearing held in case G1/19 on July 15, 2020, it seems reasonable expecting that the Infineon condition and practice may be confirmed valid without any need for increasing the bar by requiring to specify a link to the real world. It cannot be excluded that the bar may even be lowered in the future, though the minimum criteria necessary for conferring technical character to a software simulation may likely remain an open issue subject to further development of case law.

References

1. The EPO legal provisions are set out in the European Patent Convention (EPC). The text of the EPC can be found at. https://www.epo.org/law-practice/legal-texts/epc.html
2. With the term software inventions, reference is made to those inventions that make use of a computer. The EPO uses however the term Computer Implemented Invention (CII) to refer to “claims which involve computers, computer networks or other programmable apparatus, whereby at least one feature is realised by means of a program”, see the Guidelines for Examination at the EPO, F-IV 3.9
3. Article 52 EPC: (1) European patents shall be granted for any inventions, in all fields of technology, provided that […]. (2) The following in particular shall not be regarded as inventions within the meaning of paragraph 1: […] (a) discoveries, scientific theories and mathematical methods; […] (c) schemes, rules and methods for per-forming mental acts, playing games or doing business, and programs for computers; (d) presentations of information. (3) Paragraph 2 shall exclude the patentability of the subject-matter or activities referred to therein only to the extent to which a European patent application or European patent relates to such subject-matter or activities as such
4. The Board of Appeal represents the second instance of the EPO entrusted with reviewing decisions issued by patent examiners; unless otherwise stated, reference will be made to decisions of the Technical Boards of Appeal. The EBoA can be said to represent the highest instance of the EPO entrusted with ensuring a uniform application of the law and with clarifying any point of law that is of fundamental importance; a Board of Appeal (or the EPO president, but not directly a party to proceedings) may refer questions to the EBoA if it considers that a decision is required for ensuring uniform application of the law, or if a point of law of fundamental importance arises. Decisions of the BoA and of the EBoA can be found at. http://www.epo.org/patents/appeals/search-decisions.html. BoA and EBoA cases are numbered according to the format T1234/YY and G12/YY, respectively

5. G 2/07, reasons 6.4.2 "[h]uman intervention, to bring about a result by utilizing the forces of nature, pertains to the core of what an invention is understood to be"
6. T1173/97, IBM, see e.g. r. 6.4 "[the technical character of a computer program] could be found in the further effects deriving from the execution (by the hardware) of the instructions given by the computer program. When said further effects have a technical character or where they cause the software to solve a technical problem, an invention which brings about such an effect may be considered an invention which can, in principle, be the subject-matter of a patent"
7. Borrowing the conclusion from T1227/05, Infineon, r. 3.3: for a computer program to be patentable, it needs to show "the potential for a technical effect going beyond basic hardware/software interaction in a computer"
8. In r. 9.2 of G3/08 dealing with patentability of computer implemented inventions, the Enlarged Board of Appeal stated: "We do not attempt to define the term "technical". […] the Enlarged Board only makes the assertions that "a computer-readable data storage medium" and a cup have technical character and that designing a bicycle involves technical considerations […]. It is to be hoped that readers will accept these assertions without requiring a definition of exactly what falls within the boundaries of "technical"."
9. It is noted that the German Federal Court of Justice (Bundesgerichtshof, the German highest instance) adopted a general definition of technical in the Rotetaube decision, and that this has been referred to and recognized as still holding valid in r. 6.4.2.1 of G2/07: "The term technical teaching was characterised as "a teaching to methodically utilize controllable natural forces to achieve a causal, perceivable result" […]"
10. More precisely, the term technology "was deliberately not defined by the legislator in order not to preclude that adequate protection would be available for the results of developments in the future in fields of research which the legislator could not foresee", see G2/07, r. 6.4.2.1
11. For a general discussion on EPO case law, see e.g. Computer Implemented Inventions under the European Patent Convention and Practice in Japan, M. Baccelli, M. Hiratsuka, AIPPI e-News No. 1, April 2018. https://aippi.org/enews/2008/edition01/computer_implemented_inventions_japan.html
12. T423/, Microsoft
13. The German company Infineon applied for a European patent that then became subject of appeal case T1227/05
14. In fact, Article 52 EPC prohibits patentability of mental activities, see Article 52(2) EPC: "(2) The following in particular shall not be regarded as inventions within the meaning of paragraph 1: […] (c) schemes, rules and methods for performing mental acts", see also note 3 above
15. T1256/09, IEX, reason 1.13: "Leaving aside the question of whether these conditions are indeed sufficient to contribute to a technical character, […]"
16. T625/11, Areva, reason 8.1 (reading translation from the French language in which the decision was issued): "[…] the process claimed […] could serve non-technical objectives or technical objectives, but not necessarily linked to the functioning of a nuclear reactor. […] The claimed process could also, as a second example, be implemented in order to establish, with competent authorities, that a given reactor fulfills the requirements in force required for its operation. The operation, entrusted to a design office, would then have an exclusively administrative purpose. […]"
17. The text of the referral can be found at: https://www.epo.org/law-practice/case-law-appeals/pdf/t140489ex1.pdf. The patent application underlying the referral is EP03793825.5
18. See also "Zur Patentierung von Entwurfs- und Simulationsverfahren in der EPA-Rechtsprechung", Rainer Moufang, GRUR Int. 2018, pages 1146ff (the author is also the legal member of the referring Board in case G1/19)

19. On the interrelationship, at least for question of patentability, between computer simulations and AI, see e.g. section II of "Software and Artificial Intelligence – Old and New Challenges for Patent Law – Conference Report on the 3rd Binational Seminar of the TU Dresden and the Charles University in Prague, November 20, 2018", in GRUR Int. 6/2019, pages 560ff, wherein reference is expressly made to the Infineon decision when discussing whether patentability of AI has been already clarified by the current legal framework
20. Official minutes of the hearing are not available at the time of writing; hence, the present notes reflect the authors' views and impressions from the online attendance to the hearing
21. The representatives of appellant and the representatives of the President of the EPO, who is party to the proceedings
22. In the amicus curiae brief submitted by Siemens, the following is noted under (ii) on page 3: "With regard to the explosively increasing importance of software in […] the Internet and digitization, there is a serious risk that a criterion that is too traditional and without recognizable justification as "necessary" will exclude the entire field of digital future technologies from patent protection could be." (informative translation from German)
23. In "Framing new technical problems in AI inventions", by Rachel Free, CIPA Journal, October 2018, it is argued that there are a number of new technical problems arising from AI inventions, which are not properly reflected in the classic formulation of technical problems. It thus seems important adapting existing examination practice to reflect the central role of computers in the conception of new inventions
24. In "Autonomous Machines and their Inventions", Ryan Abbott, Mitteilungen der deutschen Patentanwälte, October 2017, pages 429ff, the author addresses the scenatio where the invention is directly conceived by the computer and argues that "creative computers require a rethinking of the criteria for inventiveness, and potentially of the entire patent system"

IPR Risk Assessment of Companies Implementing Standard Essential Patents and Validation

Kazuto Kobayashi[1](✉), Mitsuyoshi Hiratsuka[2], and Michele Baccelli[3]

[1] Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan
kobayashi@sangaku.titech.ac.jp
[2] Tokyo University of Science, Shinjuku-ku, Tokyo, Japan
hiratsuk@kb3.so-net.ne.jp
[3] Hoffmann Eitle, Munich, Germany
mbaccelli@hoffmanneitle.com

Abstract. An essential patent or standard-essential patent (SEP) is a patent that claims an invention used to comply with a technical standard. The patent policy of the standards setting organization specifies a condition called FRAND as licensing terms and conditions for SEP. However, it has been difficult for companies that adopt standards to grasp the total royalties, because the specific royalties of FRAND are usually not defined in the patent policy of the standards setting organization. The potential royalties payment may thus represent an intellectual property risk to those companies that implement technology subject to standardization. In this paper, we derive a formula for calculating the intellectual property risk of the implementer of SEP by modeling SEP holders. In addition, the IPR risk of IEEE 802.11 is calculated and the result of the calculation is validated. Disclaimer. While the present paper has been prepared with HCC14, The authors are not liable for any direct, indirect, special or consequential losses or damage of any kind, or loss of profit, derived from using any of the information contained in the present paper.

Keywords: Patent · Technical standard · SEP · Royalty · FRAND

1 Introduction

An essential patent or standard-essential patent (SEP) is a patent that claims an invention used to conform to a technical standard. Many standards setting organization (SSO) have established a patent policy that deals with the SEP when developing standards. If a member of an SSO considers that it owns a SEP, that member is required to declare its intention to grant a license free of charge or under reasonable and non-discriminatory (FRAND) conditions in other words, such a (FRAND) declaration implies that the SSO member is ready to allow other parties to use its invention, if those other parties implement the standard technology and comply with certain conditions. The details of the patent policy are different depending on each SSO. While many if not all refer to FRAND terms and conditions, none of them specifies the amount of the

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 77–85, 2020.
https://doi.org/10.1007/978-3-030-62803-1_7

license fees. Usually, SSOs also do not determine whether a standard is covered by the alleged SEP.

After a standard is established, many SEP are licensed in patent pools formed by multiple patent holders [1]. A patent pool is a system where a number of SEP holders concentrate their SEPs and collectively license those rights to third parties and/or to members of the pool. However, some companies do not participate in patent pools but engage in patent licensing activities on their own, by engaging in a number of bilateral negotiations and bilateral licensing programs with other users of the standard [2]. There have been a number of cases where SEP holders outside of the patent pools file infringement lawsuits against companies that sell products conforming standard specifications and demand high royalties. It cannot also be excluded that patent holders have been filed lawsuits following failed negotiations under a patent pool [3–6].

For this reason, it is desirable for companies who incorporate certain standards into their products to identify the total royalties as an IPR risk, which is the sum of royalties of patent pools and royalties that may be paid in the future to companies outside the patent pools. However, it is difficult to grasp the total potential royalties, because it is not clear nor well-established how much the royalty per SEP is, whether all patents should be entitled to the same royalty rate and what the total number of SEP is. One reason lies in the difficulty for SSOs to commit to an amount of royalty rates under FRAND due to regulations under the anti-trust law. In order to calculate the total number of SEP, it is also necessary for experts to determine the essentiality of SEP candidates after screening all candidates including those shown on SSO databases and those that are not recorded in such databases.

2 Research Subject

Since the total royalties payable for SEP is the license fee per SEP multiplied by the total number of SEP, the difficulty of estimating the total royalties may be addressed by separately considering the respective causes of royalty per SEP and the total number of SEP.

First, it is unclear how much the royalty per SEP is. This is because the definition of FRAND in the patent policy of the SSO is not clear nor given. However, if an SSO clarifies the licensing fees of standards, it may raise doubts to violate the anti-trust law. In relation to some standards, the agreement of the major SEP holders on the total royalty for SEP is expressed, as made public by press releases; in some cases, including Microsoft v. Motorola (U.S. District Court, 2013) and TCL v. Ericsson (U.S. District Court, 2017), courts decided the royalty for each essential patent by adopting total royalty announced and the top-down approach. However, many SSOs do not specify such an agreement.

Secondly, it is difficult to calculate the total number of SEP for any standard. The number of SEP owned by patent pool licensors is usually provided by patent pools. The number of SEP owned by a patent holder who has made a FRAND declaration may be calculated from the patent information published in databases of FRAND declarations of SSOs. However, it is not easy to determine the number of other SEP because this requires an expert to evaluate on the essentiality of all possible granted patents.

3 Calculation Hypotheses for Essential Patents

Based on the research subject illustrated in Sect. 2, in order to utilize this approach for the calculation of IPR risk as a total royalty, we will extract major cases concerning patent policies and FRAND declarations of SSOs and examine the method of calculating the total number of SEP and royalty per patent.

3.1 Example of Trial Calculation of Patent Fees Based on Patent Pools

Microsoft v. Motorola case (U.S. District Court, 2013) is the first trial in the world that assessed the royalties for SEP. In this case, the benchmark was calculated based on the ratio of the number of patents in the patent pool to the number of Motorola's patents that is multiplied by the royalty fee of the patent pool. In this case, the license fee for the patent pool was pointed out to be lower than the average license fee in the market.

3.2 Comparative Approach

In the United States, the Georgia-Pacific factor is known as a criterion to calculate reasonable royalties. In Microsoft v. Motorola case (U.S. District Court, 2013), the application of Georgia-Pacific factor to SEP was considered, and the licenses of the case on a similar background could be referenced, and multiple licenses, including patent pools, were adopted as a benchmark for royalties. In Unwired Planet v. Huawei case (United Kingdom High Court, 2017) comparable licenses were selected and royalty rates were calculated as a relative number ratio of patents. CSIRO v. Cisco case (United States, CAFC, 2015) also adopted a comparative approach for a license fee for SEP. Thus, if the total number of SEP is estimated and a comparative approach is used to the known royalties, such as patent pools, then it is possible to calculate royalties for the total number of SEP.

3.3 Calculating the Total Number of Essential Patents

In calculating the total number of SEP, it is necessary to estimate the number of patents of SEP holders who are involved in standardization but do not participate in patent pools, or who are not involved in standardization. In Dell case (FTC 1996) the standard for personal computer Bus was developed in VESA and Dell's failure to disclose a patent was the issue. In particular, Dell exercised its patent right after VESA Bus became commercially viable. FTC ruled that Dell's actions violated Sect. 5 of FTC Act. In Rambus case (FTC 2008), FTC charged that the company violated the disclosure requirements of its patents.

Then, cases of violations of the FRAND commitments have subsided, and the duty of good faith as a contract has come to be discussed in the point of view of FRAND in subsequent infringement trials. Therefore, SEP holders are now well aware that if they intentionally evade the FRAND declaration, they would not be permitted to exercise their rights with high chances of success. In other words, no SEP holders seem to evade the FRAND declaration on purpose. At least, for the purpose is an estimation of the IPR, it would give a little impact on the calculation to exclude SEP holders who evaded the FRAND declaration on purpose. We can therefore summarize the hypothesis regarding the royalty and total number of SEP as follows:

- If a patent pool exists for a standard, the patent pool royalty can be used as the basis for the royalty per patent in calculating the IPR risk.
- The use of the comparison approach allows the calculation of the total license fee for SEP by multiplying the ratio of the number of patents in the patent pool to the number of all SEP by the license fee for the patent pool.
- When totaling the number of SEP for the purpose of estimating the IPR risk, there is little impact on the calculation even if the number of patents of SEP holders who evaded the FRAND declaration on purpose is excluded.

4 SEP Holders Model

4.1 Attribute Information

Prior to tabulating the total number of SEP based on the hypothesis in Chapter 3, the following attributes are extracted for SEP holders, in relation to their activity in the standardization and the patent pools.

- Participation in standardization: Many SEP holders participate in SSO committee. It is difficult to make inventions resulting in SEP without getting in touch with the drafts and discussions in the SSO committees.
- FRAND declarant: Among the participants in SSO committees, SEP holders have made FRAND declarations to comply with the patent policy of SSO.
- Licensor in patent pools: Some of FRAND declarants are licensor in patent pools.
- Licensee in patent pools: Some FRAND declarants are licensee and do not license in patent pools.

4.2 Hypothesis SEP Holders Model

Based on the attribute information above, SEP holders are classified into the following models 1 to 6, when considering the aim of calculating intellectual property risks [7].

1. Pure licensors: Patent holders who participate only as licensor in patent pool. Licensors in patent pools are actual SEP holders and are required for calculation.
2. Cross-licensors: Patent holders who participate as Licensors and licensees in patent pool. As with pure licensors, they are actual SEP holders and are essential for calculation.
3. Pure licensees: Patent holders who participate only as licensees in patent pool. FRAND declaration was done but the patent might not finally cover the standard. Since their patents are not licensed in patent pool, they are excluded in the calculation.
4. Non-participants in patent pool: Patent holders who have declared FRAND and have not participated in patent pool. This entity may be engaged in patent licensing activities alone and are required for calculation.

5. FRAND non-declarants: Patent holders who participate in standardization committee and have obtained SEP,but have not done FRAND declaration and not have participated in any patent pool. FTC prosecuted those who evaded the obligation of the FRAND declaration on purpose in the past. Now there does not seem to be such SEP holders, and they are excluded from the calculation.
6. Non-participants in standardization: Patent holders not involved in standardization activities. Although there may be researchers who have made basic inventions outside of SSOs, this number is considered to be small in comparison with the number of licensors of patent pools. So they are excluded in calculation.

Therefore, SEP holders model to be considered in calculating IRR risk are 1, 2, and 4.

5 Risk Calculations for Patent Implementers

5.1 Defining Variables

IPR risk is the total royalty that a company must pay to sell a product that incorporates a standard. In deriving the formula for calculating IPR risk, Ri, variables are defined as follows.

Ro_n: fee for patent pool n (Dollars)
P_n: number of patents licensed in patent pool n,
AR: Average royalty in patent pools (Dollars)
NP: Number of patents of non-participants in patent pool (corresponds to Model 4 in the previous section)
Ri: IPR Risk
N: Total number of patent pools
NPk: Number of patents owned by non-participants in patent pool with known number of patents
NPu: Number of patents owned by non-participants in patent pool with unknown number of patents
MNPu: Number of non-participants in patent pool with known number of patents
ANPk: Average number of patents owned by patent holders with known number of patents

5.2 Considering Only One Pool of Patents

Based on the calculation hypothesis and SEP holder classification model and variables presented in the previous sections, IPR risk is calculated with N = 1. Then, IPR risk is calculated by multiplying the relative ratio between the sum of the number of patents P_1 and the number of patents NP to the number of patents P_1 by the royalty Ro_1.

$$\mathrm{R}i = \mathrm{Ro}_1 \bullet (\mathrm{P}_1 + \mathrm{NP})/\mathrm{P}_1 = \mathrm{Ro}_1 \bullet (1 + \mathrm{NP}/\mathrm{P}_1) \tag{Eq1}$$

5.3 Considering Multiple Patent Pools

When there are multiple patent pools (N), the average royalty AR is derived by averaging the royalties of each patent pool with the weigh of the number of patents held (Eq 2), Ro_1 of the (Eq 1) is replaced by AR, and P_1 of (Eq 1) patent pool is replaced by the sum of the number of all N patents P_n, expanding (Eq 1) to the formula to calculate Ri that constitutes the IPR risk (Eq 3).

$$\mathrm{AR} = \sum\nolimits_{n=1}^{\mathrm{N}} (\mathrm{Ro}_n \bullet \mathrm{P}_n) / \sum\nolimits_{n=1}^{\mathrm{N}} \mathrm{P}_n \tag{Eq2}$$

$$\mathrm{Ri} = \mathrm{AR} \bullet \left(\sum\nolimits_{n=1}^{\mathrm{N}} \mathrm{P}_n + \mathrm{NP}\right) / \sum\nolimits_{n=1}^{\mathrm{N}} \mathrm{P}_n = \mathrm{AR} \bullet \left(1 + \mathrm{NP} / \sum\nolimits_{n=1}^{\mathrm{N}} \mathrm{P}_n\right) \tag{Eq3}$$

5.4 Estimation of NP When the Number of Patents Held by Non-Participants in Patent Pool Is Unknown

If the patent number is not disclosed in a SSO that permits the “blanket” FRAND declaration, NP cannot be calculated only from the information of the patent number disclosed in the FRAND declaration. In this case, NP are estimated by the sum of the total number of patents NPk of patent holders with known number of patents and the total number of patents NPu of persons with unknown number of patents. Then NPu is estimated by multiplying the average number of patents owned by patent holders with known number of patents ANPk by the number of non-participants in patent pool with known number of patents MNPu. This procedure finally leads to (Eq 4):

$$NP = NPk + NPu = NPk + ANPk \cdot MNPu \tag{Eq4}$$

6 Standard 802.11 IPR Risk Calculations

6.1 Standards: 802.11

802.11 is an international standard for wireless LAN (Wi-Fi) developed by IEEE, like for instance 802.11b which was the most popular in the early stage, improved then by versions 802.11a, 802.11n, etc. (Hereinafter collectively referred to as 802.11 including the improved versions). Although there are many products on the market, licensing of SEP in patent pools has not been so active, and there is no agreement on the total license fee. The FRAND Declaration Database (LOA) of IEEE discloses SEP patent-number and other information, and the number of patents can be counted. However, some SEP holders have not disclosed the patent-number. Subsidiaries and affiliates will be counted as a single corporation. The published patents are also calculated in the total number [8].

6.2 Patent Pool – Via Licensing

Via Licensing is a patent pool covering information and communications technology, and licenses 802.11 SEP in its program. The standard rate is $0.55 per unit, which is used as Ro_1 to calculate [9]. Via Licensing doesn't disclose the patent-numbers of 3 licensors, so we estimate such numbers. For NTT we adopt the number of patents disclosed in the FRAND Declaration. For LG we adopt the number of patents that LG license in Sisvel. For ETRI we adopt the number of patents that ETRI has licensed in Sisvel in 2015. As a result, 3 patent holders were found to have licensed 272 patents.

6.3 Patent Pool – Sisvel

Sisvel is a European headquartered patent pool that licenses 802.11 SEP. The patent- is available on their website [10]. The standard rate is €0.3 per unit, so we calculate Ro_2 as $0.34. The transferees of patents shall be regarded as FRAND declared entities, etc., and shall be counted. As a result, 8 parties have licensed $P_2 = 469$ patents.

6.4 Calculating IPR Risk

Substituting $P_1 = 272$, $P_2 = 469$, $Ro_1 = 0.55$, and $Ro_2 = 0.34$ into (Eq 2) yields an average royalty rate AR = $0.41. To Fix NP of (Eq 4), one candidate of ANPk is the average number of Sisvel licensors 59 and one other candidate is the average number of non-participants in patent pool who disclose their patent-numbers as being 14. Because there is a wide gap between the two candidates, both shall be adopted as the upper and lower bounds of the range, and IPR risk shall be calculated as the range. Then, substituting ANPk = 14–59, NPk = 470, and MNPu = 55 into (Eq 4), NP = 1240–3715.

Consequently, NP = 1240–3715, AR = 0.41, $P_1 = 272$, and $P_2 = 469$ are substituted for (Expr3) to derive IPR risk Ri. However LG's patents are overlapping in the two patent pools, so the sum of P1and P2 shall be reduced by the number of LG patents 131. As a result, Ri is estimated as $1.26–2.96.

7 Validation

7.1 Validation Approach

The IPR risk of 802.11 patent licensees $1.26–2.96 calculated in the previous chapter is validated by information such as royalties in actual cases that have not been referenced in the calculation step. VRo is defined as the royalty per SEP in the case, VPt is the expected total number of SEP, and VRi is defined as IPR risk for validation. By multiplying VRo by VPt to derive VRi (Eq 5), the range of the IPR risk for validation VRi for 802.11 is determined from multiple cases and the calculated IPR risk is validated.

VRi: Validation IPR Risk
VRo: royalties per SEP in a case
VPt: Expected total number of SEP

$$VRi = VRo \cdot VPt \tag{Eq5}$$

7.2 Expected Number of 802.11 Essential Patents

There is no authorized or official estimation of the total number of 802.11 SEP. In re Innovatio (U.S. District Court, 2013) it was examined the reports from plaintiffs and concluded that there were approximately 3000 patents. The accuracy of this numerical value is not high; however, it is enough to use for the purpose of the validation. So VPt is determined as 3000.

7.3 Example of Calculation of License Fee in Actual Cases

Microsoft v. Motorola (U.S. District Court, 2014) held Motorola's royalties of 802.11 SEP for 11 patents to be 0.8 cents ($0.008), which leads to $0.00072 per patent. In re Innovatio (U.S. District Court, 2013) held the license fee for Innovatio's SEP to be 9.565 cents ($0.0956) for 19 patents. This leads to $0.005 per patent. Thus, $0.00072 and $0.005 are adopted as the lower and upper limits of the range of royalty per patent VRo.

7.4 Validation

VPt and VRo derived in the previous section are input into (Eq 5) and the range of the IPR risk VRi for validation comes to be $2.16–15.0. Comparing the $1.26–2.96 of Ri from (Eq 3) with the $2.16–15.0 of VRi range, then the matched range is only $2.16–2.96.

One of the reasons for the difference is that the patent pool fees used, as the basis for calculation is lower than market rates generally. This is confirmed in Microsoft v. Motorola (U.S. District Court, 2014). On the other hand, only 2 cases were used in the validation, so the validity of the range of validation would not be statistically sufficient. We conclude that the risk of intellectual property under Standard 802.11 is $2.16–2.96 per unit, and we consider the reliability of the formulation of validation to be a future issue, including validation by other judicial examples.

8 Closing Comments

The IPR risk calculation of the implementer of SEP is derived. We also calculated 802.11 IPR risk (Eq 3) and validated it by (Eq 5) to find that the matched range is $2.16–2.96 per unit. The reliability of the formulation of the validation is a future problem, because calculated range does not always match to the range of the validation. We would like to express our gratitude to Akihiko Ohwada, Takeshi Misawa, Hideo Koike, and others of the Next-Generation Patent Platform Study Group for their discussion.

Disclaimer. While the present paper has been prepared with HCC14, The authors are not liable for any direct, indirect, special or consequential losses or damages of any kind, or loss of profit, derived from using any of the information contained in the present paper.

References

1. Kato, H.: Patent Pool Summary, Hatsumei-kyokai (2006)
2. Oyamada, K.: Trends in IT Pro-patent/Anti-patent, Impress R & D (2016)
3. Fujino, J.: Patent and Technical Standard: Cross Case and Legal Relationship, Hassaku-sha (1998)
4. Wakui, M.: Legal System for Technical Standards, Shoji-Homu (2010)
5. Kobayashi, K.: Case report on RAND royalty rates for standard-essential patents. Patent **67** (7), 46–57 (2014)
6. Kamiike, M., Kobayashi, K., Hiratsuka, M.: A study on the calculation of the license fees for standard essential patents in the FRAND case. Patent **68**(119), 10–133 (2015)
7. Kobayashi, K.: Risk calculation of patent holder model of essential patents of technical standards for patent implementers to contribute to technology management, Institute for the Synergy of Arts and Sciences, vol. 23 No.1 (2019)
8. IEEE-SA RECORDS OF IEEE STANDARDS-RELATED PATENT LETTERS OF ASSURANCE. https://standards.ieee.org/about/sasb/patcom/patents.html
9. Via Licensing 802.11 (a-j) Program: http://www.via-corp.com/jp/ja/licensing/ieee-80211/overvi ew.html
10. Sisvel Wi-Fi Joint Licensing Program: http://www.sisvel.com/licensing-programs/wireless-communications/wi-fi/introduction

Qualitative Analysis of Interviews with Municipal Officers Toward the Human-Centered Improvement of the eLen Regulation Database System

Aki Shima[1](✉) and Tokuyasu Kakuta[2]

[1] The Institute of Education and Student Affairs, Niigata University,
Niigata, Japan
shimaaki@me.com
[2] Faculty of Global Informatics, Chuo University, Tokyo, Japan
kaku@tamacc.chuo-u.ac.jp

Abstract. The objective of this study is to analyze results of interviews that we conducted with Japanese municipal officials who have engaged in legislative drafting and to present main issues addressed by the interviewees. Moreover, based on the results of analysis, it aims to clarify remaining problems that legislators face with during legislation, which will be necessary conditions for expanding and improving human-centered functions of e-legislation systems. Using qualitative analysis of interviews with municipal officers, this paper identifies the following four issues addressed by interviewees: (1) inconsistency among ordinances; (2) inconsistency of an ordinance; (3) insufficiency in consideration of legislative objectives and facts; (4) inadequacy of legal research. Based on the results of interview analysis, this study clarifies whether the eLen regulation database system copes with them and discusses remaining problems. Overall, it illustrates that some functions included in the eLen are helpful for diminishing those issues. However, in order to overcome the problems with which legislators face in the process of legislation, the results of this study show that it is significant to provide legislators education, such as trainings for the way to use the system and benchmarking method or for learning legislation process. Although the eLen has already implemented several instructive mechanisms, we will improve further the system so that users can learn the proper process of legislation through the usage of the system.

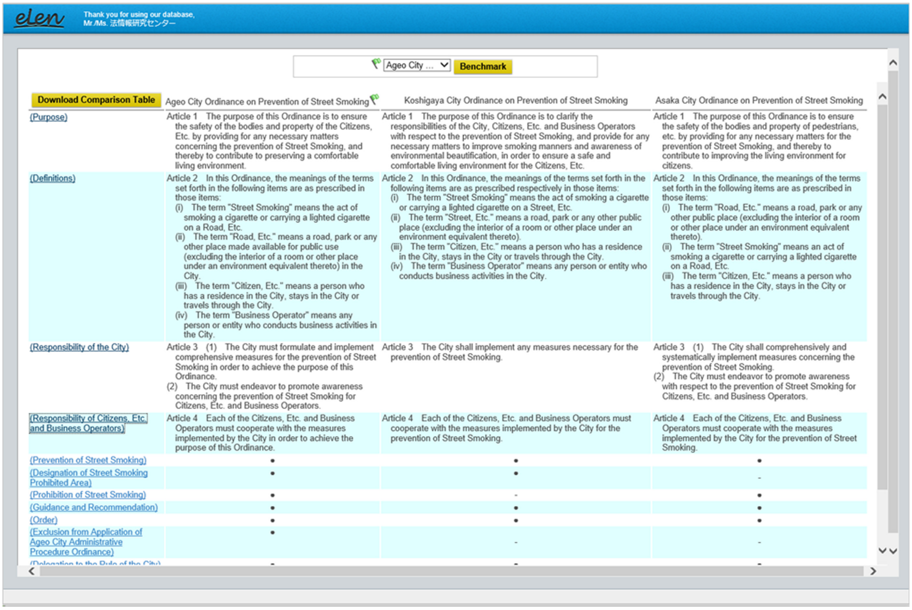
Keywords: Legislation · e-Legislation · Regulation database · Municipality · Local government · Qualitative analysis · Interview

1 Introduction

The authors have been developing and operating the eLen regulation database system to support legislation in municipalities. This system has a built-in database that covers more than 90% of all local governments (about 1790) in Japan and has been used by many municipalities since 2013. Moreover, it realizes automatic creation of “Benchmarking tables” for comparing regulations enacted by different municipalities, which

D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 86–99, 2020.
https://doi.org/10.1007/978-3-030-62803-1_8

are created by automatic clustering of similar regulations [1]. The screenshot of this eLen is shown in Fig. 1.



Download Comparison Table	Ageo City Ordinance on Prevention of Street Smoking	Koshigaya City Ordinance on Prevention of Street Smoking	Asaka City Ordinance on Prevention of Street Smoking
(Purpose)	Article 1 The purpose of this Ordinance is to ensure the safety of the bodies and property of the Citizens, Etc. by providing for any necessary matters concerning the prevention of Street Smoking, and thereby to contribute to preserving a comfortable living environment.	Article 1 The purpose of this Ordinance is to clarify the responsibilities of the City, Citizens, Etc. and Business Operators with respect to the prevention of Street Smoking, and provide for any necessary matters to improve smoking manners and awareness of environmental beautification, in order to ensure a safe and comfortable living environment for the Citizens, Etc.	Article 1 The purpose of this Ordinance is to ensure the safety of the bodies and property of pedestrians, etc. by providing for any necessary matters for the prevention of Street Smoking, and thereby to contribute to improving the living environment for citizens.
(Definitions)	Article 2 In this Ordinance, the meanings of the terms set forth in the following items are as prescribed in those items: (i) The term "Street Smoking" means the act of smoking a cigarette or carrying a lighted cigarette on a Road, Etc. (ii) The term "Road, Etc." means a road, park or any other place made available for public use (excluding the interior of a room or other place under an environment equivalent thereto) in the City. (iii) The term "Citizen, Etc." means a person who has a residence in the City, stays in the City or travels through the City. (iv) The term "Business Operator" means any person or entity who conducts business activities in the City.	Article 2 In this Ordinance, the meanings of the terms set forth in the following items are as prescribed respectively in those items: (i) The term "Street Smoking" means the act of smoking a cigarette or carrying a lighted cigarette on a Street, Etc. (ii) The term "Street, Etc." means a road, park or any other public place (excluding the interior of a room or other place under an environment equivalent thereto). (iii) The term "Citizen, Etc." means a person who has a residence in the City, stays in the City or travels through the City. (iv) The term "Business Operator" means any person or entity who conducts business activities in the City.	Article 2 In this Ordinance, the meanings of the terms set forth in the following items are as prescribed in those items: (i) The term "Road, Etc." means a road, park or any other public place (excluding the interior of a room or other place under an environment equivalent thereto). (ii) The term "Street Smoking" means an act of smoking a cigarette or carrying a lighted cigarette on a Road, Etc. (iii) The term "Citizen, Etc." means a person who has a residence in the City, stays in the City or travels through the City.
(Responsibility of the City)	Article 3 (1) The City must formulate and implement comprehensive measures for the prevention of Street Smoking in order to achieve the purpose of this Ordinance. (2) The City must endeavor to promote awareness concerning the prevention of Street Smoking for Citizens, Etc. and Business Operators.	Article 3 The City shall implement any measures necessary for the prevention of Street Smoking.	Article 3 (1) The City shall comprehensively and systematically implement measures concerning the prevention of Street Smoking. (2) The City must endeavor to promote awareness with respect to the prevention of Street Smoking for Citizens, Etc. and Business Operators.
(Responsibility of Citizens, Etc. and Business Operators)	Article 4 Each of the Citizens, Etc. and Business Operators must cooperate with the measures implemented by the City in order to achieve the purpose of this Ordinance.	Article 4 Each of the Citizens, Etc. and Business Operators must cooperate with the measures implemented by the City for the prevention of Street Smoking.	Article 4 Each of the Citizens, Etc. and Business Operators must cooperate with the measures implemented by the City for the prevention of Street Smoking.
(Prevention of Street Smoking)	•	•	•
(Designation of Street Smoking Prohibited Area)	•	•	-
(Prohibition of Street Smoking)	•	-	•
(Guidance and Recommendation)	•	•	•
(Order)	•	•	•
(Exclusion from Application of Ageo City Administrative Procedure Ordinance)	•	-	-
(Delegation to the Rule of the City)	-	-	-

Fig. 1. Benchmarking table

There is previous research that introduces the eLen database system with the functions illustrated above [2], which was proposed and created by the authors based on interviews with many local government officers in Japan [3]. The salient issues raised by those interviewees concern preparation for legislative work, such as investigation and comparison of existing regulations. It is also pointed out that collecting similar regulations and producing tables for comparison among regulations are ad hoc, intuitive and time-consuming. The most obvious finding of the previous research that emerges from the interviews is that all respondents answer that they have enacted regulations by referring to precedent and similar regulations, and sometimes regarding them as models. Additionally, an analysis of regulation data as well as operational results of the eLen are also indicated in the paper above [2]. However, a detailed analysis of those interviews conducted by the authors was not made when the eLen database system was designed, and only functions that were frequently requested by municipal officers who have engaged in legislation were realized. Therefore, the previous research indicates that the eLen covers the outstanding needs of municipal officers, but questions remain as to whether these functions alone could make legislators' work easy and provide sufficient human-centered legislative support.

Furthermore, the authors' study has been conducted as part of "legal engineering" [4] and "e-legislation" [5] studies, aimed at "applying information science and software engineering to laws in order to support legislation" (p. 322) [4] as these laws can be regarded as specifications in society. In this paper, the term “e-legislation” is used to refer to introduction of IT and ideas of information science to legislative work as well as rulemaking in general. Since much information and legislative work process will be accumulated as intermediate products and the history of an e-Legislation process in the form of digital data, those products of legislative work can be visualized and organized more objectively. Thus, e-Legislation can make legislation process more precise and more efficient. For instance, it would be able to help people who engage in legislation discover new issues and identify mistakes.

Moreover, not only can e-Legislation contribute to streamlining of administration in Japanese municipalities, but it also has the potential to export the e-Legislation system itself to other countries. The data that are stored as intermediate products of e-Legislation are not only the one such as texts of articles and proceedings of legislation process, but also structures of policies and rule description methods that are formulated as abstract models in the e-Legislation process. In other words, e-Legislation intends to develop the data that include semantic structure instead of superficial text information. The study that handles such semantic information was born in the 1970s in the field of artificial intelligence, developed as “knowledge engineering,” and has been called “ontology” since the mid-1980s. This research aims to make use of the data that are called “knowledge” in such fields [5].

Thus, the objective of this study is to analyze results of interviews that we conducted with Japanese municipal officials who have engaged in legislation in local governments. Moreover, based on the results of analysis, it aims to clarify remaining problems in the process of legislation, which will be necessary conditions for pursuing ways to coordinate and integrate interface between human and information technology (IT). Since there is no other previous research on needs surveys of Japanese municipal officers regarding legislation except our research [3], this paper will use the interview data gathered through the previous research.

The paper has been organized in the following way. The second section is concerned with the methodology used for this study (Sect. 2). In the Sect. 3, first, an overview of the legislation situation as well as the common process of legislation in local governments in Japan will be given (3–1). There are two types of ordinance legislation in municipalities: new enactment and partial revision of ordinances. This paper will not deal with the latter as there is already support system provided by private companies. In the next part of the Sect. 3 (3–2), it will identify issues addressed by interviewees who have engaged in legislative work. The Discussion section will assess whether the functions included in the eLen regulation database and its extensions cope with those issues and clarify the remaining issues that will be expected for human-centered improvement of the eLen regulation database system (4). Finally, this study will conclude with a brief summary (5).

2 Methodology

The study uses qualitative analysis of interviews conducted by authors in order to gain insights into issues addressed by officials in Japanese municipalities in the legislation process. In-person and semi-structured interviews with municipal officers who belong to a division of legislation were implemented by authors in all local governments (19 cities, 13 towns and a village[1]) in Kanagawa prefecture in Japan. Additional in-person and phone interviews were conducted to reinforce the needs surveys. Moreover, a short questionnaire was designed to ascertain the participants' ways of legislative work. There are 66 notes taken during and after interviews as well as 33 answers of questionnaires collected from September of 2011 to May of 2014.

The reason we selected neither members of assemblies nor lawyers but local government officials as interviewees is that there is a situation in Japan that officials usually prepare for legislative work and engage in legislative drafting. For instance, in a survey conducted by the National council of Municipal Councils, the total number of ordinances submitted by members of assemblies in all cities (814) in Japan was 687 during the year from January 1, 2017 to December 31, 2017. It means that only 47% of the cities (386) have submitted legislative drafts [6]. In addition, the average number of submissions in a city where there was a case of legislative proposals was 1.8. Looking at this in all local governments, including prefectures, cities and towns, ordinances proposed by the head of municipalities occupy about 85% of the total [7]. Moreover, even though regulations are drafted by members of assemblies, those who help the members with legislation as staff in Assembly Secretariat are also local government officials in Japanese municipalities. Moreover, we chose officers in a division of legislation as interviewees, since they deal with a wide range of legislation across many divisions in a municipality and that all drafts of ordinances and regulations are scrutinized by this division.

After coding transcribed texts of 66 notes, qualitative analysis of the data was conducted to identify the issues addressed by interviewees regarding legislation in municipalities. Further data collection is required, but as there are no other qualitative data of interviews with Japanese municipal officers regarding legislation process, the survey run by authors would be useful sources to know needs of legislators when the e-legislation support system is developed.

3 Legislation in Local Governments

3.1 Background of Legislation in Local Governments

It is provided that local governments shall have the right to manage their property, affairs and administration, and to enact their own regulations as far as laws and regulations are not violated (Article 94, The Constitution of Japan). Thus, each municipality in Japan, the total number of which is 1772 in 2020, has established ordinances

[1] The respondents of surveys are as follows: City of Yokohama, Kawasaki, Sagamihara, Yokosuka, Hiratsuka, Kamakura, Fujisawa, Odawara, Chigasaki, Zushi, Miura, Hatano, Atsugi, Yamato, Isehara, Ebina, Zama, Minamiashigara, Ayase, Town of Hayama, Samukawa, Ooiso, Ninomiya, Nakai, Ooi, Matsuda, Yamakita, Kaisei, Hakone, Manazuru, Yugawara, Aikawa and Village of Kiyokawa. Although an interview and questionnaire were also conducted with the Kanagawa Prefectural Government during the same survey period, the results have been excluded due to many differences between prefectures and other scales of municipalities. In addition, respondents' comments are not representative of the municipality's views, but their opinions. The municipality's names are hidden in this paper, since some of them would not like to disclose their names.

[8]. Ordinances are enacted, amended and abolished by decisions of assemblies. The right to propose these ordinances is given to both the head and members of an assembly of a municipality, but most of which are proposed by the former [6, 7].

It is the local government officials who play roles of drafting ordinances submitted by the head of municipalities. Members of a division in charge of a specific area of ordinances have a responsibility of drafting ordinances. Then, the drafts are passed to the division of legislation in order to scrutinize them carefully. After reviews by the legal division, they are submitted to the Assembly [9]. Although most of the staff in the divisions that have responsibility for ordinances do not have technical knowledge on drafting ordinances, Japanese ordinances are supposed to be written, following unique and detailed legislative drafting rules as well as using specific language in accordance with laws and regulations. These rules and terms are different from everyday language and are not easily learnt.

Thus, the role of the staff in the division of legislation is to help the staff who has to make legislative drafts and is not familiar with such complicated rules and manners of drafts. Specifically, the legislative divisions examine drafts in terms of violation or conflicts of laws, objectives of making new regulations, applicable structures of ordinances, usage of legal words, influences of the concerned ordinances on other ordinances and regulations in the municipality and so on [10]. However, in general, even legislative staff acquires knowledge and skills in legislative drafting through OJT. They are not also law professionals and, even those with many experiences in legislation usually have to move to other departments a few or several years later, just like any other staff members, so that new staff members need to be trained from the beginning.

Ordinances play a very important role in implementing policies in local governments and in setting rights and obligations of residents. Nevertheless, the staff who has responsibility for preparing ordinances is not necessarily fully experienced, especially in small municipalities where it is difficult to secure sufficient human resources. Such current situation could lead to overlook mistakes of ordinances after their enforcement.

Despite such difficult situations in human resources and training of staff in local governments, in Japan, there is a situation that local governments have been expected to formulate ordinances more spontaneously because of the movement from centralization to decentralization. The basic idea of this movement was that administrative services that are close to residents' daily lives should be managed by local governments, so that "the autonomy and independence of local governments would be enhanced and enable them to fulfill the responsibilities to develop unique and dynamic local societies" (Article 2, Decentralization Promotion Act, 1995). However, it is also pointed out that the number of ordinances that municipalities should enact individually has increased because of the decentralization, which has also led to increase burdens on officers as regards legislation.

3.2 Results: Situation of Legislation and Issues Addressed by Municipal Officials

In order to clarify the remaining problems that municipal officials face with during legislation, this paper will present four main issues ((1)–(4)) addressed by the interviewees who engaged in legislation process. According to interview surveys of the legislative section staff in local governments, the following two types of legislation were cited as cases where ordinances were drafted: revisions and enactment of ordinances. As mentioned above, this study will describe only the new enactment cases.

In the case of new enactment, municipal officers answered that they had mainly drafted ordinances in two patterns: by modeling standard examples to be followed if they are provided by the central government or prefectures, and by referring to the similar ones that were enacted by themselves in the past or by other precedent municipalities (C, L, Q City; A, D Town and many others).

When new ordinances are planned to enact in municipalities, all the respondents answered that they always referred to similar ordinances enacted by themselves or precedent municipalities, unless there are standard examples provided by the central government or prefectures. The Government used to provide such models of ordinances to municipalities, and local governments obeyed them before decentralization. However, from the viewpoint of decentralization and local autonomy, standard examples have been offered only on the limited matters. The following examples were indicated by respondents (F, K, M and Q City; E, F and J Town; A Village): ordinances on the police, firefighting, tax, allowances, national health insurance, officials' salary, all of which need to be enforced impartially among municipalities.

Issue: (1) Inconsistency among Ordinances. In the case of referring to standard examples provided by the central government or prefectures, an issue addressed by the respondents is that it is laborious for drafters to follow both the unique rules of legislative drafting and usage of legal words that each local government has decided by custom (G and K City; B and I Town).

> Each municipality traditionally has kept its own rules on how to draft ordinances. For example, even ordinances about establishment of public facilities are different in detail among municipalities. Overall, they are similar, but there are some different parts (B Town).

> When we refer to standard examples, I think there are a few problems. Take ordinances about salary, for example, there is a huge difference in prescribing among municipalities…(A Village).

Even when drafters intend to follow the standard examples, inconsistency among previous ordinances that were enacted by a municipality could cause. Therefore, officials have to tackle laborious amendment tasks to avoid inconsistency in terms of drafting rules and usage of legal words.

Issue: (2) Inconsistency of an Ordinance and (3) Insufficiency in Consideration of Legislative Objectives and Facts. Makise [8] suggests that municipal officials do not refer to standard examples under the current situation, but to precedent ordinances enacted by other municipalities when they consider legislative drafting [8]. In the interviews conducted by the authors, most of respondents, at first, hesitated to admit that they drafted ordinances by referring to the precedent ones produced by other municipalities. However, some remarks show that officials think this way of legislative drafting positively as follows:

> We have no choice but to prepare for drafts by referring to precedent ordinances, but it's not like just copy and paste. I think it would be allowed if we try to cut out a good part and make it our own (D Town).

> It is true that small towns cannot store know-how for legislation, or that officials don't have time to study at all. The staff of each division has to do various projects, often alone. The situation is the same in legislative drafting. So, after all, we couldn't help gathering similar ordinances of the same prefecture, comparing them, and picking up some articles (D Town).
> When you put a penalty in or out of an ordinance, considering which degree is applicable, honestly, it is not so easy to decide, so after all, it is safe if we follow the others (D Town).
> Looking at ordinances of other local governments, I sometimes notice the background or policies of the ordinances are different from my own, and I think that will also deepen the policies of my city (D City).

Despite many merits of the method mentioned above, other respondents raise some problems with this method: (2) the issue of inconsistency that caused by doing "patchwork" and (3) the issue of insufficiency in consideration of legislative objectives and facts. These issues are clearly indicated by an official of a city (D City) as follows:

> I think there are two problems. One is the problem of terms. For example, there are cases where it is not or difficult for even a term to be consistent in the same draft. As the same term could have a different meaning, the procedure or targets of support [provided by the ordinance][2] could change according to the meaning. Also, if a city defines a term, but the other doesn't, and then the draft is created by copy and paste, such a draft would be inconsistent.
>
> The other problem is there are "ordinances without soul." Namely, ordinances should be drafted based on aims. And then, methods to carry out the aims should be stated in the following articles. So, without considering aims, "doing patchwork" would create "ordinances without soul." But, after all, such things are not problems of a system, but those of users.

The second issue mentioned above in the quote is also addressed by several officials as follows:

> The persons in charge of legislative drafting in each division often explain about the legislative facts or contents of ordinances by intuition, so I always need to ask them to think reasons of legislation properly. But, I have trouble with this issue the most, what's more, it takes time the most (K City).
>
> It is often the case that drafters just think it's not a problem to copy and paste of precedent ordinances. There are many examples that drafters bring drafts to the legislative division and want us to check them. But, such drafts are frequently difficult to read. When I asked them the reason of that, they answered they copied and pasted good parts of ordinances enacted by this and that city. It happened a lot (B Town).

Issue: (4) Inadequacy of Legal Research. The last issue pointed out by many respondents is common in all the cases of new enactment of ordinances. It is the issue of inadequacy of legal research. Many respondents presented that searching relevant laws and ordinances of other municipalities is hard and an endless task that they spend a lot of time for (F, G, H Town and many others). As explained above, under the situation of decentralization, the new era began, in which municipalities can enact their own ordinances with the bounds of laws. It can be allowed that a municipality makes a regulative ordinance if there is rationality that the region of the municipality needs its regulation. Specifically, making ordinances is regarded as a range of municipal discretion under

[2] The sentence in a bracket was inserted by the authors for giving a supplementary explanation of the remark of the interviewee.

certain conditions, even if it regulates more strictly than national laws and regulations with the same objectives and restrictive methods (“Uwanose”), or even if it creates additional regulation standards to the national standard (“Yokodashi”). Thus, officials who mainly conduct legislative work in Japanese municipalities are required to consider whether the ordinance violates such rules very carefully in the process of legislation. Moreover, if new enactment including revisions of ordinances has effects on current regulations enacted by the municipality, they should be also revised simultaneously.

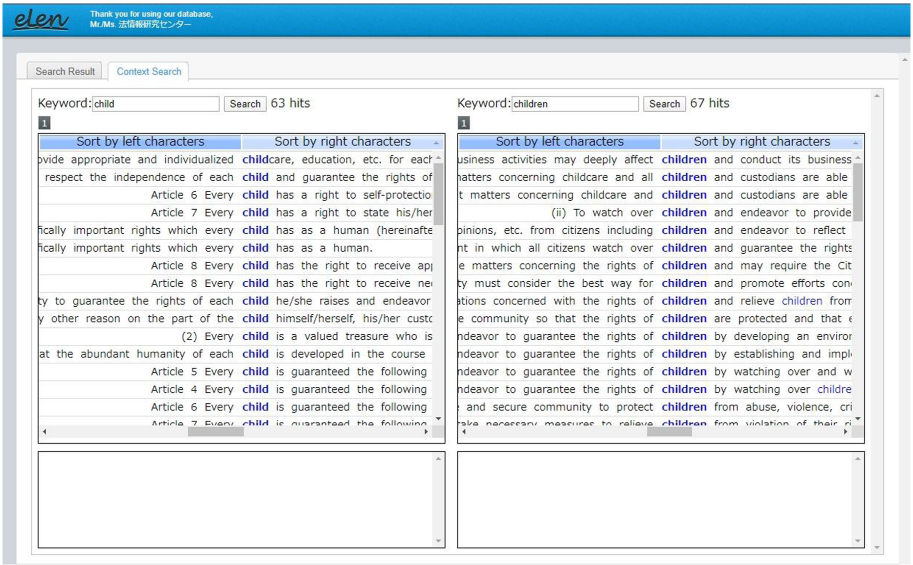
> When we make ordinances, it’s so difficult and troublesome in checking relevant laws and regulations. In short, “Yokodashi” and “Uwanose”. So, if the system checks such points, that will be great. I don’t know whether it is possible on the system or not. (B Town).

4 Discussion

The results mentioned in the previous chapter indicate that there are four main issues raised by municipal officials who have engaged in legislation. In this section, this paper will discuss possible responses to each issue. Particularly, it will identify whether functions included in the eLen regulation database system cope with those issues and clarify remaining problems.

4.1 Response to the Issues Addressed by Interviewees

Response to the Issue (1). A possible response to the laborious tasks that municipal officials need to tackle to avoid inconsistency in legislative rules and usage of legal terms is the function of “Context Searching” that the eLen provides as shown in Fig. 2.



Keyword: child Search 63 hits

1

Sort by left characters	Sort by right characters
ovide appropriate and individualized	**child**care, education, etc. for each
respect the independence of each	**child** and guarantee the rights of
Article 6 Every	**child** has a right to self-protectio
Article 7 Every	**child** has a right to state his/her
fically important rights which every	**child** has as a human (hereinafte
fically important rights which every	**child** has as a human.
Article 8 Every	**child** has the right to receive ap
Article 8 Every	**child** has the right to receive ne
ty to guarantee the rights of each	**child** he/she raises and endeavor
y other reason on the part of the	**child** himself/herself, his/her custo
(2) Every	**child** is a valued treasure who is
at the abundant humanity of each	**child** is developed in the course
Article 5 Every	**child** is guaranteed the following
Article 4 Every	**child** is guaranteed the following
Article 6 Every	**child** is guaranteed the following
Article 7 Every	**child** is guaranteed the following

Keyword: children Search 67 hits

1

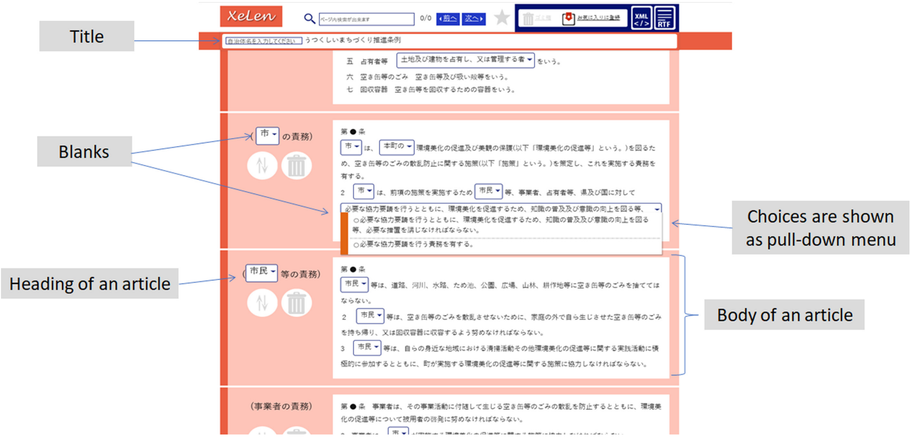
Sort by left characters	Sort by right characters
usiness activities may deeply affect	**children** and conduct its business
atters concerning childcare and all	**children** and custodians are able
t matters concerning childcare and	**children** and custodians are able
(ii) To watch over	**children** and endeavor to provide
pinions, etc. from citizens including	**children** and endeavor to reflect
nt in which all citizens watch over	**children** and guarantee the rights
e matters concerning the rights of	**children** and may require the Cit
ty must consider the best way for	**children** and promote efforts con
ations concerned with the rights of	**children** and relieve children from
e community so that the rights of	**children** are protected and that e
ndeavor to guarantee the rights of	**children** by developing an enviror
ndeavor to guarantee the rights of	**children** by establishing and impl
ndeavor to guarantee the rights of	**children** by watching over and w
ndeavor to guarantee the rights of	**children** by watching over childre
and secure community to protect	**children** from abuse, violence, cri
take necessary measures to relieve	**children** from violation of their ri

Fig. 2. Context searching

This function might be useful to check the manner of expressions and the way to prescribe regulations in a municipality if users narrow the searching range to the own municipal regulations.

Moreover, there are some types of ordinances that tend to enact repeatedly. For example, specific ordinances, such as those on establishment and management of public facilities, are required to be enacted under the law (Article 244-2, Local Autonomy Act). Every time when a new public facility is established in a municipality, it must make an ordinance on establishment of the facility in accordance with the law.

For those who do not have an experience in enactment and have difficulties in drafting even typical ordinances, an extended function of the eLen, which will create templates automatically as shown in the Fig. 3, would be helpful. The "template" function of the eLen[3] would enable drafters to enact ordinances more efficiently as they could start to consider the draft by referring to an example. In addition, drafters can create a standard sample of each municipality by narrowing targets of municipalities searched in the eLen to its own ordinances, so that it would make officials work more easily on checking consistency among ordinances enacted by the municipality before.



自治体名を入力してください うつくしいまちづくり推進条例

五　占有者等　土地及び建物を占有し、又は管理する者 をいう。
六　空き缶等のごみ　空き缶等及び吸い殻等をいう。
七　回収容器　空き缶等を回収するための容器をいう。

（市 の責務）

第●条
市 は、本町の 環境美化の促進及び美観の保護(以下「環境美化の促進等」という。)を図るため、空き缶等のごみの散乱防止に関する施策(以下「施策」という。)を策定し、これを実施する責務を有する。
2　市 は、前項の施策を実施するため 市民 等、事業者、占有者等、県及び国に対して
必要な協力要請を行うとともに、環境美化を促進するため、知識の普及及び意識の向上を図る等、
○必要な協力要請を行うとともに、環境美化を促進するため、知識の普及及び意識の向上を図る等、必要な措置を講じなければならない。
○必要な協力要請を行う責務を有する。

（市民 等の責務）

第●条
市民 等は、道路、河川、水路、ため池、公園、広場、山林、耕作地等に空き缶等のごみを捨ててはならない。
2　市民 等は、空き缶等のごみを散乱させないために、家庭の外で自ら生じさせた空き缶等のごみを持ち帰り、又は回収容器に収容するよう努めなければならない。
3　市民 等は、自らの身近な地域における清掃活動その他環境美化の促進等に関する実践活動に積極的に参加するとともに、町が実施する環境美化の促進等に関する施策に協力しなければならない。

（事業者の責務）

第●条　事業者は、その事業活動に付随して生じる空き缶等のごみの散乱を防止するとともに、環境美化の促進等について被用者の啓発に努めなければならない。

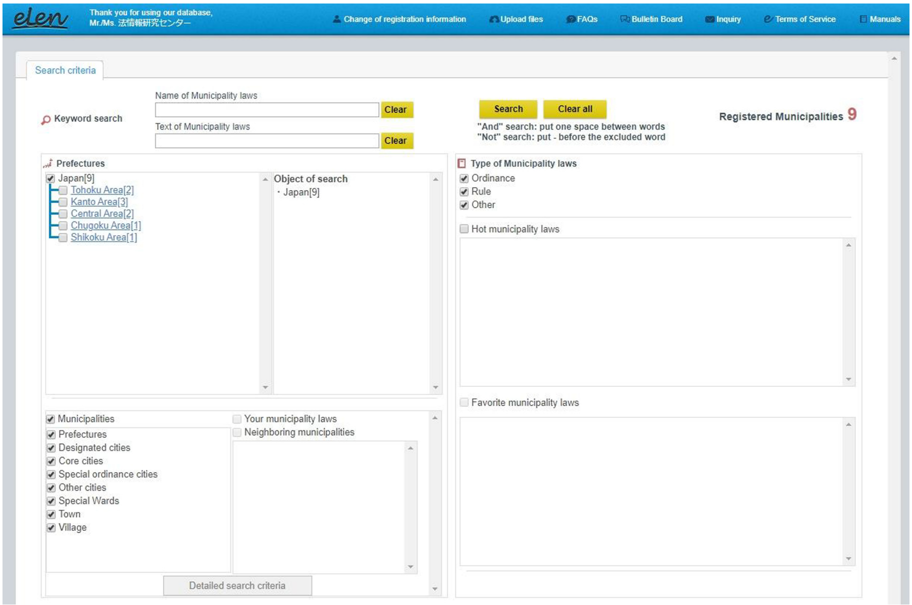
Fig. 3. GUI of a template in the extended version of the eLen

Response to the Issue (2) and (3). The eLen was developed based on the needs of officials in local governments, which were those of making laborious work on searching and gathering ordinances among different municipalities more efficient, and those of making time-consuming jobs on creating comparison tables among referring ordinances more easily. There was no tool by which enabled drafters to search

[3] The extended version of the eLen can automatically create a template by extracting common parts from similar regulations clustered by AI. In the template, different parts of regulations are shown as blanks and users can select one of choices to complete the regulation. If you change clustered groups, various templates can be automatically created by the eLen.

ordinances among different municipalities and to create comparison tables easily before the eLen was released to every municipality in Japan. Drafters used to do the task by looking for other municipal websites one by one, and to create a table by copying and pasting parts of ordinance texts.

After the release of the eLen, local government officials were released from many laborious tasks by using functions of the eLen, such as cross-searching, conditional search, clustering ordinances for display, and benchmarking table to compare some ordinances. Users can research regulations including ordinances enacted by approximately 1700 municipalities with key words. The results of searching can be narrowed down by putting additional conditions on attributes of municipalities, such as the size of municipalities, population, industry, and the area of municipalities as indicated in the Fig. 4. The list of the searching results also can be classified by the function of clustering according to the descriptive similarity. Moreover, the eLen provides a comparison table of each article that is lined up in accordance with the same heading as seen in the Fig. 1. This function of creating comparison tables automatically, which are named as "Benchmarking Tables," has acquired a high reputation from many legislative drafters (D and Q City; D Town and so on).



Thank you for using our database,
Mr./Ms. 法情報研究センター

"And" search: put one space between words
"Not" search: put - before the excluded word

Registered Municipalities 9

Prefectures
☑ Japan[9]
□ Tohoku Area[2]
□ Kanto Area[3]
□ Central Area[2]
□ Chugoku Area[1]
□ Shikoku Area[1]

Object of search
· Japan[9]

Type of Municipality laws
☑ Ordinance
☑ Rule
☑ Other

□ Hot municipality laws

□ Favorite municipality laws

☑ Municipalities
☑ Prefectures
☑ Designated cities
☑ Core cities
☑ Special ordinance cities
☑ Other cities
☑ Special Wards
☑ Town
☑ Village

□ Your municipality laws
□ Neighboring municipalities

Fig. 4. Cross searching

Originally, the "Benchmarking Method" was introduced as one for business improvement in 1990s [11]. "Benchmarking" is a method by which an organization finds targets to be referred ("Benchmark") and compares them with its own achievement in order to identify problems and points to be improved [12]. In the area of local administration, this method has been developed in the US as a project that measures performances of local governments by comparison among those of other governments [12].

Some researchers propose the "Benchmarking Method" as a useful method in the case of legislation [8, 13]. According to Tanaka [14], "benchmarking of ordinances is defined as a method that municipalities continuously compare best practices of ordinance system developed by other municipalities with their own situation so that they can utilize such models in designing and operating ordinance system" (p. 204). He also suggests that legislative drafters should follow the process of benchmarking for legislation, such as clarifying the purpose of ordinance legislation, analyzing the current situation of its own municipality, selecting benchmarking points, choosing municipalities to be benchmarked.

If benchmarking method is used properly in legislation process, the issues of patchworking (3) or insufficiency in consideration of legislative purposes (4) could be avoided. Both copy and pasting and patchworking during legislative work are not human but mechanical activities without thorough consideration. The reason such actions were taken should be investigated further. Nevertheless, it is possible to think that government officials who engage in legislative work do not know the "benchmarking method" and the way to use it sufficiently, as it is often the case that those in charge of drafting ordinances are not familiar with this task.

Therefore, a possible response to these two issues is to provide education for legislation. In terms of the eLen regulation database system, offering legislators training occasions would be valuable in order to learn how to make the best of benchmarking functions in the process of legislation. In fact, we have provided such education to government officials, members of assemblies, and support staff for legislation such as librarians. Moreover, to study e-legislation, we have planned to create a consortium in cooperation with Judicial Policy Education Research Center in Kagoshima University, by which the eLen has been provided to all Japanese municipalities.

Another response to these issues is to design the system more instructively so that users can learn the proper method and process of legislation through using the database. The eLen has already implemented instructive functions. For instance, the functions of benchmarking tables and templates could be worked as a device for making users realize errors or problems of their own ordinances as these functions can highlight not only similarities but also differences among various ordinances. In our recent survey conducted in 2019, several officials said that the pull-down menu of templates included in the eLen was useful as an instructive tool (Mie Prefecture, Ishikari City and Satsumasendai City). An important comment was made by an official in Ishikari City. He pointed out that the function would provide users with other choices clearly, so that it could clarify thinking points that give legislators hints and clues. We have thus improved the eLen to implement educational functions.

Response to the Issue (4). Regarding inadequacy of legal research, first, it is necessary to provide legislators with education on legal research since some specific knowledge and skills for research are required. The authors have engaged in education on legal research for students at universities as well as municipal officials for a long time. Furthermore, we have a plan to implement a function for tracing relevant laws and regulations. It would enable users to check their research.

4.2 Education for Legislators

Overall, the eLen has been developed to provide legislators several functions such as "Context Searching," "Cross Searching," "Benchmarking Table" and "Template" that can be utilized to diminish four main issues addressed by interviewees. However, no matter how human-centric computing is aimed through paying attention to real users' feedbacks and developing the system based on them, it is essential that human beings themselves act properly when they use technology. The objective of human-centric e-legislation is not to provide an automated legislative system that will replace human work in legislation completely, but to create a system that streamlines laborious tasks, reduces human errors, and provides clues for legislative work in order for legislators to focus on tasks that only human can accomplish. Thus far, it is significant to provide education in legislation, such as trainings for learning the way to use the system, benchmarking method, or legislation process. Although the eLen has already implemented several instructive mechanisms, we will improve further the system so that users can learn the proper process of legislation through the usage of the system.

5 Conclusion

The objective of this study was to analyze results of interviews that we conducted with Japanese municipal officials who have engaged in legislation of ordinances or regulations and to present main issues addressed by the interviewees. Moreover, based on the results of analysis, it aimed to clarify remaining problems that legislators face with during legislation, which will be necessary conditions for expanding and improving human-centered functions of e-legislation systems. Using qualitative analysis of interviews with municipal officers, this research identified the following four issues: (1) inconsistency among ordinances; (2) inconsistency of an ordinance; (3) insufficiency in consideration of legislative objectives and facts; (4) inadequacy of legal research. Based on the results of interview analysis, this study assessed whether the eLen regulation database system coped with those issues and discussed remaining problems. Overall, it illustrated that some functions included in the eLen were helpful for diminishing those issues. However, in order to overcome the problems with which legislators face in the process of legislation, the results of this study showed that it was significant to provide legislators with education, such as trainings for learning the way to use the system, benchmarking method or legislation process.

There are two main limitations in this paper. First, the interview data are not up to date. As further interviews with local government officials have continued to be conducted by the authors, it will be expected to analyze the new qualitative data in the next study. In addition, although there have been a few cases of interviews with members of assemblies and lawyers implemented by the authors, it would be useful to expand the range of interviewees to grasp the different aspect of legislation in Japan. Second, the situation of the legislation process in municipalities might be unique to Japan.

However, there are few studies on legislation process in municipalities in other nations, especially focusing on real voices of people who engage in legislation by using qualitative analysis. Thus, in order to conduct an international comparative study on legislation process in municipalities, it will be valuable to run our own survey of those who engage in legislative work in other countries. There may be some similarities with Japanese legislative approach. Japanese benchmarking method for legislation might be also adopted or deserves consideration as a new legislative method for other countries. Since the method of clustering and template production adopted in this study are not influenced by languages, it would be possible for the eLen database system to be transplanted into different countries with different languages. Therefore, further qualitative research in this field would be of great help in exploring possibility of the eLen as well as our research.

Acknowledgements. This study was supported by SECOM Science and Technology Foundation and JST, CREST, Grant Number JPMJCR1782, Japan.

References

1. Nakamura, M., Kakuta, T.: Development of the eLen regulation database to support legislation of municipalities. In: Proceedings of International JURIX 2014 Conference, pp. 185–186 (2014)
2. Kakuta, T., Shima, A., Saito, D., Otani, T.: Zenkoku jichitai reiki database eLen no kouchiku to teiryouteki reikichosa [Development of eLen database for regulations of the local governments and a quantitative survey of the regulations]. Inf. Netw. Law Rev. **13**, 14–33 (2014). [in Japanese]
3. Shima, A., Kakuta, T.: Jichitai reiki sakuseiji niokeru tareiki no sanshou ni kansuru chosahoukoku: kanagawa kennaino zenshichoson wo sanpuru nishite [Report on the legislative drafting work compared with other regulations in municipalities: a sample survey on municipalities in Kanagawa prefecture]. Nagoya J. Law Polit. **259**, 383–409 (2014). [in Japanese]
4. Shimazu, A.: Hourei kougaku: anshin na shakai shisutemu sekkei notameno houhouron: houreibunsho no kaiseki wo chushin ni [Legal engineering: methodology for designing trustworthy social systems: legal document analysis]. IEICE Found. Rev. **5**, 320–328 (2012). [in Japanese]
5. Kakuta, T.: e-Legislation kankyo no kouchiku he mukete: jouhoukagaku wo ouyou shita rippoukatei no sagyousien [Toward developing e-Legislation environment: supporting law-making works based on information science]. Inf. Net. Law Rev. **11**, 13–32 (2012). [in Japanese]
6. Hasegawa, T.: Nigen daihyou no seisaku rippouryoku bunseki [Comparison of legislative abilities between tops of local government and members of assemblies]. Koukyou-seisakukenkyu **17**, 108–119 (2017). [in Japanese]
7. Katou, Y., Hiramatsu, H.: Giin jourei shuran [The collection of ordinances enacted by assembly members]. Koujinsha, Tokyo (2011). [in Japanese]
8. Makise, M.: Kata kara surasura kakeru anata no machi no seisaku jourei [Let's try to draft an ordinance to implement policies in your city: using templates]. Daiichihoki, Tokyo (2017). [in Japanese]

9. Kitamura, Y., Yamaguchi, M., Izuishi, M., Isozaki, H.: Jichitai seisaku houmu [Legal work for implement of policies in municipalities]. Yuhikaku, Tokyo (2011). [in Japanese]
10. Gyoseihouseishitsumukenkyukai: Houseisitsumu nyumon [Introduction to legislative drafting with illustrations]. Gyosei, Tokyo (2013). [in Japanese]
11. Diamond Harvard Business Review: Benchmarking no riron to jissen [Theories and practices of benchmarking]. Diamond, Tokyo (1995). [in Japanese]
12. Tanaka, H.: Jichitaihyouka niokeru benchmarking no kanousei: beikoku no torikumi ga shisasuru mono [Future prospects of benchmarking in local government evaluation: the implications of practice in the United States]. Jpn. J. Eval. Stud. **11**(2), 3–12 (2011). [in Japanese]
13. Tanaka, T.: Jourei zukuri eno chousen [Challenge of drafting ordinances by using benchmarking method]. Shinzanshashuppan, Tokyo (2002). [in Japanese]
14. Tanaka, T.: Jichirippou no shuhou [Methods of autonomous legislation]. In: Kisa, S. (ed.) Jichirippou no riron to shuhou [Theories and Methods of Autonomous Legislation], pp. 203–211. Gyosei, Tokyo (1998). [in Japanese]

Japan-EU Passenger Name Record Negotiations and Their Implications

Toru Maruhashi

Meiji University, Kanda-Surugadai, Chiyoda-ku, Tokyo 101-8301, Japan
torumaruhashi@meiji.ac.jp

Abstract. EU and Japan started negotiations of an agreement on passenger name record ("PNR") while EU-CANADA PNR negotiations are concluding. Responding to the call from United Nations Security Council, International Civil Aviation Organization is discussing amendments on the PNR to the Standards and Recommended Practices on Facilitation. In these bi-lateral and global circumstances, what should be desirable outcome of negotiations on Japan-EU PNR Agreement and could be broader issues left behind?

In this contribution, PNR and the nature of its processing are overviewed (Sect. 2), and their current legal and practical framework in Japan is critically confirmed (Sect. 3). Then Japanese PNR system is compared with the original draft negotiating directive of Japan-EU PNR Agreement with author's perspective (Sect. 4) and finally global implications for the PNR and importance on (relatively legacy) technology and practice of algorithmic pattern-based search are explored.

Keywords: Passenger name record (PNR) · PNRGOV · Pattern-based search · Transparency

1 Introduction

On 26 July 2017, the Grand Chamber of Court of Justice of the European Union ("CJEU") delivered Opinion 1/15[1]. The Court declared that the envisaged agreement between European Union and Canada on the transfer and processing of passenger name record data ("PNR data" or "PNR") is incompatible as to sensitive data possibly included in the PNR. It also instructed that various other points must be amended to be compatible with Articles 7 and 8, and Article 52(1) of the Charter of Fundamental Rights ("CFR").

In July 2019, EU and CANADA jointly declared that they have concluded negotiations for a new PNR Agreement while Canada noted its requirement for legal review[2].

[1] CJEU *Opinion 1/15* ECLI:EU:C:2017:592. *See* Maruhashi T (2019) Draft PNR Agreement between CANADA and EU; CJEU Opinion 1/15 - Distance from Mass Surveillance and Data Retention – Information Network Law. 17:63–91. (in Japanese).

[2] Para. 11 of Canada-EU Summit Joint Declaration July 17–18, 2019, Montreal.

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 100–114, 2020.
https://doi.org/10.1007/978-3-030-62803-1_9

Globally, International Civil Aviation Organization (ICAO) has been working on new standards and recommended practices (SARPs) on PNR to be contained in Section D, Chapter 9, Annex 9 to the Chicago Convention responding to the call by United Nations Security Council Resolution 2396. European Commission has been participating in that SARPs' drafting process following the negotiation position approved by Council.[3] The draft SARPs seems to incorporate various points from that EU position.[4]

Japan has been on the EU waiting list for the negotiations on PNR Agreements next to Canada.[5]

The Article 37 of Japan-EU Strategic Partnership Agreement ("Japan-EU SPA") is explicit on the partners' endeavour to use PNR.[6]

European Commission, without waiting its conclusion of PNR Agreement with Canada, recommended the Council to authorise the opening of negotiations for Japan-EU PNR Agreement with negotiating directives ("NDs")[7].

In these bi-lateral and global circumstances, what should be desirable outcome of negotiations on Japan-EU PNR Agreement and could be broader issues left behind?

PNR is a unique set of personal data in several dimensions. These data are internationally transferred from private operators to governments. They are used and analysed for governmental purpose of preventing and combating terrorism and other serious transnational crime. That analysis involves some extent of algorithmic profiling and prediction. They are further transferred to another domestic or foreign government. If we find problems or shortcomings in these processing, that could generally be applicable to other kind of data similarly situated and the way of their regulation.

[3] Council Decision (EU) 2019/2107, which preamble 14 states "The position of the Union is established in accordance with the applicable Union legal framework on data protection and PNR data, namely Regulation (EU) 2016/679, Directive (EU) 2016/680 and [PNR] Directive (EU) 2016/681, as well as the Treaties of the European Union and CFR as interpreted in the relevant case law of the CJEU, in particular Opinion 1/15".

[4] Proposed by a task force in ICAO information paper FALP/11-IP/1(December 2019) and approved with amendments at ICAO Facilitation Panel Eleventh Meeting. See Meeting Report FALP/11 (January 2020).

[5] A. Iizuka, Director of the Customs and Tariff Bureau, Ministry of Finance (Government response in the Diet on 23 March 2018 in Japanese). European Commission expressed its view on the negotiations that "Having arrangements in place in time for the 2020 Olympics would bring a real security dividend." in its Twentieth Progress Report towards an effective and genuine Security Union COM(2019) 552 final 30.10.2019.

[6] Japan-EU Strategic Partnership Agreement was signed in July 2018 and partly became effective on 1 February 2019. Its Article 37 reads 'The Parties shall endeavour to use, to the extent consistent with their respective laws and regulations, available tools, such as passenger name records, to prevent and combat acts of terrorism and serious crimes, while respecting the right to privacy and the protection of personal data'.

[7] European Commission, COM(2019) 420 final (September 2019). The Economic and Financial Affairs Council authorised revised version of the NDs (12762/19 + ADD 1) on 18 February 2020, which is still confidential as of 24 June 2020. *See also* EDPS Opinion 6/2019 on the negotiating mandate of an Agreement between the EU and Japan for the transfer and use of Passenger Name Record data 25-Oct-2019, a*vailable at* https://edps.europa.eu/data-protection/our-work/publications/opinions/eu-japan-passenger-name-record-data-agreement_en.

In this contribution, PNR and the nature of its processing are overviewed (Sect. 2), and their current legal and practical framework in Japan are critically confirmed (Sect. 3). Then Japanese PNR system is compared mainly with the original draft NDs of Japan-EU PNR Agreement with author's perspective (Sect. 4) and finally, global implications for the PNR and importance on (relatively legacy) technology and practice of algorithmic pattern-based search are explored.

2 PNR and the Nature of Its Processing

2.1 Passenger Name Record

PNR "is the generic name given to records created by the aircraft operators for each flight a passenger books. PNR records contain information provided by the passenger and information used by the aircraft operator for their operational purposes."[8] Contracting States of ICAO requiring PNR access should align their data requirements and their handling of such data to guidelines contained in ICAO Document 9944 (ICAO-WCO-IATA PNR Guidelines).

2.2 PNR Push and PNRGOV Standard

Governments use PNR to conduct analysis that helps to identify possible high-risk individuals that may have been otherwise unknown to government authorities and make, where appropriate, the necessary interventions. PNR information can be provided by aircraft operators by sending the information electronically ("push" method) or allowing the appropriate authorities to access the parts of their reservation systems where the PNR information is stored ("pull" method). However, internationally there is an agreement to utilize the "push" method, for data privacy reasons (See footnote 8).

To ensure interoperability for reporting to the appropriate government authorities, a push method message format called PNRGOV based on EDIFACT rules and syntax is developed 'as the international standard that must be used for the transmission of PNR' (See footnote 8). The standards[9] contains complete description of the message structure, segments and elements as well as the relationship between messages.

2.3 Targeting

Each passenger's PNR data is transferred using push or pull method from the airline's reservation system to the border authority's system such as immigration and customs before entry and departure. In the relevant system, the risk of a passenger is automatically evaluated by a program that incorporates an algorithm or scenario for

[8] Management Summary on Passenger-related Information ['Umbrella Document' version 2.0 – July 2017] published by the International Civil Aviation Organisation (ICAO), the World Customs Organisation (WCO) and the International Air Transport Association (IATA).

[9] *See* PNRGOV EDIFACT Message Implementation Guides updated, *available at* https://www.iata.org/en/publications/api-pnr-toolkit/#tab-3.

matching the pattern of high-risk persons, as a pre-processing step called targeting. Once targeted by the risk evaluation process, further risk assessments and investigations are performed and the corresponding passengers are taken additional interrogation at entry and exit. Explanations on Canadian and the US examples of these pattern-based (or rule-based) searches[10] follow:

The Scenario Based Targeting (SBT) program of Canada Border Services Agency uses advanced analytics to evaluate [PNR] against a set of conditions or scenarios, [which are] made up of personal characteristics derived from [PNR], such as age, gender, travel document origin, places visited and length and pattern of travel. If an individual matches a scenario, further manual risk assessments are conducted by National Targeting Centre officers. Risk assessments include checking individuals against international and domestic law enforcement and intelligence partners' databases, and may result in the individual being referred as a "target" for closer questioning or examination by a Border Services Officer at the port of entry.[11]

The Department of Homeland Security (DHS) U.S. Customs and Border Protection (CBP) operates the Automated Targeting System (ATS) [to manage] the shared threat to the homeland posed by individuals ... that may require additional scrutiny prior to entering or exiting the United States. [In identifying such individuals] ATS compares existing information about individuals [including PNR] ... entering and exiting the country with patterns identified as requiring additional scrutiny. The patterns are based on CBP Officer experience, trend analysis of suspicious activity, law enforcement cases, and raw intelligence[12].

In the latter example of the U.S. ATS, currently PNR is just one of the various source information ingested to the system; other information comes from, for example, Border Crossing Information (BCI), Electronic System for Travel Authorization (ESTA), Secure Flight Passenger Data (SFPD).

3 Current Legal and Practical Status of PNR in Japan

In Japan PNR is collected through NACCS system[13] operated by Nippon Automated Cargo And Port Consolidated Systems, Inc.[14] Airline operators are obligated to report PNR in a format designated for NACCS directly or indirectly through its Service

[10] *See* Zarsky TZ (2013) Transparent predictions. U Ill L Rev 1503, arguing importance of modest transparency in prediction process focusing on pattern-based searches.

[11] Office of the Privacy Commissioner Canada, *2016–17 Annual Report to Parliament on the Personal Information Protection and Electronic Documents Act and the Privacy Act.*

[12] Privacy Impact Assessment Update for the Automated Targeting System DHS/CBP/PIA-006(e) January 13, 2017.

[13] NACCS is a system for online processing of procedures taken with Customs, ISA and other relevant administrative authorities or related private-sector services for arriving/departing ships and aircraft or import/export cargo. *See* The service homepage, *available at* https://bbs.naccscenter.com/naccs/dfw/web/ (Japanese).

[14] A Special Corporation governed by Act on Processing, etc. of Business Related to Import and Export by Means of Electronic Data Processing System (Act No. 54 of 1977).

Provider[15] in EDIFACT based PNRGOV format to NACCS. Relevant user authorities such as Immigration Service Agency ("ISA"), and Customs and Tariff Bureau, Ministry of Finance ("Customs") pull PNR reported to NACCS center as described in more detail below.

3.1 PNR and Immigration Control

According to Paragraph 8 of Article 57 of the Immigration Control and Refugee Recognition Act ("ICRRA")[16], an Immigration Inspector may request to report inbound PNR if he or she finds it necessary for securing the enforcement of landing examination … from an aircraft operator, a chartered aircraft operator or a joint carrier.

PNR under ICRRA are matters relating to (i) the person making the reservation, (ii) the details of the reservation pertaining to (i), (iii) the baggage of (i) and (iv) the procedures for (i) to board the aircraft. PNR entries are detailed in its Regulation[17].

Inbound PNR shall be reported within 60 min from the request. From 1 January 2016, inbound PNR became able to be reported to immigration authorities through NACCS; since then they have had electronic access to PNR reported to NACCS.

3.2 PNR and Customs

The Customs Law ("CL")[18] covers both inbound and outbound PNR.

According to Paragraph 12 of Article 15 of CL, a Director-General of Customs-house ("DG Customs") may request to report inbound PNR if he or she finds it necessary for securing the enforcement of Article 69-11 (embargoed goods on import) or other provisions of CL, from an aircraft operator or a joint carrier and according to Article 15-3, from a chartered aircraft operator.

According to Paragraph 3 of Article 17 of CL, DG Customs may request to report outbound PNR for securing the enforcement of Article 69-2 (embargoed goods on export) or other provisions of CL, from an aircraft operator or a joint carrier and according to Paragraph 2 of Article 17-2, from a chartered aircraft operator.

Under CL, both inbound and outbound PNR are (i) names of the personal who reserved the tickets, (ii) the details of the reservation, (iii) their accompanying luggage, and (iv) any information with regard to boarding procedures. PNR entries are detailed in Cabinet Order[19] and Ordinance[20].

The Passenger and the Reservation Information shall be reported within 60 min and the Belongings and Check-in Information shall be reported within 30 min from the request.

[15] Currently only ARINC can connect NACCS Center for PNR "Push".

[16] Cabinet Order No. 319 of October 4, 1951.

[17] Regulation for Enforcement of the Immigration Control and Refugee Recognition Act (Ministry of Justice Order No. 54 of 1981).

[18] Act No. 61 of 1954.

[19] Cabinet Order No.150 of 1954.

[20] Regulation for Enforcement of the Customs Law (Ministry of Finance Order No. 55 of 1966) ("RECL").

Inbound PNR has been available for DG Customs' request since October 2011 and electronically through NACCS since April 2015. Outbound PNR has been available for DG Customs' request since June 2017 and through NACCS since March 2019. In practice, airline operators are required to report both inbound and outbound PNR twice, 72 h before scheduled time of departure and immediately after the departure.

The major purpose of obtaining outbound PNR is explained as 'grapping the behavior of re-entry passengers by comparing departure information and entry information'[21].

Since March 2019, the electronic reporting of both inbound and outbound PNR through NACCS has been mandatory[22].

3.3 Processing of PNR Through NACCS

Once an airline operator inputs PNR directly or indirectly to NACCS, it satisfies legal "push" requirement under ICRRA and CA.

NACCS just operates as a common proxy server making PNR received from airline operators available for ISA and Customs to "pull" it for these authorities' statutory purpose. PNR is just retained in NACCS for six days according to the business process specification specified by the Customs[23].

Headings for PNR in NACCS format mainly consists of brief texts, numbers and IATA codes, but exceptionally, two headings of 'other information' are in free text format and can be long one, corresponding to SSR and IFT tags in Segment Group 1 of PNRGOV.

NACCS allows airline operators to push their PNR in PNRGOV format processed based on PNRGOV standard.

Japanese PNR system is compatible with ICAO-WCO-IATA PNR Guidelines.

3.4 Retained PNR as Personal Information File

Particulars of PNR collected by ISA and Customs shall in accordance with Articles 10 and 11 of the Act on the Protection of Personal Information Held by Administrative Organs (APPIHAO) be notified to Ministry of Internal Affairs and Communications (MIC) and published as a Personal Information File (PIF) Register, elements of which includes:

Name of the PIF; Name of the Administrative Organ and the name of the organizational section in charge of the processes for which the PIF will be used; Purpose of Use of the PIF; Particulars recorded in the PIF; the scope of individuals that are

[21] 'Enhancement for Reporting Scheme of "Advance Information on Passenger" (ANNEX)' presented by Customs to Customs Subcommittee, Tariffs and Foreign Exchange Committee 24 Nov 2016 (in Japanese).

[22] Section 14, Article 15 of CL and Article 2-5 of RECL. *See* also Enhancement of Reporting Scheme of Advance Electronic Information (AEI) on Passengers and Crews (May 2017), *available at* https://www.customs.go.jp/mizugiwa/ryogu/kekka03.pdf.

[23] https://bbs.naccscenter.com/naccs/dfw/web/data/customs/jimu/toriatsukai_index_tetsu_k_1.html (in Japanese).

recorded in the PIF; The means of collecting the Personal Information recorded in the PIF; Whether the Recorded Information contains Special Care-required Personal Information; If the Recorded Information will be routinely provided to a party outside the Administrative Organ, the name of that party.

However, PIF is exempted from these notice and publication requirement (Articles 10(2) and 11(2) of APPIHAO), if it contains particulars concerning the security, diplomatic secrets, and other important interests of the State or it is prepared or obtained for criminal investigation, investigation of tax crimes based on the provisions of laws related to tax, or instituting or maintaining a legal proceeding.

Processing and retention of PNR are governed by APPIHAO and Public Records and Archives Management Act ("PRAMA"). Individuals (including foreign nationals living abroad) have in principle a right to disclosure, correction (including deletion) and suspension of use or provision under the APPIHAO[24].

Under Article 5 of PRAMA, when an employee of an administrative organ has prepared or obtained an administrative document, the head of administrative organ must set the retention period of document and the date on which the relevant retention period expires.

Current retention period of PNR seems to be 5 years (ISA) and 7 years (Customs) respectively according to Administrative Document File Management Registry. After the retention period, they will be deleted outright. Unlike EU-Canada PNR agreement, no masking operation is used.

3.5 PNR Transparency/Data Minimisation

According to Article 49 of APPIHAO, the MIC "may collect reports on the status of enforcement of this Act from the heads of Administrative Organs" and annually publishes a summary of the reports.

The PIFs on PNR of ISA and Customs describes their purposes as 'equitable control over the entry into and departure from Japan and residence of all persons' and 'prevention of embargoed goods from importation into Japan etc.' respectively; both of them do not contain Special Care-required Personal Information i.e. sensitive information nor routinely provided to other administrative organs.

Data formats of Japanese PNR is compatible with ICAO-WCO-IATA PNR Guidelines and can be processed based on PNRGOV standard through NACCS. To achieve data minimisation, some 'other information', which are expected as filled in free text information corresponding to SSR and IFT tags in Segment Group 1 of PNRGOV shall be automatically filtered out through NACCS processing.

Customs publish a webpage entitled 'Summary of Passenger Name Record (PNR)'[25] and specifies the purpose of use of transmitted PNR as "customs enforcement purpose including preventing the smuggling of terrorism related goods and illicit drugs".

[24] *See* preamble 165-170 of Commission Implementing Decision (EU) 2019/419 of 23 January 2019 pursuant to Regulation (EU) 2016/679 of the European Parliament and of the Council on the adequate protection of personal data by Japan under the Act on the Protection of Personal Information ("Adequacy Decision").

[25] https://www.customs.go.jp/english/procedures/advance4_e/index_e.htm.

3.6 Targeting Use

PNR has been analyzed for purposes of immigration and customs examination, but there is little public information about how it has been used and useful for targeting purposes, other than just for blacklist matching by supplementing it with other information such as Advanced Passenger Information.

According to Basic Plan for Immigration Control and Residency Management [26], as to the use of PNR, its "Immigration Control Intelligence Center[27] … receive PNRs [through NACCS since January 2016] and, then become able to conduct advanced analysis using the information held by the Ministry of Justice along with other information, and these results are being used in the border measures by the regional immigration control and residency management offices at the port of entry".

ISA "conduct[s] advanced analysis using [PNR] held by [it] along with other information, and these results are being used in the border measures by the regional immigration control and residency management offices at the port of entry… [, and is] taking measures to identify persons who pose a security risk, …and preventing their entry, and will continue to strengthen use of such information and to conduct smooth and prompt entry examinations for foreign nationals who do not pose a problem."[28]

In September 2019, ISA was "considering the introduction of a system, a so-called rules-based engine that generates automatic trend analysis and categorization of foreigners to be cared and performs automatic matching"[29]. This presumably mean that automated targeting using PNR has not actually progressed at least on the ISA side.

3.7 Sharing and Further Use by Domestic Authorities

As to sharing of PNR data with other relevant domestic agencies, "[i]n accordance with the [APPIHAO], [Customs] share PNR data with other relevant government agencies solely to the extent necessary for the customs enforcement purpose including preventing from smuggling of terrorism related goods and illicit drugs and [Customs] determine the necessity of sharing for each case individually"[30].

Under ICRRA, the Director-General and other official of ISA are encouraged to share information with relevant governmental agencies (Article 61-7-7), and may request necessary corporation from them (Article 61-8).

APPIHAO does not prohibit an administrative organ from sharing Retained Personal Information ("RPI") contained in PIF solely for the original purposes, or as otherwise provided by laws and regulations (Article 8(1)).

In addition, an administrative organ may provide another person with RPI for purposes other than the original purpose, if RPI is provided to another administrative

[26] http://www.immi-moj.go.jp/seisaku/pdf/2019_kihonkeikaku_english.pdf.

[27] A Passenger Information Unit (PIU) for ISA.

[28] Ministry of Justice 'Basic Plan for Immigration Control and Residency Management (April 2019). http://www.immi-moj.go.jp/seisaku/pdf/2019_kihonkeikaku_english.pdf.

[29] Minutes of the 16th meeting of 'The seventh Immigration Policy Discussion Panel' and its material no. 3 'immigration control'(19 September 2019) (in Japanese).

[30] fn. 27.

organ or local public entity, who uses it only to the extent necessary for executing the processes or business under its jurisdiction provided by laws and regulations, and there are reasonable grounds for the use of that RPI (Article 8(2)(iii)).

Accordingly, Customs or ISA may, within their mandate, share PNR with national or prefectural law enforcement agencies or agencies responsible for national security, and within the latter's mandate, if there are reasonable grounds for the use of PNR.

However, there are few public records regarding governmental use of PNR for the purpose of law enforcement or national security except for a few examples of an abstract narrative description. In 2019, enormous numbers of foreign dignitaries came to Japan for the emperor's throne. The National Police Agency reported, "The police collected and analyzed comprehensive terrorism-related information in close cooperation with foreign security intelligence agencies, etc. to prevent illegal acts such as terrorism against these key persons. At the same time, in cooperation with the [ISA] and [Customs]…countermeasures against terrorism, such as border measures, utilizing [PNR] were taken."[31]

Actually, no such out of the purpose sharing is published in the past Article 49 APPIHAO annual summary reports compiled by the MIC.

Probably, police and intelligence agencies actually and routinely have been provided from ISA and Customs and have used PNR for the original purpose, but they do not retain PNR as subset of NACCS based database, rather it has somehow been entered as a new entry into or flagged as PNR-originated on an existing entry in their own black lists, or if it is suspect's one, is incorporated into a criminal case records, which is outside the scope of APPIHAO PIF disclosure and reporting scheme.

3.8 Sharing and Further Use by Foreign Authorities

Customs share PNR data with other foreign customs administrations, "in accordance with the Article 108-2 of the [CL] and Customs Mutual Assistance ("CMA") Framework, solely to the extent necessary for the customs enforcement purpose including preventing from smuggling of terrorism and related goods and illicit drugs and [Customs] determine the necessity of sharing for each case individually"[32]. The nuances of the actual provisions of that Article 108-2 are slightly different; reciprocity principle and its purpose limitations are stated as:

Customs, when sharing information must confirm the foreign counterpart authority can provide the information equivalent to the information to be shared, the same level of confidentiality as in Japan is guaranteed by the laws of the relevant foreign country for the information to be shared, and the information to be shared is not used for any purpose other than contributing to the performance of the foreign counterpart's duties.

Minister of Justice, i.e. ISA can share PNR with other foreign immigration administrations, in accordance with the Article 61-9 of the ICRRA solely for their enforcement purpose and appropriate measures shall be taken to ensure that shared information is not used for purposes other than helping the foreign counterpart

[31] "Review and Outlook of Security (2019)", Security Bureau, National Police Agency (in Japanese).

[32] fn. 27.

authorities execute their duties. Unlike CL, no reciprocity or confidentiality is required here.

As to the privacy and data protection in international agreements, most of the CMA Agreements and Cooperative Frameworks has very simple provisions. For example, Paragraph 2 of Article 16 Japan-European Community on Co-Operation and Mutual Administrative Assistance in Customs Matters provides reciprocally for:

Personal data may be exchanged only where the Contracting Party which may receive it undertakes to protect such data in at least an equivalent way to the one applicable to that particular case in the Contracting Party that may supply it. The Contracting Party that may supply the information shall not stipulate any requirements that are more onerous than those applicable to it in its own jurisdiction. The Contracting Parties shall communicate to each other information on the laws and regulations of each Contracting Party, including where appropriate, those in the Member States of the Community.

Thus, PNR obtained from airline operators may be provided (or feed-backed) from Japan Customs to EU member states customs solely for the latter authority's duty, which is not restricted to prevention of smuggling of terrorism and related goods and illicit drugs.

3.9 Data Security; Security Breach Notification

The head of an administrative organ must take necessary measures to prevent the leakage, loss, damage, and other appropriate management of the RPI (Article 6 of APPIHAO). MIC "Guidelines on Measures for Appropriate Management of Personal Information Held by Administrative Organs" breaks down data security measures and a data breach incident shall be promptly reported to Chief Privacy Officer of the Organ. If the fact on the data breach will be publicly disclosed, the information will be promptly provided to the MIC. If the data breach is caused by staff of the agencies, disciplinary action will be taken (Article 82 of the National Public Service Act) and penalized in case of intentional leakage of confidential information (Article 109 (xii) of the National Public Service Act).

As to access control, Customs disclose, "Limited officials in the centralized unit solely utilize PNR data for analyzing and targeting of passengers. To limit the access to PNR data, the data is used in the closed system and the access to the facility in which the unit is situated and the office of the unit is strictly limited" (See footnote 34). The centralized unit is Passenger Information Unit ("PIU"), though it is not in a cross-agency style.

4 Comparative Analyses of EU NDs and Japanese PNR Processing Legislation and Practice

As to PNR negotiations with EU, Japanese government recognises the necessity and importance of the use of PNR (**para. 2 of NDs**).

As a relatively early adapter, Japan has already extended its PNR system coverage to most of the airline operators other than EU-based one. Through NACCS, airline operators are using PNR ‘push’ and can transmit it in PNRGOV standard format twice (72 h before the scheduled departure time and immediately after the departure) without heavy burden or inconvenience (**paras 10–12**). As far as Japanese laws and regulations are concerned, EU airline operators have a legal basis for them to transfer PNR via NACCS to ISA and Customs (**para. 7**).

The other negotiation points are, as elaborated in Sect. 3 above and for the reason not controllable by Japanese government, not simple or easy. Author’s perspective follows.

4.1 Purpose Limitation (Paras. 3 and 5)

Limiting use of PNR for - although Article 37 of Japan-EU SPA is in line with - sole purpose of preventing and combating terrorism and other serious transnational crime defined in EU Legislation, would lessen the discretion of ISA and Customs in using and sharing PNR because they are not subject to any such limitation under Japanese laws and regulation. We need special provisions for this limitation under ICRRA and CL.

As to defining the categories of crime, recent working arrangement between NPA and Europol (“WA”) [33] covers exchange of specialist knowledge and strategic analysis on terrorism and serious crime, and lists areas of crime specified by Europol though NPA reserves its position. The list could be a starting point for the negotiations.

4.2 Transfer of Analytical Information (Para. 4)

NDs requires flow of ‘analytical information’ from competent authorities of Japan to police and judicial authorities of the Member States, Europol and Eurojust[34]. This flow is not specifically covered by any of the previous agreements or frameworks regarding the immigration control or customs. In contrast, WA encourages exchanging analyses, but does not provide for the legal basis for the transfer of personal data. If ‘analytical information’ includes personal information, we need to comply with APPIHAO[35].

4.3 Clear and Precise Safeguards and Controls (Para 8)

Respect for Fundamental Rights and Freedoms (Paras. 3 and 6). As CJEU in its Opinion 1/15 on the envisaged EU-Canada PNR Agreement used the strict necessity test to ensure proportionality of that Agreement to protect fundamental rights and freedoms, we need to and probably it is enough to consider that its guidance is broken down in para. 8.

[33] Working arrangement on establishing cooperative relations between the National Police Agency of Japan and the European Union Agency for Law Enforcement Cooperation signed on 03 December 2018.

[34] From EU to Japan, *See* EDPS Opinion fn. 7 paras 22–23

[35] *See* II. (b) (2) Limitations flowing from APPIHAO of Appendix 2 to Adequacy Decision

The Categories of PNR, Data Minimisation and Proportionality/Sensitive Data. As elaborated in Subsect. 3.3 above, Japanese PNR is considered to be already minimised in line with PNRGOV and to the proportionate to the purpose of the PNR Agreement except for headings allowing input of long free text information.

APPIHAO does not prohibit Administrative Organs from processing sensitive data within the meaning of EU law. However, as far as both PIFs are concerned, personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership or concerning a person's health or sexual life or orientation cannot on its face be included in headings other than two 'other information' headings in free text format.

If the free text information in these headings can be automatically filtered out through NACCS processing, there will be no room for sensitive data remaining in PNR and data minimisation requirement will be satisfied at the same time.

Data Security, Security Breach Notification. The general data security requirement under APPIHAO is as described in Subsect. 3.9.

Here, the addressees of the data breach notification are designated as European national data protection supervisory authorities; it seems difficult for Japanese PIUs in Customs or ISA to give the notification directly to such supervisory authorities.

A concession would be to construct communication route via Japanese supervisory data protection authorities, if any, or Personal Information Protection Commission (PPC)[36] on their behalf.

Transparency, Right of Access. As discussed above, transparency to passengers are to some extent achieved by publishing PIFs. In addition, Customs have certain level of notice to the passengers on its website. We shall improve the notification content.

APPIHAO has provisions on right to access, rectification and deletion, where appropriate, however, the problem is how to secure the right of individual notification of the use of PNR as instructed by CJEU[37] citing analogically *Tele2 Sverige*[38]. EU does not legislate in PNR Directive or Law Enforcement Directive[39] nor included in PNR agreements with the U.S. or Australia this right by which an individual must be notified of the use of information by competent authorities as soon as that information is no longer liable to jeopardise the government investigations. At maximum concession, a mechanism similar to that presented in the Adequacy Decision[40] under which an individual who suspects that his/her PNR has been collected or used by public authorities in Japan can submit a complaint to the PPC[41].

[36] *See* Subsect. 4.4 below.

[37] *See* Opinion 1/15 fn. 1 paras. 221-225, fn. 1.

[38] *Tele2 Sverige and Watson and Others*, C-203/15 and C-698/15, EU:C:2016:970, (21 December 2016) para. 121.

[39] Directive (EU) 2016/680.

[40] fn. 26.

[41] *See* Adequacy Decision preamples1412013143.

Effective Redress. As to the effective administrative and judicial redress concerning PNR generally, legal framework explained in preambles 11–170 of Adequacy Decision will apply.

Automated Decision-Making, Database Compared. As preamble 93 of Adequacy Decision describes, Act on the Protection of Personal Information (APPI), APPIHAO nor relevant sub-statutory rules contain general provisions addressing the issue of decisions affecting the data subject and based solely on the automated processing of personal data. We need some regulations on these criteria, especially when ISA introduces 'rules-based engine'. Although practically PNR targeting would follow human review, algorithms used for pattern-based search for the purpose of targeting and these patterns need some scrutiny in various phases of development from political, legal and technological viewpoint as well as ex-post control such as review of usage and production of statistics on false positives and negatives and their disclosure or publication[42].

The Use of PNR Data by the Japanese Competent Authority Beyond Security and Border Control Checks. As discussed, ISA and Customs seems to routinely share PNR with police organisations for their original purpose. APPIHAO does not prohibit these agencies from sharing PNR with police to the extent necessary for law enforcement and there are reasonable grounds for the use of PNR. We need to clarify that enriching blacklists used by police with shared PNR needs certain level of additional supervision.

The Period of Retention of the PNR Data. PNR seems to be retained by 5 years (Customs) or 7 years (ISA) respectively and deleted outright upon expiration of these retention periods and no masking operation is used in the interim.

As CJEU does not allow retention of PNR after departure, justification of the purpose of retaining outbound PNR for behavioral analyses for the future re-entry is necessary[43].

Note that EU PNR Directive obligated PIU to retain both inbound and outbound PNR for 5 years (Article 12(1)). The negotiations in this regard would be highly dependent on the formal outcome of EU-Canada PNR Agreement.

Transfer to Other Authority, onward Transfer. Both ISA and Customs have statutory power to transfer PNR to domestic authorities or foreign counterparts and both of them are obligated to take certain measures to protect confidentiality or purpose limitation. We need to improve the level of protection in such transfer to the one in the PNR Agreement.

[42] *See* Zarsky fn. 12.

[43] *See* fn. 14. "The ATS 15-year retention period is based on CBP's historical encounters with suspected terrorists and other criminals, as well as the broader expertise of the law enforcement and intelligence communities. It is well known, for example, that potential terrorists may make multiple visits to the United States in advance of performing an attack. It is over the course of time and multiple visits that a potential risk becomes clear."

For limiting countries to which PNR is transferred to those countries of adequacy decision or concluding PNR agreement, we need to consider if the U.S., Canada, and Australia is sufficient for Japan to exchange PNR.

4.4 Oversight by an Independent Public Authority (Para 9)

As explained above, and more thoroughly described in the Adequacy Decision regarding police or intelligence purpose use, PNR obtained by Customs or ISA is under the supervision of several government agencies whose independence is questionable[44].

In this regard, since December 2019 Japanese government has been discussing in gathering and consolidating regulations regarding personal information protection relating to the private sector (i.e. APPI), administrative organs (i.e. APPIHAO), and incorporated administrative agencies and toward having the PPC centrally govern the consolidated systems. However, this consolidation process is expected to take long. In the interim, we need to have some tentative solution for independent oversight on PNR.

4.5 Joint Review of Targeting Algorithm? (Para 13)

Presumably, the regular joint review of the PNR Agreement will be conducted at the same level as the Adequacy Decision in general, but a review of the reliability and topicality of PNR targeting may be new thing. It is necessary to prepare for this kind of review of algorithm and develop a supervision strategy.

5 Conclusion

PNR and pattern-based search is developed on the relatively legacy technology. It seems perhaps that relative importance of PNR in the context of combat of terrorism and international organized crime might have been lowered.

Nevertheless, how to control pattern-based algorithmic search by government continues to be important, because, if we cannot control this level of technology and its use by government, we would face much more difficulties in controlling more sophisticated technology based on machine-learning (or artificial intelligence).

If the PNR analysis still has some effectiveness against terrorism and international organized crime, it is undoubtedly important that global development and collaboration supporting UN Security Council Resolutions and the ICAO SARPs discussion.

There is a need to standardize the way of controlling the PNR pattern-based search that should be implemented in democratic countries as well as to increase the number of countries that use PNR globally.

Should the partners of EU PNR Agreements including Japan create a system in which peers review each other in terms of technology and governance of PNR targeting instead of one-way audit of PNR?

[44] US supervision over PNR has a similar structure that spans multiple agencies. The shortcoming of that structure is criticized by Article 29 working party in its letter of 11 April 2018.

There is a need to deepen academic discussions while making efforts to increase the transparency and interpretability of algorithmic decision-making tools, such as statistics on the practicality efficiency and topicality of the PNR.

This paper prioritizes a detailed introduction to Japan's PNR system and does not conduct a deep legal doctrinal analysis. Author believes that this level of detail will work as lenses through which we view possible improvement in processing small set of, but important categories of personal data like PNR. If the conventional and orthodox approach on retaining, processing and sharing personal data by public agencies adopted by Japan will be refined through these PNR negotiations with the EU, it will bring a 'security dividend' to Japanese government and citizens and globally to participants in a forum such as ICAO.

The Data-Driven Society

The Data-Driven Society

Gopal T. V.

Anna University, Chennai, India
gopal@annauniv.edu

Centered around the human are arts, beliefs, customs, institutions, other outcomes of human work and thought with regard to a specified time or a given community. In a sense, it is a culture that evolves with the progress of the human towards a chosen goal in a chosen path either individually or collectively with fellow humans.

The ‘data-driven’ is an attempt to make a culture where the actions are based on data. The process involves building tools and abilities necessary for processing the data to support the choices made by humans to progress. Computers have been the mainstay in the collection and manipulation of data to produce meaningful and purposeful information.

The 14th IFIP TC9 Human Choice and Computers Conference (HCC 14) attracted several submissions with data as the pivot for human-centric progress. We present seven submissions that have been selected after careful review by the members of the Program Committee of HCC 14.

Martin Warnke reports a data-driven exploration of the connections between the internet technologies such as cloud computing and the societal governance based on a regulatory framework.

Identity in the context of electronic transactions has always been a challenging task. Rapid advances made in the recent past triggered a wide range of debates around the terms such as public, private, personal, non-personal, impersonal, and interpersonal that abstract the human in a connected social system. Kurihara Yusuke proposes the Self-Sovereign Identity model that explores this cyberspace in the context of blockchain technologies.

Vassilis Galanos and Mary Reisel illustrate the efficacy of ethics in the domains of Artificial Intelligence and Robotics using the Japanese Toy Dog AIBO as a case study. The Japanese Ethnographic Observations provide unique focus in this research.

Two data-driven approaches in Healthcare and Managing Water Supply in a social system are well reported and grounded in practice.

People and performance are closely related through an organizational structure. Dennis Grenda and Anne-Marie Tuikka report the “smart workflows” that knit the people, technology, and the organizational structure in Industry 4.0.

The design and development of technology for the differently abled has always been considered very important. A good testing method for Refreshable Braille Display has been reported from India.

The call for papers for HCC 14 had 40% of the suggested list of topics that explicitly mention data. 22% of the accepted submissions could be categorized in this section that has data as the primary focus. Data becoming sublime to the human in a majority of accepted submissions reflects the freedom of choice that makes it human-centric.

Heaven and Earth – Cloud and Territory in the Internet

Martin Warnke(✉)

Leuphana University Lüneburg, Lüneburg, Germany
warnke@leuphana.de

Abstract. When John Perry Barlow declared Cyberspace as independent from the common territory and sphere of influence by economy and state in 1996 he could not know what happened afterwards. All the dreams of immateriality and exclusion from state power had to recede from a reality of internet infrastructures that became more and more massive, resource consuming and dominant.

What developed into the metaphor of The Cloud at the same time settled as a highly influential new power rivalling with democratic institutions of the state on its territory. According to Galloway it is the new means of power, the protocol, that reigns technology and society.

The key question of this paper is if there is a close connection between the family of internet protocols and the regulatory means of state. My tentative answer is: yes, indeed, there are close connections between the two spheres that seemingly were incompatible: heaven (The Cloud) and earth (the state territory) couple strongly.

The approach is done by Luhmannian Systems Theory, the term being "interpenetration" and "structural coupling". The example will be the Chinese Social Credit System.

The main thesis is that the Chinese Social Credit System couples its protocol to the laws and regulations of state and economy, also conceptualized as being protocols, via the internet as its common language. It turns out that the coupling and interpenetration could be highly efficient thus for the first time enabling a co evolution of a highly mobile turbo capitalist economy and a strict authoritarian state territory, a marriage between heaven and earth.

Keywords: Cloud · Systems theory · Internet protocols

1 A Very Short History of the Internet

The internet, as we all know, is a result of the fear of the US government to be overtaken by the Soviets. It is a child of the cold war, an immediate effect of the so-called Sputnik Shock.

A research agency, the ARPA, did investigations for the US government on how the US could catch up to the Soviet Union. Especially if there were a nuclear war how the military still could function. Because it was known that a thermonuclear strike would lead to a nuclear electromagnetic pulse that brings down electronics like telecommunication devices. It would not be possible any more to maintain the command chain, orders from the generals would not reach the soldiers. Absolute chaos.

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 117–129, 2020.
https://doi.org/10.1007/978-3-030-62803-1_10

So, the RAND Corporation in 1959 started work for ARPA on concepts of a communication system that would probably withstand the electromagnetic pulse. It was Paul Baran who published a series of reports on this in 1964, at the same time Donald Davies had the same technical idea in Great Britain. It was the idea of the packet switching instead of the common line switching that was used for the telephone connections. The lines are too vulnerable, a strike could destroy cables, no transmission through them would then be possible before repair. Packet switching could stand losses of packets because they could be sent repeatedly.

The packet switching brought redundancy to telecommunications, and with it brought resilience, the capability to withstand damage. And it brought diversity, since the packets could travel on whatever route, through cables, over landlines, by radio, you even could implement packet switching with carrier pigeons! This immense flexibility is a signature of the internet of today. No single carrier could monopolize the infrastructure, we now talk and exchange data by WiFi, Ethernet or Cellular Phone technology, to name a few, operated by different companies. Very many kinds of gadgets could use the internet: phones, computers, watches, tablets, fridges, cars. But how could all this work? Why isn't there chaos all about?

It is by conventions. In the case of the internet and its connectivity, we call the conventions: protocols! They leave freedom for the inside by maintaining strict regulations at the borders to the outside. In the internet it's all about the protocols. Protocols take care for stuff to happen that would be very improbable to happen without them. People call concepts like these: media. Protocols like the internet protocols are a hugely important type of digital media. They allow for things to happen that would never occur without them.

At this point we should employ Luhmann's notion of media:

"We would like to call media the evolutionary achievements that enter at those possible breaks in communication and that serve in a functionally adequate way to transform what is improbable into what is probable" ([6], p. 160).

An open systems theoretical question for me, by the way, is whether the internet could be qualified a s medium of distribution alone. It also has aspects of a language and of a medium of success.

One of the basic protocols of the internet is: the internet protocol, issued by Jon Postel in 1981. The document is a so called "Request for Comments", an openly negotiated proposal. You can look up all those RFCs in the internet, these are the documents that make up a huge part of the history of the internet. It is interesting to read these documents. They are not written in a formal language but in plain english. I quote directly from RFC 791:

> 1.4. Operation
> The internet protocol implements two basic functions: addressing and fragmentation. The internet modules use the addresses carried in the internet header to transmit internet datagrams toward their destinations. The selection of a path for transmission is called routing ([10], p. 2).

Protocols address people. They are no algorithms but conventions written in natural language. And exactly this is one of the reasons why they could persist over such a long period of time: from 1981 up until today. These are at about four decades now, an eternity. These conventions all have to and can be implemented in whatever programming language that is in current use on rapidly evolving hardware.

An internet packet header looks like this:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version|  IHL  |Type of Service|          Total Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Identification        |Flags|      Fragment Offset    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Time to Live |    Protocol   |         Header Checksum       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source Address                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Destination Address                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Options                    |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fig. 1. An Internet Packet Header ([10], p. 11)

Let’s have a look at one of the fields:

> Time to Live: 8 bits
> This field indicates the maximum time the datagram is allowed to remain in the internet system. If this field contains the value zero, then the datagram must be destroyed. This field is modified in internet header processing. The time is measured in units of seconds, but since every module that processes a datagram must decrease the TTL by at least one even if it process the datagram in less than a second, the TTL must be thought of only as an upper bound on the time a datagram may exist. The intention is to cause undeliverable datagrams to be discarded, and to bound the maximum datagram lifetime ([10], p. 14).

By issuing these RFCs everybody capable to do so could influence the development of the internet. The internet became a place of innovation, of independence, a world made up of freedom and of ideas. The internet was once part of the counter culture, like LSD, and the hippies.

The interesting aspect in respect to systems theory is: a protocol constitutes a binary relation. The service works in case the protocol is obeyed, and it is denied if not followed strictly, thus it has also the nature of a code that governs connectivity, *Anschlussfähigkeit*, the very same word in English for the technical and the sociological term. Luhmann did a very similar characterization of the way structural coupling takes place between autonomous subsystems:

> Structural coupling in this context means transforming analog (simultaneous, continuous) relations into digital relations that can be handled in accordance with an either-or schema, and also intensifying certain mutual irritation channels with a high degree of indifference toward the environment ([7], p. 110).

I take this analogy as a hint that protocols have an immense importance for structural coupling, even if we do not mix up the digital with computers.

2 Grassroots

In 1996, the lyricist of the Rock Band Grateful Dead that also was part of the American counterculture, this John Perry Barlow published a manifesto to declare cyberspace independent. It starts like this:

> Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather [1].

What was the reason for such an emphatic outcry? The Clinton administration in 1996 passed a law to regulate the internet, mainly to open it up for regular business. This enraged the Internet people that wanted to be left alone. Although the beginning of this declaration still comes as a request, almost a little shyly, similar to the style of the founding documents of the Internet, the “Requests for Comments” [5], the strong utopia of an immaterial territory arose from this manifesto, which was – or at least should be – free of the power relations of late capitalism and socialism and could thus make everything new and better. The internet, according to its founders, should not be part of the state, it should remain somewhere else, at its own.

We now know, two and a half decades later, that this idea shattered to pieces.

How the Internet as a technical-social utopia relates to the state, whether it itself has a territory, if it lives in a cloud, or whether it perhaps behaves exactly the other way round, is what I want to lay out before you in the following.

3 The (Im)materiality of the Internet

> Cyberspace does not lie within your borders. Do not think that you can build it, as though it were a public construction project. You cannot. It is an act of nature and it grows itself through our collective actions. […] Cyberspace consists of transactions, relationships, and thought itself, arrayed like a standing wave in the web of our communications. Ours is a world that is both everywhere and nowhere, but it is not where bodies live. […] there is no matter here.
>
> Our identities have no bodies, so, unlike you, we cannot obtain order by physical coercion [1].

The idea of the immateriality of information technology, in particular the Internet, here once again in the version by John Perry Barlow, has often been emphatically represented, especially in art, and is one of the pillars of the phantasm of cyberspace as a domination-free area. Jean Francois Lyotard spread this idea in the legendary exhibition “Les Immaterieaux” at the Centre Pompidou in 1985, which he supervised, and later said something about it:

> The term ‘immaterial’ is now a somewhat daring neologism. This merely expresses the fact that today - and this has prevailed in all areas - the material can no longer be regarded as something that opposes a subject like an object. Scientific analyses of matter show that it is nothing more than a state of energy, i.e. a connection of elements which themselves are not tangible and are determined by structures which each have only a locally limited validity […] The increasing mutual penetration of matter and spirit – equally clearly through the use of word processing systems – now causes the classical problem of the unity of body and soul to shift [8].

This did not remain without contradiction. Horst Bredekamp, the famous art historian for example, commented on ideas of immateriality like this:

> It is an abstruse thought that a picture on a screen would be material-free. Video artists, especially those of the first generation, have pointed this out by using the television as a sculpture. The moving or non-moving images of the screens are afflicted with a logistics that exceeds the Florentine Pietà of Michelangelo many times over in material gravity ([4], pp. 39).

And so it seems at least doubtful how Barlow [1] situated the role of the digital media in the political field: "In China, Germany, France, Russia, Singapore, Italy and the United States, you are trying to ward off the virus of liberty by erecting guard posts at the frontiers of Cyberspace. These may keep out the contagion for a small time, but they will not work in a world that will soon be blanketed in bit-bearing media."

Was this utopian or naïve?

4 The Cloud and Technical Gravity: Protocols, Networks and Territories

In the debate about the Internet, the structure and logic of the Internet has been the subject of public debate, and this may have led to the false impression that it is somehow immaterial. And indeed: although computers and their functionality can be described physically within the framework of solid-state physics, they gain their epochal significance above all through the fact that they perform symbol processing. The computer deprived humankind of its special position of being solely responsible for the symbolic. Computers operate in the symbolic and exist in the real!

It is no exaggeration to say that the "protocol" is the central type of document that constitutes not only the technology but also the social and political aspects of the Internet. The term was originally coined in diplomacy, protocols describe the rules that must be observed to ensure the smooth functioning of a process. And what applies to the facilitation of diplomatic encounters even among opponents of war also applies to technical components and their competing operators: as long as one adheres to the protocol, the exchange can be continued, if one violates it, the communication breaks down, and everything stops. It is a hardly noticeable, non-violent form of domination, but it is without mercy. No complaints or appeals help: disregarding the Internet protocol, for example, is immediately punished by an error message and the termination of the data exchange.

Alexander Galloway wrote an important book about protocols. He describes how our society has reached its current state of development, and its title is: protocol [3].

> How would control exist after decentralization? In former times control was a little easier to explain. In what Michel Foucault called the sovereign societies of the classical era, characterized by centralized power and sovereign fiat, control existed as an extension of the word and deed of the master, assisted by violence and other coercive factors. Later, the disciplinary societies of the modern era took hold, replacing violence with more bureaucratic forms of command and control.
>
> Deleuze has extended this periodization into the present day by suggesting that after the disciplinary societies come the societies of control ([3], p. 3).

"[P]rotocol is how technological control exists after decentralization." ([3], p. 8). Control and thus power are maintained by protocols.

What is special about the technical protocols of connectivity, which, according to Galloway, has so strongly shaped our society today? The Internet Protocol IP, for example, regulates how the data packets that are sent across the Internet must be structured. It does not matter to the IP which technical infrastructure is used, as long as it adheres to the protocol rules. There is also no explicit coding of the geographical location of the sender. And it doesn't matter what the content of the data stream actually is, the protocol levels above regulate it, like the ones for email, for the web, or for the social media. The downward indifference subsequently made possible innovative network technologies, such as the Internet via mobile telephony, which is now used in every smartphone and was not yet conceivable in 1983; the upward indifference finally made possible content and functions that are dominant today but still unthinkable in 1983: social media, television on the net, surveillance of the entire population.

So, do we live basically in a virtual disciplinary society with voluntary discipline through smart self-monitoring techniques – in other words, an incorporation of the panoptic principle into the individual? Probably not, since the differences between control and discipline are too great, as are the differences between voluntariness and coercion. The observance of protocols expands possibilities, violence limits them. Protocols appeal to desire, coercion to the mind. The techniques of self-optimization seduce, laws force us. Protocols disappear from our attention – also a typical characteristic of a medium –, against orders one can rebel. In Frieder Nake's categorization, protocols act on the "subface", which, unlike the surface, remains invisible:

"The surface is visible. It is for us. The subface is invisible. It is for the computers. The computer can change the subface, it can manipulate it. The surface does not have this property" ([9], p. 3).

The crucial point for the relationship between the Internet and the state seems to me to be precisely the quality that is at stake here: the complete indifference of Internet technology towards its political environment. This has nourished the idea that the Internet is somewhere, nowhere: in The Cloud, not on the earth. On the one hand, this has led so many founding mothers and fathers of the Internet to egalitarian grassroots fantasies, but on the other hand it has also led to the current situation in which the Internet appears both as a driving force of social change and as a means of repression for authoritarian regimes. The Internet simply doesn't care whom it serves. And it doesn't matter to it that it promotes all communication processes and the concentration of power as rapidly as no media technology before it [13]. As a communications infrastructure, the Internet does not favor any form of society or state; its connectivity acts more like a catalyst or accelerator of social development, supporting and empowering those who design and operate it.

The fact that there is at all a connection between state power and Internet technology is not due to the rules of protocol, but to the operation of the network itself, which we are now coming to.

All those who still believe in the immateriality of the Internet are recommended to visit a data centre. These are highly secured industrial plants that consume an enormous amount of electrical energy and therefore have an enormous need for cooling, which not only operate the central servers that store and organize the content, but also mediate and regulate data traffic, all nodes of the network. Every node on the Internet that governs the data flow for a subnet marks the relationship between inside and outside for this subnet. When the Internet was built, the protocols did not regulate national responsibilities, nobody even thought that a data packet should take a different route than the fastest one from sender to recipient. In every switch or router, there are address and connection directories, routing tables for this purpose, on the basis of which the routing in the Internet happens automatically, unaffected by the borders of a territory.

But the world did not remain libertarian and hippy-like as imagined by the nerds who invented the Internet in the first place. The active network components, the switches and routers, are located somewhere and are operated by someone, and the indifferent protocols do not care whether the routing tables follow political rules of geographical restrictions at borders or whether there are economic goals of companies or whether simply the original efficiency of data traffic should be the main idea. If you request a website in a country with Internet external border control that is inadmissible in this state, then there is no technical connection to the associated network components in the accessible routers, and the query runs into the void. If right now memories of your last visit to China come to your mind then this is significant. Even circumvention practices, tunneling mechanisms such as VPN or proxy servers could be blocked by the routers. This is how the Internet is earthed. And this grounding characterizes the respective variety of the local Internet: state-regulated vs. neo-liberally economically oriented or net-neutral indifferent according to its founding idea.

The Internet has its head in The Cloud and its feet firmly on the ground.

John Perry Barlow observed the Internet as a cloud in contrast to the background of the territory ([11]):

$$\text{internet} = \overline{\text{cloud}\,\Big|}\;\text{territory}$$

From the point of view of the operators of terrestrial networks it rather looks like this:

$$\text{internet} = \overline{\text{territory}\,\Big|}\;\text{cloud}$$

I use Spencer Brown’s forms in the metaphorical way as Dirk Baecker does. It is well a matter of dispute whether I should do so, but this very condensed, almost poetic form has a certain appeal to me.

5 The Many Borders of the Internet: The Case of China’s Social Credit System

Because of the functional differentiation of a modern society, state borders and the borders of the functional subsystems of society no longer fall into one. While in the disciplinary societies with strictly hierarchical organization all spheres of power were united under the absolute ruler, the king or emperor, and the borders of the empire were also those of trade, political power and law. But the control societies built their own subsystems with their own borders: languages marked nations, trade is getting global by complex regulatory institutions at the state borders: customs, immigration, international telecommunications organisations, capital flow seems to have almost no limits at all.

Niklas Luhmann writes:

> The distinction […] between the environment as a whole and the systems in a system's environment, explains how boundaries are put under pressure to improve their performance, that is, explains how a more exacting determination and preservation of boundaries becomes necessary. System boundaries always separate out an environment, but the requirements for this vary if the system must distinguish other systems (and their environments) within its own environment and adjust its boundaries to this distinction. In the simplest case, the system treats its environment as another system. Thus national boundaries are frequently conceived as boundaries with another nation. But this becomes increasingly illusory when relations with an economic, political, scientific, or educational 'abroad' no longer correspond to these same national boundaries. Under such circumstances, the boundary definition moves inside; this is confirmed in self-referentially closed systems, which determine their boundaries by their mode of operation and mediate all contact with the environment through other levels of reality ([6], p. 30).

The tension that we can observe between statehood and technical communication infrastructure is that between conflicting boundaries between systems and their environments. Boundaries make themselves felt very concretely where, for example, commercial enterprises need an Internet that functions globally and is only committed to economic needs in order to coordinate their global research and development and marketing activities, but where they come up against state boundaries that prevent a libertine exchange of data, for example because they open a dependency in the People's Republic of China. This creates boundaries within the subsystems: management must be staffed with party functionaries, for example, and communication must be interrupted because of state censorship, which runs counter to the autonomy of the economic system. The idea of the cloud where information is meant to live hits the reality of a state territory quite roughly.

In a decentralised world society, whose technical constitution is regulated by protocols, it can be speculated, however, whether the idea of protocol might not bring the solution here. In any case, a digital data optimism of Barlowian coinage is no longer a rational position that can be maintained.

Now to our small case study. It deals with the Chinese "Social Credit System", planned for 2020, that works with precisely the same digital media that were supposed to be the absolute contrary to any state power in the Declaration of Independence of Cyberspace 1996. Interestingly enough, the term "credit" is central here, as it also used by Luhmann when pointing out that central control could not work for a differentiated society:

> Finally, there was and remained world society, in which all this was expected to take place within territorial boundaries, a functionally differentiated system that owed its own effectiveness to the autonomy of functional systems and was not to be combined with any sort of central control. This applied above all to the credit system of international finance, which could guarantee a certain flexibility in the timing of investment and consumption, which was crucial for growing regional development. Naturally, it also applied to international affairs, to scientific research, and not least to everything that interested intellectuals. Any insistence on

> organizationally controlled, regional autonomies is quite simply not compatible with this. The attempt to introduce this sort of modernization in modern society had to be at the cost of stagnation, a drain on last power resources, diminishing acceptance, and finally the collapse of the 'system' ([7], p. 307).

So, we could wonder: how could a modernization of the Chinese society work out without central control for which it acquired some fame from its communist history? My impression is: by a protocol.

The whole thing, the Social Credit System, is based on the implementation of a protocol that is absolutely not socially indifferent because it regulates social issues precisely and absolutely rationally. As you might know, it intends to implement the massive collection of very personal data from all sources, its valuation into social credit points from which, in the end, positive or negative feedback is generated by the power of the state. All this is technically based on the connectivity of the internet. It is a social protocol, one that works on the basis of all the politically blind technical Internet protocols invented by the nerds and hippies in the 1980s, grounded by sovereign control over the network infrastructure on state territory.

The fact that the Chinese Social Credit System is a phenomenon of the involvement of political and economic boundaries is stated explicitly:

> Our country is in a period of expansion in which the openness levels of the economy are rising on an even greater scale, across even broader fields, and at even deeper levels. Economic globalization has enabled an incessant increase of our country's openness towards the world, and economic and social interaction with other countries and regions is becoming ever closer. Perfecting the social credit system is a necessary condition to deepen international cooperation and exchange, establishing international brands and reputations, reducing foreign-related transaction costs, and improving the country's soft power and international influence, and is an urgent requirement to promote the establishment of an objective, fair, reasonable and balanced international credit rating system, to adapt to the new circumstances of globalization, and master new globalized structures [12].

The social credit system is called a "necessary condition", i.e. the indispensable prerequisite for the territorial borders of the Chinese state being able to cross the system environment boundary of the global market in The Cloud without dissolving both interpenetrating systems through lethal boundary damage.

The claim of the planning is truly comprehensive, here it is not only about financial creditworthiness, as we know it from the German Schufa, here the etymology of the word "credit" of credo, the credibility in the others, again comes to honor. The beginning of the declaration of the Council of State of the People's Republic of China reads:

> A social credit system is an important component part of the Socialist market economy system and the social governance system. It is founded on laws, regulations, standards and charters, it is based on a complete network covering the credit records of members of society and credit infrastructure, it is supported by the lawful application of credit information and a credit services system, its inherent requirements are establishing the idea of an sincerity culture, and carrying forward sincerity and traditional virtues, it uses encouragement to keep trust and constraints against breaking trust as incentive mechanisms, and its objective is raising the honest mentality and credit levels of the entire society. And Completely moving the construction of a social credit system forward is an effective method to strengthen social sincerity, stimulate mutual trust in society, and reducing social contradictions, and is an urgent

> requirement for strengthening and innovating social governance, and building a Socialist harmonious society [12].

Sincerity is pretty much the most commonly used word in this text. It is the positive side of a code to be enforced through the social credit system. It is not only about finances, but also about morality, about the values of a desirable state, and this in all conceivable fields of society. The list is long and tiring, ranging from the allocation of credit lines to the correct behavior of civil servants, care of the elderly, science, education and tourism. Incentives should be given, and violations should be punished. For road traffic there is the passage: "Enter citizens' traffic safety and law-breaking situation into sincerity files, stimulate all members of society to raise their consciousness about traffic security", and to this end we imagine the massive surveillance of the Chinese public space with video cameras that support individual facial recognition.

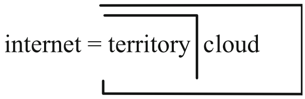
The decisive question is probably whether it can actually come to the co-evolution of socialism and the market, a development that no classical theory of society or economy thought possible and which can perhaps only be realized through a hierarchy of technical and social protocols. The technical medium of the interpenetration emerging here seems to be the Internet and its global connectivity by protocols, which allows a coupling both to politics and to the economy, precisely because of its blindness to both spheres.

The market capitalization of Internet companies alone shows that the Internet is perfectly linked to the economy. But that it also offers perfect connectivity to (police) state methods of surveillance is now becoming obvious through the profiling by the social media, the omnipresence of geo localization and the voluntary body searches by self-optimizers, all three notions stemming from a typical Police Recognition Department but also part of our digital lives. We thus have a classic case of the interpenetration of two functional subsystems of society, and the medium that couples both to the state and to the economy is the Internet: grounded to the territory, global like The Cloud. Here issues of theory once again emerge. How could we name this? Is this a matter of language as in the case of other system interpenetrations?

The SCS is supposed to appear impersonal and silent and invisible like Kafka's Schloss bureaucracy. People and faces appear when you look at the members of the Chinese State Council, and even then, you can't get rid of the impression of watching masks.

China is also technically exactly the right nation for such interpenetration: the state is autocratic, and the IT infrastructure is a monoculture. Almost everything is paid with one of the very few online payment systems, mostly in WeChat, the monopolistic social medium, where all photos, all chats, all financial transactions, all geodata are assembled and state monitored. The design and use of the Internet are becoming a state goal, and the State Council of the Communist Party of China has adopted such a goal in 2014 in the form of the social credit system, as already mentioned. Its reign shall be silent and unshakable. Violations are punished non-violently by the fact that social participation becomes impossible in a relentless binary manner. Anyone who can no longer get a loan from a bank because of insufficient social credit, can no longer buy train or plane tickets, can no longer rent an apartment and can no longer get a job, has got himself a social error message and the termination of the social process, impersonal, merciless, inescapable. The parallelism in the functioning of the social credit system with the technical protocols of the Internet is obvious. On the other hand, it differs from Bentham's Panoptikon, as Foucault described it in his "Surveillance and Punishment", in that it takes place invisibly and without physical coercion, on the subface. Once the social credit system planned for 2020 has been implemented, participation in social life will be tied to submission to it. It works piquantly with the very digital media that Barlow's declaration of independence in 1996 was supposed to guarantee: separation from the state and individual freedom.

China's view on the internet could be characterized as the unity of the difference between territory and cloud, a re-entry in the terminology by Spencer Brown:



6 The Internet as a State Signature

If we try a comparison of systems, if we distinguish Far Eastern socialist state capitalism from the liberal parliamentary democracies of Europe with functioning informational self-determination and, finally, from the unbounded neoliberal USA, which permits an unbridled land seizure by Internet corporations, the central importance of state regulation or non-regulation of the Internet for a state constitution becomes obvious. Due to the fundamental significance of the Internet with its protocols, its grounding in the territory and its mobility in The Cloud, its function as a coupling medium between autonomous social subsystems, it is almost the signature of a state whether and how it regulates its Internet infrastructure. For, according to Dirk Baecker: "And […] the next society? Since the introduction of the computer, their problem is no longer the surplus of criticism, but the surplus of control." ([2], p. 169. Transl. MW) And truly: functional subsystems can be connected to each other via the Internet and controlled via cross system databases. Machine learning and artificial intelligence play the role of automatic omnipresent guards at the data interfaces. The Internet is the completion of the decentralized control society. It also marks the unity of the contradiction between centrality and decentrality, between seduction and control. And this is precisely why societies and their states differ fundamentally in the way they deal with control. The achievement of informational self determination then becomes the antithesis of a market liberal data liberation as well as a total control and sanctioning of state capitalism.

The platforms, the surfaces of the Big Data subfaces, on which functions of the subsystems meet, play a special role here: Google, Facebook, Twitter, Amazon and the People's Republic of China's platforms like Baidu, Renren, WeChat and Alibaba. Whoever controls them, who have grown so boisterously on the Internet, controls a crucial infrastructure of society. But that also means: if, e.g., we Germans really want to enforce our German constitution, the Grundgesetz, and thus preserve the article on freedom of the press, Article 5 GG, or the article on postal and telecommunications secrecy, Article 10 GG, then we have to deal with the social media platforms! The Network Enforcement Act, Netzwerkdurchsetzungsgesetz, of 2017 has thus made a start, and the outcry from parts of the network industry has confirmed that the legislator has made a decisive point. But this will probably not be enough, as the example of the Social Credit System of the People's Republic of China shows, which can only function because a platform supported vertically consistent and monopolistic collection of the data up to the provision of the services with the right of access of the state power enables the control of all social subsystems equally. The separation of powers will also have to be called: platform splitting!

The inter-, counter- and superimposition of globally and territorially operating platforms on the basis of Internet based data flows creates new border relationships that differentiate themselves, crossing each other autonomously. State borders have become tremendously complex, they can no longer be drawn on political maps, they not only end at border fences, but also at access possibilities to data infrastructures, right up to the chips that are built into network components. State hacking is an act of territorial border violation that is becoming increasingly apparent. A citizen remains under the control of their state platforms even abroad when using them. And how should they not!

And so Perry Barlow was right in a way, despite all the failure of hopes: so that cyberspace could have remained a place of mind and freedom, the state and the weary giants of flesh and steel, and to add: of data, should have kept away.

But they didn't.

References

1. Barlow, J.P.: A Declaration of the Independence of Cyberspace (1996). https://www.eff.org/cyberspace-independence. Accessed 1 Oct 2018
2. Baecker, D.: Studien zur nächsten Gesellschaft. Frankfurt am Main (2007)
3. Galloway, A.R.: Protocol. How Control Exists after Decentralization (2004)
4. Huber, H.D., Kerscher, G.: Kunstgeschichte im "Iconic Turn" – Ein Interview mit Horst Bredekamp. In: Huber, H.D. (ed.) kritische Berichte. Zeitschrift für Kunst- und Kulturwissenschaften, Sonderheft Netzkunst. **26**(1). Transl. by MW (1998)
5. Internet Society (2018). https://www.rfc-editor.org/
6. Luhmann, N.: Social Systems. Stanford. Originally: 1984: Soziale Systeme. Grundriß einer allgemeinen Theorie. Frankfurt am Main (1995)
7. Luhmann, N.: Theory of Society, vol. 2, transl. by Barrett, Rhodes. Stanford. orig. Die Gesellschaft der Gesellschaft (1998)
8. Lyotard, J.F.: In kunstforum international. Köln, June/July 1990. https://www.kunstforum.de/artikel/jean-francois-lyotard/. Accessed 21 June 2020. Transl. by MW
9. Nake, F.: Zeigen, Zeichnen und Zeichen: Der verschwundene Lichtgriffel. In: Hellige, H.D. (ed.) Mensch-Computer-Interface: Zur Geschichte und Zukunft der Computerbedienung, pp. 121–154 (2008). Transl. by MW
10. Postel, J.: Internet Protocol. Darpa Internet Program. Protocol Specification (1981). https://tools.ietf.org/html/rfc791
11. Spencer-Brown, G.: Laws of Form. London (1969)

12. State Council of the People's Republic of China. 社会信用体系建设规划纲要 (2014). http://www.gov.cn/zhengce/content/2014-06/27/content_8913.htm. Official translation: Creemers, Rogier. https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/. Accessed 1 Oct 2018. I very much thank Prof. Martin Woesler for his support on all of this
13. Warnke, M.: Databases as citadels in the web 2.0. In: Lovink, G., Rasch, M. (Hrsg.): Unlike Us Reader, Social Media Monopolies and Their Alternatives, Amsterdam (2013)

Self-Sovereign Identity and Blockchain-Based Content Management

Yusuke Kurihara[1,2](✉)

[1] Graduate School of Media and Governance, Keio University, Kanagawa, Japan
yukurihara.cyberlaw@gmail.com
[2] InfoCom Research, Inc., Tokyo, Japan

Abstract. The concept that a person aims to control my own identity without the intervention of a controlling entity is called self-sovereign identity (SSI). The SSI is an antithesis to the phenomenon that certain companies and organizations, called digital platformers, collect data centrally.

This paper expands the concept of SSI and describes the possibilities and problems when applying it to self-content management using the concept of Self-Content Management (SCM). In particular, I discuss the possibility of self-sovereign management of digital content, and discuss DRM using blockchain technology as a means.

This paper is assumed that SSI can be applied to content created by myself, but by properly managing content using blockchain technology, it is possible to complete licenses and levy fees. I also clarified that it could be a substitute for the resale right.

As a result, SCM promoted the problem of "orphans' works" and facilitated the processing of rights, and revealed that it could contribute to the further development of culture. However, I also address that there are two issues. One is the lack of institutional trust, rather than the technical credibility of the blockchain. The second is consistency with the Attorney law of Japan.

Keywords: Self-Sovereign Identity (SSI) · Self-Content Management (SCM) · The Blockchain-Based Digital Right Management (DRM)

1 Introduction

1.1 Personal Data, Know Your Customer and Self-Sovereign Identity

When you buy cigarettes or alcohol at a convenience store, a clerk asks you to press an adult confirmation button in Japan's regulations. When purchasing online, you enter your password, personal account info, and address and credit card number to receive a delivery at home. This is just one example of various daily services that require the presentation of an individual's identity in a verification called Know Your Customer. This information, including purchase history, is then collected by a company, regardless of whether customers are conscious or not, and analyzed and used for direct advertising.

In the digital society, the management of digital identities is becoming increasingly important as sensitive information regarding things like financial accounts and medical

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 130–140, 2020.
https://doi.org/10.1007/978-3-030-62803-1_11

care is digitized. The “information bank” has become a hot topic in Japan.[1] The concept of “controllability” is based on the concept of companies and organizations becoming central management organizations to collect and utilize data. Consumers receive a reward equivalent to the provided personal data (which is sometimes used for the public interest).

However, what central management organizations collect individual digital identity is risk because digital security is a concern, and even trusted international companies have experienced large-scale leaks. Regulating agencies have become more involved in investigating and applying sanctions to companies involved in data breaches. In particular, general data protection regulation (GDPR), which came into force in Japan in May 2018, is prominent. The maximum fine imposed for a violation of the GDPR is (1) €1,000 million or 2% of total sales for the previous fiscal year, whichever is higher; or (2) €2,000 million or 4% of total sales. Since its implementation, the GDPR has led to significant fines, as published by national data protection agencies [1].

While there is no comprehensive data protection law in the United States, Chinese video application TikTok settled with the U.S. Federal Trade Commission in 2019 through a $5.7 million fine for allegedly obtaining personal data of children without parental consent in violation of COPPA (The Children’s Online Privacy Protection Act) [2].

In recent years, data protection legislation has grown in popularity. CCPA (The California Consumer Privacy Act) has been in force since January 2020. African countries such as South Africa and Egypt have enacted data protection laws similar to the GDPR, while Association of Southeast Asian Nations countries such as Singapore and Indonesia have also enacted new data protection laws in recent years. In Thailand, data protection legislation similar to the GDPR has been introduced in Parliament. In China and the Philippines, data localization, which mandates the domestic preservation of personal information and important infrastructure data, is stipulated. This proves the importance of the data and the regulation against the threat of data concentration in the enterprise.

Alternately, there is the concept of self-sovereign identity (SSI). This concept is intended to allow individuals to control their own identity without the intervention of a management body [3]. As an antithesis to the centralization of data in businesses and governments, it is attracting attention as one of the ideal forms of digital identity.

1.2 Non-Personal Data/SSI/Self-Content Management

In Europe, a distinction is made between personal and non-personal data, such as industrial information. In Europe, the GDPR states that the protection of personal, but not non-personal, data is a fundamental right. However, to ensure the free flow of information, the “Framework for Free Distribution of Non-Personal Data in the EU

[1] An information bank reliable entity to which a person delegates the provision of personal information to a third party to the extent that the person agrees, with the aim of promoting the distribution and use of personal data by enhancing effective personal involvement (controllability) (Ministry of Economy, Trade and Industry Study Group on Approaches to Certification Schemes for Information Trust Functions “Guideline on Authorization of Information Trust Functions ver. 2.0 [October 2019]”).

(REGULATION (EU) 2018/1807 OF EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 November 2018 on a framework for the free flow of non-personal data in the European Union)" prohibits requiring the storage and processing of data within the territory of a specific country or preventing the storage and processing of data within other countries.

The collection of industrial data also assumes that the distribution of information is an essential aspect of corporate globalization. However, if the concept of SSI is applied to non-personal data, the individual should be able to control whether the data is distributed freely. It should be up to the individual to decide whether to distribute his or her non-personal data abroad even when the company's servers are abroad.

One example of this issue is music ownership. In Japan, most music copyrights have been transferred to JASRAC in trust, and JASRAC has centralized management. It has eased facilitation of the rights management between music users and numerous composers and songwriters and has also aided in making appropriate allocations. But outside of music, things aren't going well. Content created by a professional is a manifestation of his or her own personality, and his or her own work should constitute non-personal data, based on the SSI philosophy, so that he or she can have appropriate control and involvement over the data.

In recent years, blockchain technology has been replacing centralized content management [4]. Content management is also getting attention from the content market [5]. According to the network externality of the platform service, the blockchain's copyright management system may lead to a paradigm shift in the centralized copyright management model and content delivery methods of JASRAC [6] using the digital rights management (DRM) 2.0 that SSIs bring [7].

In this paper, therefore, we examine the prospects and challenges of content creators using blockchain technology to appropriately implement SCM (Self-Content Management) based on SSI.

2 Self-Sovereign Identity (SSI)

2.1 Concept of SSI as Ideal Models for Secure Data Protection

In short, self-governing identity means that one's own identity is one's own management. Most service providers, including the four [8] biggest providers—Google, Apple, Facebook, and Amazon—collect and use some personal data, and they maintain and manage this data centrally. Naturally, this data use is optional. Google's privacy policy, for example, states that "You can adjust your privacy settings to control what we collect and how your information is used." Thus, "controllability" is left to the user. Also, in the Google Terms of Service, there is in the item of "Modifying and Terminating our Service" that reads, "We believe that you own your data and preserving your access to such data is important." The idea of digital identity is an antithesis to the centralization and use of data by companies.

With regard to data, in Japan it is used "ownership" not "Syoyu-ken" which means same definition of ownership in Japanese. However, it is well known that ownership of data cannot be conceived. Also, if there's really "ownership" in the data and it's accessible, then there's no need for Google to get involved, nor is there a need for such a sentence. This statement shows that there are principles and practices.

Christopher Allen, a leading proponent of SSI, suggests that "A vision of how we can build trust in our digital identities while maintaining individual privacy" is SSI [9].

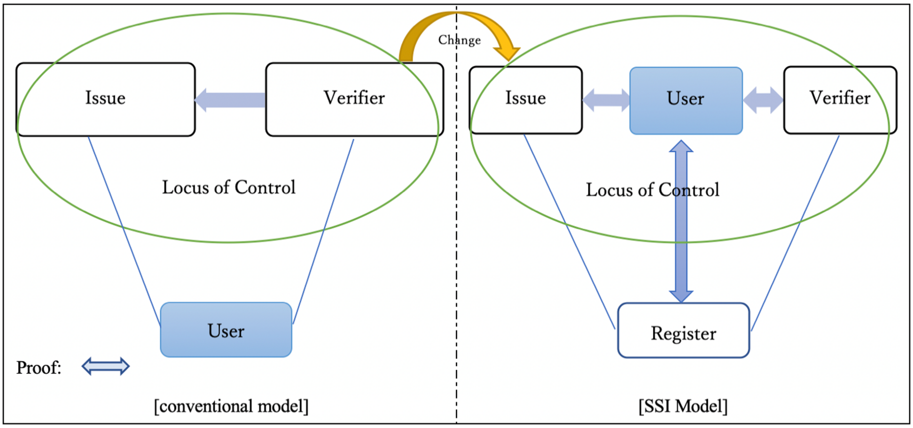


Fig. 1. Changes in the model for KYC (From conventional to SSI) [10]

As shown in Fig. 1, the user is at the center of SSI. In the conventional model, the user is not involved in managing his or her own identity, and authentication can only occur between the issuer and the authenticator. In the SSI model, the user is actively involved in managing his or her own identity, plays a central role, and without user authentication, no personal data is transferred between the issuer and the authenticator.

The concept of SSI is based on the fact that personal data are subject to governance without external factors. The following six characteristics are listed [11]:

1. Complete control of the data
2. Ensuring the security and privacy of users' personal data
3. Full data portability
4. No need for trust in central institutions
5. Ensuring data integration
6. Maintaining transparency of personal data

This concept also attracts attention because it is expected that the incidents will be reduced because companies are not centrally storing data [12].

2.2 Social Implementation Level

The SSI model is gaining attention because it provides a solution to data protection legislation such as the GDPR rather than merely providing a strong sense of empathy and support for ideas. Blockchain technology is attracting attention as a platform for social implementation of these solutions. Blockchain technology is a peer-to-peer

(P2P)-based consensus-building system that does not require a central authority and does not apply only to cryptographic assets. P2P-based blockchain technology acts as a trusted third-party replacement because, in the SSI model, blockchain technology satisfies the six elements described in Sect. 2.1.[2]

In Europe, the European Commission established the European Union Blockchain Observatory and Form in February 2018. In May 2019, a theme report entitled "Blockchain and Digital Identity" was issued [13]. It emphasizes a blockchain technology-based concept centered on personal data with autonomous and distributed identities in the EU. In Europe, personal data are defined as "right to be forgotten" and data access rights and automatic processing provisions in the GDPR as the right to ensure a person's identity.[3] In the case of blockchains, there are aspects that are difficult to implement, especially in relation to the right to correct or forget. In other words, the technical feature that it cannot be modified after the fact is a means of ensuring credibility. Therefore, in principle, it cannot be addressed in relation to the right to correct or to be forgotten.[4] It also includes the smart contract [14], which eliminates the possibility of external factors intervening in the realization of an electronic protocol for handling the terms and conditions of a contract, and blockchain is applicable [15]. Therefore, the SSI model is based on blockchain technology.

Blockchains are not centralized, but they are P2P based and are sometimes referred to as "Decentralized Identity (DID)". In the same way as SSI, the user keeps his or her own personal data but cooperates within the range permitted by the user.

2.3 ERC 725/ERC 735

In fact, the concept of DID is indispensable for the social implementation of SSI. Section 2.3 describes several use cases (See, Table 1). In one of the use cases, Ethereum developer Fabian Vogelsteller presented a draft SSI standard on GitHub called Ethereum Request for (ERC) 725 [16]. ERC 735 is the relevant standard for adding and deleting claims of ID Smart Contracts in ERC 725.

[2] Abraham [11] at 7.

[3] European Union Blockchain Observatory and Form [13] at 19.

[4] Turning away from this discussion, the GDPR response in the blockchain, in particular the right to correct and the right to be forgotten, can be cleared by:

In the real estate registration records of Japan, correction before and after correction, such as ex officio correction, is distinguished by an underline, so this should be followed. In other words, before and after the correction are clearly indicated. Since it is practically impossible to directly rewrite the original data before correction on a blockchain basis, the corrected information is written by specifying the corrected part.

Next, in relation to the right to be forgotten, it is necessary to ensure that no corrections are left even in the right to correct mentioned above. For the time being, this response could be addressed by not granting any access privileges on a blockchain-based basis. However, if it is not observable ex post facto, the reliability cannot be maintained, so it is necessary to separate the data groups that exercised the right to be forgotten. In this way, it is possible to solve the problem. However, it is necessary to examine whether leaving traces of the exercise of the right to be forgotten is sufficient guarantee of the right to be forgotten.

ERC 725 can be defined to assign IDs not only to people but also to organizations, devices, software, etc. and can be considered one of the social implementations of SSI. In addition, ERC 725 allows the transfer of assets and rights as well as personal data.

1. ERC 725–v2: Proxy Account Standards [17]
2. ERC 734: Key Management Standards [18]
3. ERC 735: Claim Holder Standard [19]

Table 1. The use case of ERC 725/735 [20]

Project name	Use case description
Origin	ERC 725 is used as the basis for Origin's user ID. ERC 725 was selected for interoperability optimization. Origin users have a public profile that includes proof of email, phone number, social networking service account, etc.
Caelum Labs	ERC 725 will be used as a basic local and national identity system for Ethereum-based blockchains. The main goal is to use it to establish trust between people, companies, and governments, so smart contracts can be used in any process between them
Hydro	Hydro plans to provide a platform layer to facilitate and improve aspects of the dApp development process, including identity management. Accordingly, Hydro plans to offer developers a standard implementation of ERC 725 in conjunction with other identity standards that can meet the specific needs of its applications. These identity standards are detailed in the ERC 1484 reference repo
LydianID	Lydian Ventures is using ERC 725/735 identities and investigations on several projects and created LydianID as an identity management dApp. LydianID is currently being used by academic institutions to certify on blockchain the studies of their students, who can share the attestation with others. ERC 725/735 identities are used in other projects where interoperability among several parties is required and a role-permission system is built with investigations
Smilo	Smilo is using ERC 725 identities and off-chain claims including Facial Biometrics. Our first implementation contains festival/opera/station facial authentication based on a DID + ticket number. We want to collaborate on a global standard to make interoperability work, and PII is shared in a secure way

2.4 Degree and Course Registration Certificate

Globally, blockchain technology is being applied to school records, such as degree certificates, used to fake academic credentials. In Japan, a survey report on the applicability of blockchain technology in universities and research institutions was compiled in April 2019 on the themes of degrees, academic records, and research data [20]. In December 2019, Muroran Institute of Technology and NTT West announced a joint research and development program in order to promote efforts to solve problems in the development of recurrent education in the future [21].

According to this NTT West's press release [22], that shows compliance with the GDPR and other regulations. However, the necessity of complying with the GDPR is not necessarily high in the case of educational attestation certificates in Japan. However, there is a possibility that issuers will disappear due to consolidation. It is an effective means to avoid such risks and to enhance credibility. It is also advantageous for school management in terms of management costs.

3 How Does SSI Fit into DRM?

3.1 DRM and Blockchain Technology

Blockchain DRM for non-personal data mainly refers to copyright management, but contents may also include music and photos. Experimental blockchain-based DRM has existed since 2015 [23]. In February 2019, JASRAC said it would introduce blockchain into its music royalty management system. Table 2 lists several current instances of blockchain DRM use around the world.

Table 2. The use case of DRM with blockchain-based systems

Name	Summary
Ascribe (Germany)	Art work management service. Artists can easily register their work and manage ownership transfers, loans, and sales history with a work certificate issued by the company, which provides a means to deliver a work that meets the artist's needs
Bound (Americas)	Blockchain image rights management platform. If an image is registered in a procedure similar to a social media posting, a unique certificate is issued, and a legitimate right can be claimed through the certificate
KodakOne (Japan)	Blockchain image rights management platform. It automatically monitors websites for unlicensed use of registered images and, if found, executes licensing (smart contract). It also automates the purchase of photos and can prove photo rights
Startbahn (Japan)	Blockchain artwork management system. This service records, manages and shares the history of works of art using the blockchain, tracking who buys art and returning a portion of the transaction value to the artist for each resale

3.2 SCM: Potential SSI for Content

There is a view that "creation", which is one of the requirements for works under the Copyright Act of Japan, is an expression of individuality. The expression of individuality means that a work has one identity, or a similar nature, representing the individual. That is why the moral rights of authors include the right to publish, indicate one's name, and maintain integrity. If the content is distributed, the right to publish is likely to be waived, but the right not to request the name of the author and not to alter the content should be protected even for digital content.

In addition, the copyright holder must be guaranteed the rights of reproduction, public transmission, and adaptation, which are among the subsidiary rights of copyright, because reproduction and modification are easy in a digital environment.

Therefore, in order to realize SCM, it is desirable to include direct contract at the outset. The centralized management system of JASRAC comes from the complexity of rights processing. The smart contract allows the company to handle a larger number of theoretical direct contracts, and the technology that supports it is positioned as a blockchain.

SCM is positioned as a means of realizing SSI's philosophy in the content field. This SCM is the realization of the fixation of subjects and objects in information goods. Koichiro Hayashi once advocated a "digital creation rights" as a copyright system for the digital age, as digitization has made it easier to reproduce content [24]. In addition, in terms of the rights to personal data, a comparison of the ownership approach and the copyright approach has been made from the perspective of the subject and object of information [25].

In conclusion, Hayashi claims the following; The developing digital environment has created this series of trends as consumers become aware that information goods can be freely and easily reproduced, that it becomes difficult to identify an object by the circulation, and by the circulation, the subject of the information goods becomes unclear.

Criticizing the centralization of traditional DRM, Lessig worried about the coming of a controlled society. Lessig's concept of commons is set as an antithesis to the managed society.

The theories proposed by Hayashi and Lessig were tools for the purpose of balancing the return of copyright privileges to creators with free content usage. The SCM fulfills these objectives. In effect, the purpose is realized by the architecture based on the current system.

In other words, according to Hayashi's theory, SCM based on SSI aims to achieve complete control on the premise that the creator takes the initiative and owns the content he or she creates. This control means that the licensing process, including the collection of license fees, is completely self-contained without the intervention of an agent. There is no need for a centralized governing body such as JASRAC, and because it is based on SSI, it is not "administered," as Lessig criticizes.

The SCM is also about blockchain-based, leak-free rights management or just compensation. In the past, in France, the patron system collapsed and painters could not make a living, so the resale right was born; the architecture of the blockchain mirrors the resale right. (At present, the resale right is not permitted in Japan, although there is an opinion for guaranteeing the legitimate interests of the artist that it should be introduced.) The SCM also provides an alternative to the resale right.

In addition, Chang points to the possibility of direct blockchain-based management.[5] However, she claims the blockchain-based DRM can solve the technical issue but it remains the problem of the commercial practice. She points out that this could help

[5] Yeyoung [26] at 255.

eliminate Orphan Works. From a broad perspective, solving this problem will also help eliminate the difficulty of rights processing in digital archiving of cultural resources.

4 Conclusion

The GDPR focuses on the sovereignty of data subjects. I explained that the corresponding blockchain technology, as well as efforts and use cases to implement it, is being discussed in the EU. DRM is this application. Until now, however, it has been difficult for creators to keep their data and manage the rights to the content they create due to the many legal and technical challenges. The concept of SSI applies to both personal data and content, which have become commonplace utilizations.

However, the following two points remain to be addressed. First, blockchain-based rights management is concerned with the "trust" of certificates issued at the time of registration. This trust is not sufficient if blockchain systems cleared the technical standard. In the absence of social recognition, there are no proof documents to use in court. Therefore, the remaining challenge to achieve SSI is not the establishment of laws but the creation of a "trust." If this trust comes from government certification, it could be back to the centralized systems for gathering personal data and content. Whether managed by a third party or by industry guidelines, best practices will be an issue for the future.

Second, there are several legal issues. Problems remain with Legal Tech (the information technology to assist the legal affairs)'s relationship with Section 72 of the Attorney Act of Japan and Section 109 the Attorney Act of Korea(provisions prohibiting, as a business, the provision of legal services, etc. for the purpose of receiving remuneration in relation to another person's legal case) [27]. Lately, in the Cologne district court in Germany, legal tech was found to be in violation of the legal services act (Rechtsdienstleistungsgesetz), which has similar provisions (LG Köln, Urt, v. 8.1.2019 - 33 O 35/19).

The concept of DID addresses the right to "own," so there are few problems with strangeness. However, in a consortium-style SCM based on the SSI concept, strangeness becomes a problem with the enforcement of other members. Furthermore, enforcement and rights disposition are problematic in relation to the nature of the case. For details, see Matsuo's thesis [27]. However, the paper concludes that there are separate considerations for services, and this is also a future issue.

Despite the problems described above, the concept of SCM is important in preserving creative works for future generations. In the digital archive of cultural resources, the problem in the rights-handling process is orphan works. Because of increasing numbers of orphan copyrighted work and more unusable content, it is a tragedy for copyright holders and authors that the lack of a copyright holder leads to a decline in culture. It is necessary to recognize that rights management by SCM leads to "cultural development" (Article 1 of the Copyright Law of Japan). We hope that this paper will contribute to the development of a society in which individuals have an identity for content and respect it, just as SSIs do.

Acknowledgments. The author wishes to acknowledge Dr. Kunifumi Saito, Associate professor of Law Faculty of Policy Management, Keio University, for his help in interpreting the significance of the results of this study.

References

1. Kurihara, Y.: Regulatory enforcement cases and analysis after one year of implementation of the GDPR. InfoCom T & S World Trend Report 365, pp. 29–33 (2019)
2. FTC Website: Video Social Networking App Musical.ly Agreements to Settle FTC Allegations That it Violated Children's Privacy Law. https://www.ftc.gov/news-events/press-releases/2019/02/video-social-networking-app-musically-agrees-settle-ftc. Accessed 21 Aug 2020
3. Nomura Research Institute: Digital Identity: Autonomous and Distributed Identity (2020). https://www.nri.com/-/media/Corporate/jp/Files/PDF/service/ips/technology_1.pdf?la=ja-JP&hash=255BF197AD405C48800CED1B7FFFDD98A93A5CDE. Accessed 21 Aug 2020. NRI Secure Technologies, Ltd., JCB
4. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System (2008). https://bitcoin.org/bitcoin.pdf. Accessed 21 Aug 2020
5. Kishigami, J.: The blockchain-based digital content distribution system. Future Gener. Comput. Syst. **89**, 746–764 (2018)
6. Savelyev, A.: Copyright in the blockchain era: promises and challenges. Comput. Law Secur. Rev. **34**(3), 550–561 (2018)
7. Finck, M., Moscon, V.: Copyright law on blockchains: between new forms of rights administration and digital rights management 2.0. IIC – Int. Rev. Intellect. Property Competit. Law **50**(1), 77–108 (2018). https://doi.org/10.1007/s40319-018-00776-8
8. Galloway, S.: The Four: The Hidden DNA of Amazon, Apple, Facebook, and Google. Portfolio (2017)
9. Allen, C.: The Path to Self-Sovereign Identity, 25 April 2016. http://www.lifewithalacrity.com/2016/04/the-path-to-self-sovereign-identity.html. Accessed 21 Aug 2020
10. Medium. Self-Sovereign Identity: Shifting the Locus of Control. https://medium.com/@trbouma/self-sovereign-identity-shifting-the-locus-of-control-10da1c8757ad. Accessed 21 Aug 2020
11. Abraham, A.: Whitepaper Self-Sovereign Identity (2017). https://www.egiz.gv.at/files/download/Self-Sovereign-Identity-Whitepaper.pdf. Accessed 21 Aug 2020
12. Ogawa, A.: What Is the Self-Sovereign Identity? The New Potential of Blockchain, InfoCom T & S World Trend Report, No. 346 (2018)
13. European Union Blockchain Observatory and Form. Blockchain And Digital Identity (2019). https://www.eublockchainforum.eu/sites/default/files/report_identity_v0.9.4.pdf. Accessed 21 Aug 2020
14. Szabo, N.: The Idea of Smart Contracts. Nick Szabo's Papers and Concise Tutorials (1997). http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_idea.html. Accessed 21 Aug 2020
15. Hopf, S.: Blockchain technology impacting property rights and transaction costs regimes. In: Twenty-Fourth Americas Conference on Information Systems, New Orleans (2018)
16. ERC-725 Ethereum Identity Standard Wecsite. https://erc725alliance.org/. Accessed 21 Aug 2020
17. Fabian Vogelsteller frozeman, ERC: Proxy Account #725. https://github.com/ethereum/EIPs/issues/725. Accessed 21 Aug 2020

18. Fabian Vogelsteller frozeman, ERC: Key Manager #734. https://github.com/ethereum/EIPs/issues/734. Accessed 21 Aug 2020
19. Roelen, E.: https://github.com/ERC725Alliance/erc725/blob/master/docs/use-cases.md
20. METI of Japan, Ministry of Economy, Trade and Industry FY 2018 Industrial Technology Survey Report. https://www.meti.go.jp/press/2019/04/20190423002/20190423001-1.pdf. Accessed 21 Aug 2020
21. Nihon Keizai Shimbun, 2 December 2019. https://www.nikkei.com/article/DGXMZO52864490S9A201C1L41000/
22. NTT West Website. https://www.ntt-west.co.jp/news/1912/191202a.html. Accessed 21 Aug 2020
23. Kishigami, J.: The blockchain-based digital content distribution system. In: 2015 IEEE 15th International Conference on Big Data and Cloud Computing (2015). https://ieeexplore.ieee.org/document/7310737. Accessed 21 Aug 2020
24. Hayashi, K.: Toward a Soft Copyright System. Copyright Law and Economics, Keiso Shobo, pp. 227–248 (2004)
25. Hayashi, K.: Relationship between personal information rights and property rights, and between subject and object. Inf. Secur. Gener. Sci. **7**, 1–40 (2015)
26. Chang, Y.: Opportunities and challenges of using blockchain technology for copyright registration and contents licensing. Dokkyo Law Review No. 105, pp. 231–256 (2018)
27. Takayuki, M.: Analysis on LegalTech and attorney act of Japan. Inf. Netw. Law Rev. **18**, 1–23 (2019)

Assessing the Japanese Turn in AI and Robot Ethics: Extracting Meaningful Principles Between Exoticism and Empiricism in the Case of AIBO

Vassilis Galanos[1](✉) and Mary Reisel[2]

[1] Science, Technology and Innovation Studies Subject Group,
School of Social and Political Science, University of Edinburgh,
Old Surgeons' Hall, Edinburgh EH1 1LZ, UK
Vassilis.Galanos@ed.ac.uk
[2] Applied Anthropology and Japanese Studies, Rikkyo University, Tokyo, Japan
reisel@rikkyo.ac.jp

Abstract. The present paper critically examines a recent recurrent pattern of Western scholarship of importing sets of Japanese ethics in artificial intelligence/data/robot ethics contexts without a deeper examination of their meaning and value. The paper's outline is unfolded as such: (1) We draw on material stemming from an ethnographic participant-observer study that followed a debate between Western and Japanese people confronting the robotic AI pet AIBO. (2) We demarcate how many of the proposed Japanese values are practically relevant to the examination of human-robot interaction and how this feeds into existing questions about privacy and safety, in the context of a global overwhelming AI hype and narrative bias. (3) Finally, we discuss how a long history of Western enthusiasm and occasional misunderstandings of Japanese values comes full circle with the recent trend, and we conclude with a set of open questions that require more dedicated empirical research in order to reach more proper and practical value system in the future design of technology.

Keywords: AIBO · AI ethics · Data ethics · Ethnography · Human-robot interaction · Japanese ethics · Roboethics

1 Introduction

Academic scholarship on artificial intelligence (AI) ethics [1, 2, 7–9], data ethics [6, 8–10], and robot ethics [3–5, 18], is flourishing with research and suggestions for the coming future, and we see an accumulation of publications that review and design proposals for a variety of ethical tools in data-driven, autonomous, and robotic technologies. General disagreement about terminology[1] across different disciplines, or even

[1] There is plenty of work to be done given that the variety of technologies and applications conveniently currently placed under terms such as "data," "AI," "robot," "algorithm," that can be interpreted with an ever-increasing flexibility and unclear boundaries. The many understandings of other umbrella terms such as "ethics," "justice," "fairness," "responsibility," and the like, only make the situation more difficult and vague.

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 141–157, 2020.
https://doi.org/10.1007/978-3-030-62803-1_12

within specific ones, paired with age-old imaginaries [13] about replicating or outsmarting the human mind seem to fuel the debate and further complicate the ability to assess well new and emerging themes that are constantly added to a growing list of challenges surrounding AI, data, and robotics [23]. Policy debates [14, 19] have been similarly aiming to "catch-up" with the AI hype, currently departing from a period (2014–2017) of largely unsubstantiated concerns shaped by science fiction [11], narrative biases [38], and influential public figures [22]. However, practical everyday case studies appear to be rare and thus missing from the existing literature[2]. Ethics appears to be mostly context-based and the construction of a unified, one-size-fits-all ethical framework is impossible. Ethical and responsible regulation is not only required but crucial, in a sense reiterating the old Collingridge dilemma of control [24, 25]: Would it be possible to implement a technology that might be risky without ethically regulating it? And can there be an ethical set of values that will be globally accepted and respected? Mundane examples and observations of existing applications appear to be the only viable approach to extract meaningful considerations about ethics in each and every one of those domains. Empirical descriptions of each and setting of boundaries are the first step in starting to understand the ethical limits and in leading to safe and friendly technological future.

Given the aforementioned challenges in generating an all-encompassing set of ethical principles for AI/robot/data-related questions, we focus on three particulars in conjunction: (1) everyday human-robot interaction (HRI) in the case of a commercially available and successful AI robotic pet, as this is observed in (2) a specific geo-cultural environment (Japan), and finally putting into test existing proposals about a Japanese turn to AI/robot/data ethics. This study started as an experimental collaboration bringing together two lines of research: one of the authors collaborated in an ethnographic study of users interacting with Sony's robotic pet AIBO, while the other worked on a critical review of an observable trend in scholarship that involves Japanese ethics in discussions about AI/data/robot ethics. Our goal is to critically explore and compare our findings, and question to what degree are the supposedly Japanese ethical principles projected in non-Japanese AI/data/robot debates accurate. Our hope is that this initial study can open the door to new ethical values and ideas that can start a more concrete and feasible discussion on ethical regulations, as well as legal laws, of HRI and the development of advanced AI.

AIBO is a good candidate object of research as, technologically speaking, it brings in aspects of all three main strands of technology ethics we have mentioned so far: it is a robot that is intended to function and communicate with humans on a daily basis (HRI, roboethics), it uses AI algorithms (AI ethics), and its latest models are in a constant connection to the Internet cloud, which means it is constantly exchanging data with other robots of its kind and constantly gaining knowledge and information (data ethics). Other robotic pets in Japan[3], are also intended to be part of a global network connected to a cloud and constantly improving itself and its interaction with the

[2] Perhaps with the exception of algorithmic-based decision-making and injustice.

[3] The broader ethnographic project of which a segment is used in the present paper, also involves the experimental stage of a robotic cat that is to be discussed in future work.

humans they serve. As described by Wikipedia authors by the time writing (19-02-2020), “AIBO (stylized aibō, Artificial Intelligence Robot, homonymous with aibō (相棒), “pal” or “partner” in Japanese)”, hence its commercial success and inclusion in numerous households justifies its selection as an optimal case study for the practical ethics in HRI. In addition, AIBO has a long history in Japan and is known by many Japanese as well as foreign users. There were several generations of AIBO so far, and it keeps developing, growing and gaining more popularity in households with each new and improved generation.

Before proceeding, and past the brief introduction of AIBO, we would like to refer to the term “AI” and the approach we have implemented in this work (including its corollaries “robot,” “data-driven,” and so on). While, as noted earlier, the term is being largely hyped and flexibly interpreted, for this work, we will use definitions that apply to the kind of AI technology specific to AIBO (to the extent that the brand itself claims to have “AI” as part of its name). Hence, AIBO can count as an application of Blay Whitby’s non-anthropocentric assertion that AI is “the study of intelligent behaviour (in humans, animals, and machines) and the attempt to find ways in which such behaviour could be engineered in any type of artefact” [40]. In particular, AIBO further satisfies the softer view of intelligence in humans, animals, and certain machines, that is, “that quality that enables an entity to function appropriately and with foresight in its environment” [41] and does so by being a mixture first- and second-wave AI [37], that is, an embodied data-driven manipulator of programmable symbols and neural networks. AI is not only about replicating intelligence (which in itself is a term and idea that is still negotiated and unclearly defined) but also – and in this case mostly – about the ability to present human intelligence in non-human life forms and things. This model of AIBO has the ability to react to emotions, and even more, to adjust to the behaviour of a unique owner and obey only the owner’s voice and commands. These patterns make it seem intelligent and obtaining the ability to behave in an intelligent form in terms of studying behaviour patterns as well as human emotions. We acknowledge, and largely sympathise with, bodies of literature appreciating intelligence (and hence AI) as an emergent behaviour of systems as opposed to intelligence as property of information processing equipment such as brains or computers [21]. In this sense, while AIBO is considered by its manufacturers to be developed with AI software, and while humans who interact with AIBO are considered by commonsense to have natural in-born intelligence, we are interested in the opportunity of studying intelligence traits as the emergent behaviour of this HRI. We abstain from metaphysical/mystical discussions and all statements about agency are effect-oriented and not ontological.

The paper unfolds in the following order: past the present introduction of our theme follows an exemplar ethnographic observation stemming from one of the two authors’ larger project, further discussed in the next section that reviews the current trend in mixing AI/data/robot ethics with Japanese ethics; the conclusions summarise our findings, acknowledge limitations of the paper, and offer suggestions for future work. We summarize the analysis with themes relevant to the current changes of COVID-19 and the adaptation of robotic pets and their sanitation that is developed in order to meet the demands of the current and future new hygiene requirements. Our different fields of knowledge (science, technology and innovation studies with applied anthropology and ethnographic methodologies of analysis) and view of Japanese culture on the one hand and our distinct approaches to issues of ethics and technology on the other allow us to unite and design what we hope will be an original creative way to look at the issues and to offer a set of adaptable multicultural mélange of values and concepts relevant to existing ethical debates on practical technology ethics.

Interestingly, in 1983, the British clinical psychologist, Neil Frude, published his book The Intimate Machine in which he proposed the development of intimate and romantic relationships between people and machines through animism as a connector [36], and his ideas raised anxiety rather than excitement in the mid-1980s Western context; waiting for the post-2010s for scholars to project the orientalist/animist vision as a valid proposal in AI/data/robot ethics. By that time, robots and artificial pets were already part of Japanese life and social relations. When the West was haunted by The Terminator, the Japanese created robot toys and games, some bearing the images of Buddha and other Japanese gods, kami, since each one can be a god in Japan. How do we learn to relate and to feel a difference between humans, objects, and non-human living life forms around us, and why? When is an AI robot a source of pleasure and when a source of fear? Are the emotions towards the objects that surround humans embedded in humanity from conception, or are they culturally and politically constructed and manipulated? After nearly a decade of scholarship in the area, what can ethics, policy, and governance learn from the real Japan about the value of Japanese ethics in AI/data/robot ethics? Is Frude's vision reincarnated? The following empirical section aims at offering a point of departure to debate these questions empirically.

2 Living with AIBO: Notes from an Ethnographic Observation

These questions were in my[4] mind when I was on my way to one of Sony's AIBO centres, located at a major department store in central Tokyo, heading for a meeting with a Japanese business venture that was interested to design a new robot pet. The business was a new startup company managed by two young people who wanted to understand well the market of robotic pets. The new and small AIBO was just out in the market and provided a good opportunity to observe different reactions of people who came to play with the "doggie". Facing a fast aging population and a generation of people who live alone, the placement of the robot as a friend, family-substitute, and close emotional support becomes increasingly important and receives substantial government support. Sociological investigations of surveys about happiness and economy conducted by the government indicate clearly that the collapse of the traditional community and the loss of the value of social connectivity (tsunagari), that used to be the main support of people throughout Japan, led to many of the social problems seen today, such as retreat into virtual spaces, declining marriages, and suicides [33].

[4] Given that this section expresses the ethnographic observations by one of the two authors, we use first-person singular to retain the personal experience where appropriate.

Artificial companions, especially AI/robotic pets, are considered a good solution in Japanese society, and the robot pets became part of a rising culture focusing on "healing" (Iyashi), intended to provide psychological cure for possible loneliness among the growing population of the "living alone (hitori-gurashi)" people.

The idea of "healing" appeared as an important cultural value already in the 1970s and 1980s and was part of the New Age wave of spiritual and holistic awareness. Yumiyama described it as an important idea contributing to the value of social harmony and better community life since healing leads to peace and smooth relationships [35]. The concept "healing" was attached to everything: healing nature, healing cooking style, healing clothes, healing therapies, everything that could help one calm down and feel better. However, with the changes of social and economic life that Japan has been going through since the beginning of the millennium, the large-scale "spiritual boom" is fading away and new forms of smaller groups and personal healing are taking over focusing on the value of self-cultivation which is a central ideology in Japanese healing methods [34]. Actually, the care of the self is so important that acceptance of help or support from outside is seen negatively and reduces dramatically people's levels of happiness and confidence [33]. These values stand is strong opposition to the West where psychology and psychotherapy encouraged mental health by talking, sharing, and seeking help.

Clearly, the task of making a friendly robot pet that can become a companion and a healer is complicated and faces many technological, as well as emotional, challenges in the process of designing a reliable illusion of self-awareness. Sony surprised its customers introducing a version of AIBO with a "self-decision capability[5]" (it can "decide" to obey orders of the owner or not), awareness of its approaching "death" when the battery runs out (it can "feel" beforehand and reload itself without human help), a flexible body that moves when being touched, and a sophisticated set of facial expressions and vocal sounds. As I was standing in front of my favourite AIBO, named Shinobu, the technological success could be explained by the personal attachment I felt towards Shinobu after playing with it only twice. Two times were enough to create attachment.

As I was observing the people around, a young child of 3–4 years, walking with his mother nearby, stopped in front of Sony's booth and wanted to play with the doggie. The mother seemed glad to take a short break, pulled her smartphone from her bag and stepped aside to check while the child went to play fetch with Shinobu which seemed to hesitate if to obey orders from the child or not. I was curious to see how a robot pet makes "decisions" in such situations. Shinobu looked somewhat confused, moved one paw to the front, then back, then stared at the people around as if unable to decide whether to proceed or not. It started barking in the cutest voice and the audience was laughing. Suddenly it sat down looking at the people around.

Right then, a group of probably foreign tourists entered into the store and quickly noticed the cute AIBO doggie and the people that gathered around it. The group

[5] Acknowledging several misleading anthropomorphisations, we stress by the use of inverted commas the strictly metaphorical use of human faculties (e.g. thinking, deciding, perceiving) that are used to express briefly their symbolic representation in computational languages.

consisted of eight people who were chatting noisily in English, joking about the artificial dogs on display. But they kept their distance. They were all Americans, and when I asked why they did not get closer and play with AIBO, one of them said: "I can't stand these artificial things, there is something scary in this." My Japanese colleague was confused. "It's a toy," he said, "it's not pretending to be a real dog." No comment came in response. One man from the group approached and tried to understand how it works. The Sony people were happy to explain and all went well. He smiled, but only until he heard about the self-battery innovation. Then he seemed worried. He turned to the other people and asked to the sound of embossed laughter, "what do you think he is doing when we go to sleep?" Another American man started talking about the fact that the Japanese feel comfortable enough with artificial things but Americans do not, though he couldn't explain well what he meant by "not comfortable." They all felt the dog's independent skills were disturbing. "Who knows what they put into these things with all their cameras and AI? Normal people don't understand, they don't explain, and this looks too weird." However, for my Japanese colleagues, AIBO was only a tool of healing and relaxing.

Suddenly, one of the women in the group noticed the young child playing with the robotic doggie and immediately became worried. The child put his finger into AIBO's mouth and the robot was pretending to suck it, another one of the doggie's unique technical skills. "Why is this child playing with this thing alone?" she asked loudly, "it can bite his finger off!" I showed her his mother was nearby but she just said, "I don't understand how a mother can let such a small kid play with this machine alone." The conversation continued in a ping-pong of fear and misunderstandings. One of the men mentioned he could not avoid the unpleasant feeling he had regarding specific gestures, being "overwhelmed by feelings of worry and discomfort," especially when AIBO seemed to think or refused to fetch its bone. At some point one of the people noticed the camera on the backside, near the tail, which enables it a full 360º vision. He seemed shocked. "Why does it need this ability to see me everywhere?" he asked and added frustrated, "Japanese are really weird, what's wrong with a real dog?" The woman who was worried about the child gave a last upset look at his mother and added, "this is the last step before a killer machine, how can't they see that? This thing can develop itself, it can become something else, it's like these drones that you can't see and they see everything about you. It's dangerous!" It was clear that for the Americans it was the materialization of the much hyped AI killers that will bring the end of humanity and life of Earth, as projected through public commentators like Elon Musk and Stephen Hawking [13, 22]. The embodiment of The Terminator. But the Japanese were only confused, "these people are strange, can't they see it's just a toy?"

3 Glimpses of the Japanese Turn in AI/Data/Roboethics: A Critical Assessment

The present section will examine the conflicting views between the given empirical incident and AI/data/robot ethics literature referring to Japanese ethics, demonstrating the occasional wrong translations and misconceptions of Japanese concepts which sadly occur quite often even nowadays. In the minds of many non-Japanese, Japan has an image of culture that is leading in advanced technology, with highly developed and sophisticated AI and robots in many different fields. From a cultural history of Tezuka Osamu's Astro Boy [15] to the great hype in the 1980s that surrounded the Fifth Generation Computers programme [26, 32] and the more recent Society 5.0 strategy [27], the interplay between artificial forms or replications of life and sustainable living have been at the forefront of Japan's industrial, governmental, and technological research interests. The global rediscovery of the uncanny valley hypothesis [28, 30], that escaped its initial domain on prosthetics applications of robotics and extended into robotic assistants, virtual reality, or aesthetics at large, is a good example of the way Japanese thinking becomes exoticised in the understandings of non-Japanese scholars and consumers of theory. Anthropologist Jennifer Robertson's recent book Robo Sapiens Japanicus: Robots, Gender, Family, and the Japanese Nation, was a detailed attempt at bringing to a non-Japanese speaking audience the different layers of robots in Japanese society, from policy to the household [15]. Robertson has been flagging out how Japanese theorisations about robotics are paired with long Buddhist and Shintoist traditions according to which all matter is to be treated in a ritualistically respectful way. It is interesting, however, that this relationship between Japanese spiritual tradition and technical innovation has been emphasised in international debates only in the last 15 years, that is, since the third revival of AI and robotics hype.

Before exploring the current wave of Japanese AI/data/robot ethics, we will refer to two thinkers from the previous century: Lafcadio Hearn and Vilém Flusser, both of them Western thinkers whose tragic life and social context led them to find shelter in Japan (literally in the case of Hearn who moved to Japan, and metaphorically in the case of Flusser who spoke about the orientalisation of the West). In that, we are following Irmela Hijiya-Kirschnereit's assertion that "[l]ooking back into the past, we can discern typical patterns of 'differences' and 'similarities' in the perception of Japan, and while these judgements remained surprisingly constant throughout the centuries, we can also see how these images were utilized on both sides for political purposes" [16] – and across the lines of "political," we read academic, ethical, manufactural, policymaking, and so on.

Lafcadio Hearn, also known by his Japanese name Koizumi Yakumo (1850–1904), after a tumultuous and life, tragic in many respects, found himself in Japan where he become one of the most well-known Japanese folklorists in the last 15 years of his life. Hearn opposed the dominance of Western industrialisation over Japanese traditions, partly being an enthusiast about this culture, and partly seeking something different, and hence safe, against the world that treated him as different. In the words of Allen Tuttle who has written the only existing scholarly article on Hearn's ethics in 1949: "Having fled the strong wills and broad shoulders of the West to a land of enchanted miniatures, where the lotus was actually eaten, he could not forget that he was still a spiritual alien, seeking absolutes in a world of relativity" [17]. Hearn's descriptions of Japan speak about "its extraordinary goodness, its miraculous patience, its neverfailing courtesy, its simplicity of heart, its intuitive charity" (from his classic book Glimpses of Unfamiliar Japan, 1894 [17]). In the same book, he further states that Japanese people of his time lose their kind manners only in ports, where in contact with European tradespersons, thus expressing his aversion to Western culture.

Vilém Flusser (1920–1991), who living a very similarly tumultuous and migratory life, developed his existentialist philosophy of media and communication built on the idea of nomadism, while a typical theme in his writings was the interplay between oriental and occidental traditions. For Flusser, who occasionally defended orientalist views in his writings, an initial harmony existed between the Western appreciation of the gigantic, with capitalist large cities, buildings, and corporations being its greatest effectuations, and the Oriental appreciation of the minuscule, the zen minimalism, and the selflessness. However, in late capitalism, the Western exploration of the East results in a perverse capitalisation of the minuscule. Although written in 1983, this passage appears to be very relevant in contemporary debates about data ethics: "The tiny is even less human than the gigantic. The gigantic may be at least 'admired,' but the tiny disappears from view, it is 'worthless.' The 'small man' and 'self-management' are even less human than the 'big men' and the multinationals. Never before has man ceased to be the 'measure of all things' so radically as with miniaturization. In miniaturization, man becomes a particle, 'information data,' 'bit,' or worthless entity" [31]. Hijiya-Kirschnereit, in her article "A Farewell to Exoticism—Japan and the Western World," is the first to make an argument that Flusser was already dismayed by the Western distortion of Oriental philosophy, precisely because he was in deep appreciation of the latter. She translates the following revealing passage from Flusser's 1973 autobiographical text:

> "Oriental tradition therefore appeared to us not so much as the antithesis of Western tradition, but as a structure into which Western tradition could be integrated [...] The East was superior to the West, not because it had a better perception of the same things, but because it had no perception whatsoever; not because it taught us higher values, but because it knew no values; not because it taught a true faith, but because it taught no faith at all... To admit this proves that we read the Eastern texts completely differently from the Western ones... Therefore, our very approach to the East contained the seeds of our later feeling of superiority towards the East." (Flusser in [16]).

Hearn criticised, in English language, Western habits by stressing the values he saw prevalent in Japan; and Flusser criticised, mainly in Portuguese language, the Western habit of misconceiving Japan. These two steps are crucial in our theoretical understanding of a stereotyping misconception on behalf of the Western culture, typical of a long orientalist tradition in the West. In a sense, the overall outcome of what Flusser denotes as Western superiority over the East, can be seen in the reactions of the American tourists who instantly boxed Japanese "weird" culture in the way they did (although it was their choice to travel there).

The first printed (although not widely cited at the time writing) account on the importance of Japanese ethics in robotics came from the Japanese scholar Naho Kitano, and was included in a 2006 information ethics journal issue dedicated to ethics in robotics [18]. Kitano's argument is that the Japanese concept of Rinri (elsewhere written as Rin-Ri), that is "the reasonable way (or course) to form the order and to maintain harmonized human relationships," and its promotion of "the superiority of social harmonization over the individual subjectivity" as an extension of the belief in "spiritual life in objects or natural phenomena" (Kitano in [18]). For Kitano, AI/robot-related ethical questions Westerners have been concerned with (moral agency, civil rights, robotic overlords, etc.) are not part of the Japanese discourse on the topic because "[i]n Japan, the direction of such discussions is more practical than theoretical/philosophical" and "this contributes to accelerate robot R&D, and after all, leads to legitimize the being of social robots in the human society with its consequent necessary regulations change." Thus, the discussion focuses on social harmony of all participants in society – humans, animals, nature, and objects – rather than an anthropocentric set of values. Kitano explains Rinri as "the study of the community, or the way of achieving harmony in human relationships" where "each individual has a responsibility toward the universe and the community" – hence, from a Shintoist perspective, as long as Rinri is taken as a fundamental principle in design, the development of such technologies cannot be non-beneficial or harmful for society. Most probably, this captures to a great extent the Japanese people's responses in the aforementioned discussion of the previous section.

Spyros Tzafestas' 2015 volume on Roboethics appears as one of most analytical collection of approaches to the field, although probably prepared and certainly published exactly on what can be perceived to be the hype transition from roboethics to AI ethics[6]. In opposition to Western ontological views and in agreement with Kitano's ideas, Tzafestas dedicates 21 pages of his book to explain and demonstrate the relevance of Japanese ethics to roboethics [5] by emphasizing Japanese culture's avoidance of "abstract concepts in several issues of life" and "straightforward emotional expression" in order to create an optimal framework for a harmonious cohabitation of humans and robots, "based on the exploitation of the relation among human, nature, and artifacts" [5].

Tzafestas further refers to the Japanese concepts of "Shinto (relation to past)," "Seken-Tei (everyday appearances)," and "Giri (duty)," which, in his view, extend to further values of Japanese culture that feed into design principles in robotics. Such values include: "Iyashi (healing, calmness), Kawai (cute), Hukushimu (living), Nagyaka (harmonious, gentle), Kizutuku-Kokoro (sensitive inner minds)" [5]. For Tzafestas, and intercultural roboethics, taking into account both Western and Japanese (as well as other cultures') ideals for designing robots is crucial for the development of the domain. It should be noted here that one of us, being fluent in Japanese, suggests that the translation of some of these concepts is inconsistent to their official dictionary translations; while some words such as Hukushimu do not appear in most dictionaries, something which raises the problem of translation at a higher level of metaphorical abstraction in Western orientalism. Tzafestas refers to these principles without specifically linking them to examples, and while values such as cuteness or healing (with "soothing" being more adequate than calmness) indeed are taken into account in Japanese robotic design, principles such as duty or the inner mind, although important for Japanese culture, are not specifically applied to robotics.

[6] Although, a sociology of AI/data/robot ethics is lacking; and is beyond this paper's scope.

Closer to the data ethics strand, Mireille Hildebrandt's 2016 book, Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology, dedicates a chapter to "The other side of privacy: agency and privacy in Japan." Hildebrandt stresses that "the Japanese legislation on privacy and data protection has been enacted merely to comply with the demands of Western trade partners[7]," and that the Western concept of privacy is external to Japanese tradition to the extent that the Japanese word for privacy is actually the imported English word – puraibashii [20], that means a gaikokugo, or "foreign word," a transvocalised version of the idea of privacy in other cultures outside of Japan. Hildebrandt's work contributes to a list of Japanese values that do not speak to HRI only, but also in terms of humans' relationship with their conceptualisation of privacy. According to her, these values are "the 'Inbetween' (Aida)," "politeness as 'Face' or 'Place' (Basho)," "Situated Discernment (Wakimae) and the Culture of 'As-If'," and "the Indulgence of Restraint and Acuity: Amae and Enryo-sasshi" [5]. While space limitations do not allow for a thorough analysis of these values (and criticisms of their translations as similar problems occur like in the case of Tzafestas), the following passage is capturing the essence of Hildebrandt's understanding of Western ownership of privacy as opposed to Japanese environmental affordance of privacy:

> "In Japan, to recognize and qualify particular interaction patterns as a manifestation of privacy requires sensitivity to privacy as a relational concept, rather than an attribute of individual persons. The question is not whether a person 'has' privacy, but whether the environment 'affords' a person a degree of privacy. The 'environment' refers, first, to the material surroundings of an individual person, such as walls made of stone or paper or the presence of specific monitoring technologies, and, second, to the interpersonal and institutional environment that determines disclosure of our various identifications" [5].

Interestingly, the concept of privacy in Japan has been explored by several scholars, especially after the adoption and implementation in Japan in 2003 of the Act on the Protection of Personal Information (sometimes translated as Data) [55], tracing the meaning of privacy in Japan as perceived in a very different way than in Western societies, however, influenced by the latter [54]. Most authors agree that privacy exists in Japan as expected in a collective society that developed based on what Takeo Doi identifies in the concept amae which explains this mutual dependency and the value of cooperation as a core principal in Japan [51]. This array of articles mention Doi's ideas of privacy and separation of uchi/soto, the inside-outside, as a key to understand the uniqueness of Japanese privacy [51, 53, 54]. Of particular interest, is the history of the development of the concept of privacy since the Meiji Restoration; as Adams, Murata and Orito point out, the Meiji Constitution of 1889 included specific references to the protection of personal information, something that is at apparent odds with the idea of trust in a collective society [53]. However, according to these authors, this complementarity of opposites is the key to understand the uniqueness in Japanese perception of privacy. In such a collective society, the individuals inside a group, the uchi group, depend on each other and this guarantees trust; with simultaneous respect to each

[7] Cf. Hearn's comment on Japanese kindness being "poisoned" only in ports where locals came in contact with European tradespersons.

member's private information. Finally, harmony is also the key to understand information privacy in the modern age. However, these already relatively contradictory presentations of Japan are rather theoretical interpretations. It is useful to take a look in more empirical examples.

The debate between American tourists and Japanese locals, in the way of fears expressed by the Americans, is indicative of the way data ethics of privacy can be part of AI ethics of safety, when, in the case of AIBO becoming a dangerous opponent equipped with a camera, privacy is at stake, and separates the toy function from the robot; as opposed to the Japanese view showing that the "toyness" of AIBO grants privacy and safety to the environment. But how much of this can be said to be valid in cases of unintended uses of technology, such as the scandal of Cambridge Analytica [7]? Privacy issues exist as part of a growing research agenda of the AI-robotics-online-media-data-digital ethics spectrum. The most common point of challenge in the user-technology interaction is the "privacy paradox" [46], that is, the increasing awareness of connected technology users who have privacy concerns about the potential harmful misuse of data provided to the systems oddly competitive with the increasing usage of such systems. To review the entire array of this literature would be futile for the scope of this paper[8], however, it would be useful to look at certain recent (post-2019) empirical studies offering sufficiently opposing results, examining similar subjects from slightly different perspectives. US-based participants surveyed by Lutz and Tamò-Larrieux [46] are "concerned about data protection on the manufacturer side, followed by social privacy concerns and physical concerns." A similar survey on general attitudes towards implementation of robots (not necessarily social robots) in everyday life showed that "attitudes towards robots in Europe have become more negative between 2012 and 2017" [48] (nonetheless, this period has been marked by the simultaneous AI-phobia hype as shown in [22]). Kertész and Turunen's international survey of AIBO users [47], offers a very comprehensive and broad scope of questions, including gender, age, and cultural background (but unfortunately does not control specifically for privacy of similar specific cultural traits). Their survey finds a rather negative attitude of Japanese people towards AIBO, thus, being antithetical to the stereotypical notion, as they emphasise, of Japanese people's likelihood towards robots. We did not meet such reactions in our research and we attribute it to the unique perception of privacy in Japanese culture.

3.1 Comments on Intercultural Responses to COVID-19 and Existing Challenges in International Policy and Research Practice in Pet Robots

Between the initial research and the final study of the current work, the world, faced the challenge of the Coronavirus disease 2019 (COVID-19), with numerous lessons to be learned in the near and far future. On a blog post, international and Chinese business lecturer Xiaobai Shen suggests that cultural differences in terms of privacy

[8] The literature on privacy challenges mutually shaped by technological advances is constantly renewed [52].

intrusiveness between the UK and China, enabled regions in the latter to return to regular working habits in a matter of weeks by accepting the use of digital health apps that have caused a series of concerns in the UK and other Western cities [45]. According to one of the statements she collected about the question concerning intrusion: “Yes and no. Are we not intruded on from the day we are born?…; we also intrude on others.” This represents well the Asian perspective on the collective society and the place of the individual. The different values are also well-observed when it comes to personal hygiene and the attendance of health issues in objects and people around.

The swift appearance of COVID-19 resulted in heightened alertness to health problems and the need for anti-viral external layers in material objects, especially metal-made objects that seemed to be able to enable the virus for long time. While many parts of the world had to learn new forms of cleanliness and go through hand-washing training, the Japanese didn’t need Covid19 to learn the value and importance of body hygiene. Purity and cleanliness are central concepts in Japan’s religion, Shinto, and they are celebrated regularly by daily rituals of hand cleaning with alcohol napkins, teeth brushing after every meal, and the night bath. The practice of purifying the body outside of the bathtub, the traditional *ofuro,* originates in the Shinto belief that people soak in clean water after they have purified themselves, inside and out, from their daily interaction with the mundane world [56]. Any interaction with the world outside is a process of contamination.

The obsession with cleanliness and high level hygiene is observed in every small aspect of Japanese daily practices: clean cloths, an endless variety of good smells and deodorants for every small part of the body and the house, the little alcohol napkins each one is expected to carry in the bag. A simple glass of water is always served with an alcohol napkin, and people automatically clean their hands thoroughly before touching anything, even their own glass of water. It is easy to understand why the eruption of the virus led to emphasis of practices that had already existed and need to be extended to new objects. Products that had already existed won an upgrade and media attention, such Maruzen’s anti-viral textile products [57] and Tokyu Hands which has launched a series of anti-bacterial and anti-viral products [58]. Walking in the department of new hygiene and sanitary products, the shop sellers were demonstrating each product and its uses for specific types of objects, lovely robots obviously included with their own special spray.

Linking this to the previous discussion about Rinri, ethics and the individual responsibility in Japan are also a core value in body purification and maintenance of hygiene and health as part of social responsibility towards other people and respect towards the heaven, in contrast to the western perspective which regards hygiene as part of the care of the Self.

From our paper’s perspective, we wish to further stress two dimensions relevant to near-future research, in order to address the disruptive nature of technological progress often problematised by the entrenchment of solid expectations: (a) the practical health issues and the role of policy and regulation of their production and distribution, interwoven with (b) large-scale cross-cultural differences that seem to be growing and gaining vast importance beyond the ones examined in this paper. During the early days of the COVID-19 the other side of exoticism was apparent in the West, as several cases of racist commentary against an abstract scapegoat notion of the East was covered, including cases of speciesism against nonhuman animals associated with the disease. In the case of a popular, embodied, interactive, and zoomorphic product manufactured in Japan, these challenges are already being resolved. Future discussions should take place at national as well as international levels since they have an impact on the materials used and the forms of interaction between humans and robots globally. Hence we recommend that lessons have to be learned from a long history of innovation systems and basic scientific research interacting with science and technology studies as well as cultural studies [43]. The symbiotic relationship between healthcare and innovation, often obstructed by political constraints and shaped by symbolic cultural narratives [44] is a growing empirical field that has to be benefitted by and benefit the investigation of post-COVID-19 robotic companions' technological trajectories, especially since robots enter the healthcare world and are becoming part of hospital care and support.

How helpful can all these frameworks be to any AI/robot user (no matter the geographical location) when the very foundational "stuff" of AI and robotics (units of separable data), refer to an already Western ontological traditions of formal logic [37] that are incompatible to equally Western idealisations of Japanese ethics? This full cycle of Western spiritual escapism to exotic morals in ages of technological uncertainty has to be studied in finer detail, and this paper offers only a very early stepping stone towards the investigation of the informationally rich but likewise complex arena of ethical and technological forking paths. The empirical work conducted in the area is growing and will most likely need be revised in light of responses to the pandemic, new regulations of merchandise, development restrictions, and the changing global market needs. Current proposals to implement algorithmic schemes based on decision trees to AI and robotic systems to render them "beneficial" [2, 49], although innovative, are far too simplistic to accommodate any of the aforementioned frameworks or to resolve any of the challenges in a purely mechanistic manner. "Interaction" in HRI should be the key concern in ethical debates that tend to place more emphasis either on the human or the robot/AI. As Grudin has shown [50], every period of distrust in AI (the so-called AI winters), equals to a human-computer interaction (HCI) summer. It can be possible that given the recent advances, this HCI summer will extend to HRI as well.

4 Conclusions and Future Work

Moral uncertainty related to Western industrialisation has led thinkers such as Hearn (in the early steps of industrialisation), Flusser, or Frude (in the turn to post-industrialism), to look for alternative, non-Western ontological frameworks, often associated with oriental worldviews, such as Shintoism, Rinri, and other sets of Japanese Ethics. A similar turn appears to be found in the post-2010 hype curve of AI/data/robot ethics, generating an observable (Western) socially constructed narrative bias [38] of exoticism as alternative to other similar biases such as the AI robot killer. We would like to maintain that in principle "[b]eing a collective society, harmony and social peace are the most important values in Japanese culture exceeding individual satisfaction, especially in the public sphere" [39], and we would aim at promoting such values beyond the stereotypically Japanese scope; in a sense, to be able to “perform” the valuable lessons from the Japanese traditions in the same way that actors around of various backgrounds appropriate Japanese theatrical styles such as Kabuki or Noh. In this paper we have reviewed such proposals and comparing them to an ethnographic experience of a single encounter between Western and Japanese cultures, we find that much of the discussion of Japanese ethics in these technological domains is redundant, and even misleading. Nonetheless, the empirical case study is very much limited; and the theoretical templates examined are focusing on the Japan-related proposals only. There exists a multitude of empirical case studies to be conducted and assessed with different technologies, and, as cited in the introduction, an abundance of AI/data/robot ethical frameworks to test. To this, we should add the ensemble of rearrangements in terms of health supervision, policies, research practices, and attitudes towards robots, in light of the COVID-19. Given that AI and robotics is often perceived as an imitation or augmentation of intelligence, more fundamental comparison between, for example, European and Japanese understandings of art and mimicry and the social history of artificial life; but also, the acceptance of death. How can these be implemented to HRI? Or better: how can we use HRI to assist human existence in a world in which the disruption of harmony becomes the main rule? It has been shown in the literature, that Japanese ethical principles such as Rinri and the value of healing are perfectly compatible with Western understandings of systems thinking, such as Gregory Bateson’s ecology of mind, or Félix Guattari’s ecosophy [12, 13]. A more fruitful and international dialogue between orientalist traditions (instead of their romantic projections by Westerners) and Western pluralist ontologies in light of everyday life applications of AI and robotics is lacking, and this paper aims to stress the need for a consistent cross- and inter-cultural examination of these.

Acknowledgements. Vassilis Galanos is indebted to Prof アスキュー 里枝 (Rie Askew) and his doctoral student Phan who, after request over email communication, scanned the rare 1949 article on Hearn’s Ethics by Allen Tuttle. He is further grateful to Mireille Hildebrandt for kindly sharing her book’s chapter on Japanese Ethics after a great conversation; to Giulia De Togni for sharing her experiences of Japan, recommending Roboticus Japanicus. We are grateful to the two anonymous referees whose careful comments we have taken into account. Nonetheless, we retain full responsibility for any omissions and unclarified passages. No grants have been received for the completion of this paper.

References

1. Vakkuri, V., Abrahamsson, P.: The key concepts of ethics of artificial intelligence. In: 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–6. IEEE (2018). https://doi.org/10.1109/ICE.2018.8436265
2. Allen, C., Smit, I., Wallach, W.: Artificial morality: top-down, bottom-up, and hybrid approaches. Ethics Inf. Technol. **7**(3), 149–155 (2005). https://doi.org/10.1007/s10676-006-0004-4
3. Poulsen, A., Burmeister, O.K., Kreps, D.: The ethics of inherent trust in care robots for the elderly. In: Kreps, D., Ess, C., Leenen, L., Kimppa, K. (eds.) HCC13 2018. IAICT, vol. 537, pp. 314–328. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99605-9_24

4. Lin, P., Abney, K., Bekey, G. (eds.): Robot Ethics: The Ethical and Social Implications of Robotics. MIT Press, Cambridge (2012)
5. Tzafestas, S.G.: Roboethics: A Navigating Overview. Springer, Berlin (2016). https://doi.org/10.1007/978-3-319-21714-7
6. Birhane, A., Cummins, F.: Algorithmic Injustices: towards a Relational Ethics. arXiv preprint arXiv:1912.07376 (2019, under review)
7. Kerr, A., Barry, M., Kelleher, J.: Expectations of AI and the performativity of ethics: implications for communication governance. Big Data Soc. **7**(1) (2020)
8. Greene, D., Hoffmann, A.L., Stark, L.: Better, nicer, clearer, fairer: a critical assessment of the movement for ethical artificial intelligence and machine learning. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, pp. 2122–2131 (2019). https://doi.org/10.24251/HICSS.2019.258
9. Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., Vertesi, J.: Fairness and abstraction in sociotechnical systems. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 59–68 (2019). https://doi.org/10.1145/3287560.3287598
10. Metcalf, J., Keller, E.F., Boyd, D.: Perspectives on Big Data, Ethics, and Society. Council for Big Data, Ethics, and Society, 23 May 2016. http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/
11. Asimov, I.: “Runaround: A Short Story”. Astounding Science Fiction, March, [Reprinted in “I, Robot”, Gnome Press, New York, 1950] (1942)
12. Bateson, G.: Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology. University of Chicago Press, Chicago (1972)
13. Galanos, V.: Singularitarianism and schizophrenia. AI Soc. **32**(4), 573–590 (2016). https://doi.org/10.1007/s00146-016-0679-y
14. AI HLEG: High-Level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI. European Commission, 09 April 2019. https://ec.europa.eu/digital-singlemarket/en/news/ethics-guidelines-trustworthy-ai
15. Robertson, J.: Robo Sapiens Japanicus: Robots, Gender, Family, and the Japanese Nation. University of California Press, Oakland (2018)
16. Hijiya-Kirschnereit, I.: A farewell to exoticism—Japan and the Western world. Forensic Sci. Int. **69**(3), 177–186 (1994). https://doi.org/10.1016/0379-0738(94)90382-4
17. Tuttle, A.E.: Lafcadio Hearn and the Ethics Beyond Evolution, English Poetry, 2, December 1949, pp. 17–21 (1949)
18. Capurro, R., et al.: Ethics in robotics. Int. Rev. Inf. Ethics **6**(12/2006) (2016)
19. House of Lords. Select Committee on Artificial Intelligence: AI in the UK: Ready, Willing, and Able? Report of Session 2017–19. Ordered to be printed 13 March 2018 and published 16 April 2018. The Authority of the House of Lords (2018)
20. Hildebrandt, M.: Smart technologies and the end (s) of law: novel entanglements of law and technology. Edward Elgar Publishing, Cheltenham, Northampton (2015)
21. Galanos, V.: Artificial intelligence does not exist: lessons from shared cognition and the opposition to the nature/nurture divide. In: Kreps, D., Ess, C., Leenen, L., Kimppa, K. (eds.) HCC13 2018. IAICT, vol. 537, pp. 359–373. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99605-9_27
22. Galanos, V.: Exploring expanding expertise: artificial intelligence as an existential threat and the role of prestigious commentators, 2014–2018. Technol. Anal. Strateg. Manag. **31**(4), 421–432 (2019). https://doi.org/10.1080/09537325.2018.1518521
23. Dwivedi, Y.K., et al.: Artificial Intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int. J. Inf. Manag. (2019, in press). https://doi.org/10.1016/j.ijinfomgt.2019.08.002

24. Collingridge, D.: The Social Control of Technology. St. Martin's Press, New York (1980)
25. Kudina, O., Verbeek, P.P.: Ethics from within: Google Glass, the Collingridge dilemma, and the mediated value of privacy. Sci. Technol. Human Values **44**(2), 291–314 (2019). https://doi.org/10.1177/0162243918793711
26. Furukawa, K., Fuchi, K.: Knowledge engineering and fifth generation computers. IEEE Database Eng. Bull. **6**(4), 17–19 (1983)
27. Fukuyama, M.: Society 5.0: aiming for a new human-centered society. Japan Spotlight **1**, 47–50 (2018)
28. Mori, M.: The uncanny valley [Bukimi No Tani Genshō [不気味の谷現象]], Enajii[Energy] **7**(4), 33–35 [original text in Japanese] (1970)
29. Mori, M.: The Buddha in the Robot: A Robot Engineer's thoughts on Science and Religion (tran. Charles S. Terry). Kosei Publishing, Tokyo (1974)
30. Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. IEEE Robot. Autom. Mag. **19**(2), 98–100 (2012). https://doi.org/10.1109/MRA.2012.2192811
31. Flusser, V.: Post-History. (trans. Rodrigo Maltez Novaes, 2013). Univocal, Minnesota (1983)
32. Garvey, C.: Artificial intelligence and Japan's fifth generation: the information society, neoliberalism, and alternative modernities. Pac. Hist. Rev. **88**(4), 619–658 (2019). https://doi.org/10.1525/phr.2019.88.4.619
33. Tiefenbach, T., Kohlbacher, F.: Happiness in Japan in times of upheaval: empirical evidence from the national survey on lifestyle preferences. J. Happiness Stud. **16**(2), 333–366 (2014). https://doi.org/10.1007/s10902-014-9512-9
34. Gaitanidis, I.: Spiritual therapies in Japan. Jpn. J. Religious Stud. **39**(2), 353–385 (2012)
35. Yumiyama, T.: Varities of healing in present day Japan. Jpn. J. Religious Stud. **22**(3–4), 267–282 (1995)
36. Frude, N.: The Intimate Machine: Close Encounters with the New Computers. Century Publishing, London (1983)
37. Cantwell Smith, B.: The Promise of Artificial Intelligence: Reckoning and Judgement. The MIT Press, Cambridge (2019)
38. Williams, R.: Compressed foresight and narrative bias: pitfalls in assessing high technology futures. Sci. Cult. **15**(4), 327–348 (2006). https://doi.org/10.1080/09505430601022668
39. Reisel, M.: From "Galapagos Syndrome" to globalization: Japanese businesses between tradition and virtual reality. Int. J. Bus. Anthropol. **7**(2) (2018)
40. Whitby, B.: Artificial Intelligence: A Beginner's Guide. Oneworld, Oxford (2003)
41. Nilsson, N.J.: The Quest for Artificial Intelligence: A History of Ideas and Achievements. Cambridge University Press, Cambridge (2010)
42. Gómez-Urrego, J.D.: The intersections between infrastructures and expectations: repair and breakdown in Yachay, the city of knowledge in Ecuador. Tapuya: Lat. Am. Sci. Technol. Soc. **2**(1), 495–539 (2019). https://doi.org/10.1080/25729861.2019.1649963
43. Williams, R.: Why science and innovation policy needs Science and Technology Studies? In: Canzler, W., Kuhlmann, S., Simon, D. (eds.) Handbook of Science and Public Policy, pp. 503–522 (2019). https://doi.org/10.4337/9781784715946
44. Mittra, J., Mastroeni, M., Haddow, G., Wield, D., Barlow, E.: Re-imagining healthcare and medical research systems in post-devolution Scotland. Sociol. Res. Online **24**(1), 55–72 (2019). https://doi.org/10.1177/1360780418823221
45. Shen, X.: The digital health app got China back in business within six weeks after Covid-19 lockdown - food for thought for the UK? Medium, 31 May 2020. https://medium.com/@xiaobai.shen1/the-digital-health-app-got-china-back-in-business-within-six-weeks-after-covid-19-lockdown-food-5def59447801

46. Lutz, C., Tamó-Larrieux, A.: The robot privacy paradox: understanding how privacy concerns shape intentions to use social robots. Hum.-Mach. Commun. **1**, 87–111 (2020). https://doi.org/10.30658/hmc.1.6
47. Kertész, C., Turunen, M.: Exploratory analysis of Sony AIBO users. AI Soc. **34**(3), 625–638 (2018). https://doi.org/10.1007/s00146-018-0818-8
48. Gnambs, T., Appel, M.: Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. Comput. Hum. Behav. **93**, 53–61 (2019). https://doi.org/10.1016/j.chb.2018.11.045
49. Russell, S.: Human Compatible: AI and the Problem of Control. Penguin, UK (2019)
50. Grudin, J.: AI and HCI: two fields divided by a common focus. AI Mag. **30**(4), 48–57 (2009). https://doi.org/10.1609/aimag.v30i4.2271
51. Doi, T.: The Anatomy of Dependence. Kodansha International (1971)
52. Edwards, L., Schäfer, B., Harbinja, E.: (eds.): Future Law: Emerging Technology, Regulation and Ethics. Edinburgh University Press, Edinburgh (2020)
53. Adams, A.A., Murata, K., Orito, Y.: The Japanese sense of information privacy. AI Soc. **24** (4), 327–341 (2009). https://doi.org/10.1007/s00146-009-0228-z
54. Miyashita, H.: The evolving concept of data privacy in Japanese law. Int. Data Priv. Law **1** (4), 229–238 (2011). https://doi.org/10.1093/idpl/ipr019
55. Japanese Law Translation. Act on the Protection of Personal Information, Act No. 57 of 2003. http://www.japaneselawtranslation.go.jp/law/detail/?id=2781&vm=2&re=02&fbclid=IwAR22BRnIwaYtUFiQ9oFEKJRJbRGkFCcXECKFf1a3ePOoXDWXG0h32jmwOAo
56. Boyd, J., Williams, R.: Artful means: an aesthetic view of shinto purification rituals. J. Ritual Stud. **13**(1), 37–38 (1999)
57. Maruzen Co. Ltd. Launching various antiviral textile products that have been proven as effective and safe by third-party organizations (2020). https://shinkachi-portal.smrj.go.jp/en/webmagazine/352mp/
58. Hint Magazine. Disinfectants/Anti-bacterial Products: From stand-alone to spray-on types, the 5 leading companies in the industry have gathered here today to show us all of the newest items in the market! (2019). Tokyu Hands. https://www.tokyu-hands.co.jp/en/hintmagazine/cleaning-and-laundry/jyokin.html

Intelligent Monitoring of Chronic Illness for the Ageing Rural Population: Opportunities and Cautions

Jenni Greig[1](✉), Anwaar Ul-Haq[2], Greg Dresser[3], Oliver K. Burmeister[1], and Sabih-Ur Rehman[2]

[1] School of Computing and Mathematics, Charles Sturt University, Bathurst, Australia
jgreig@csu.edu.au
[2] School of Computing and Mathematics, Charles Sturt University, Port Macquarie, Australia
[3] Research and Innovation, LiveBetter Community Services, Orange, Australia

Abstract. Globally there are an increasing number of older people who require care for a range of health concerns, the most significant of which for our health systems are those with chronic illnesses, or multiple chronic or complex conditions. Even in countries with the best rated healthcare systems, this change in demographics and health care needs poses a significant challenge. Many older people, particularly those in non-urban locations, currently experience a range of vulnerabilities which can impact on their health status. Technological solutions are required to support health systems to be economically, socially and environmentally sustainable. In this context, socially accountable care needs to empower older people to make choices which align with their values, while also taking into account professional and familial care-givers, equitable care provision in what are often large and disjointed systems, and resource constraints. Intelligent technologies offer the potential to reduce some of the burden on health care systems, while simultaneously providing person-centered care, enabling improvements to older people's wellbeing. Through the findings of a relatively simple technology-based health intervention we explore how these benefits will only be realized if such technologies are designed and implemented with exceptional social accountability in place.

Keywords: Ageing · Assistive technology · Vulnerabilities · Social accountability

1 Introduction

Although the intelligent augmentation of human care has been considered from various perspectives, this article's contribution is focused on how the lens of 'vulnerability' has implications for both the intelligent technological and human care that have yet to be adequately considered in integrated solutions. In response to the rapidly ageing population, the World Health Organization released the first World report on ageing and health in 2015. This report emphasized the need to create environments in which the

D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 158–170, 2020.
https://doi.org/10.1007/978-3-030-62803-1_13

capacities of older people are maintained. There is a growing demand on primary health care in particular. In Australia, the number of standard GP consultations per person per year has increased for the over-65 population, while there has been a decrease in these visits for all other age groups [1]. It has also been acknowledged that older people, even those who engage in healthy behaviours earlier in their life, are more likely to experience one or more chronic diseases as they age. Geographic location impacts on the accessibility of primary health care, with waiting times to see a GP increasing as communities are increasingly distant from urban centres, or where there is more socio-demographic disadvantage [2]. This challenge of providing adequate health care as the population ages is a world-wide phenomenon [3]. Moreover, there is an expectation in many parts of the world that health care professionals and health care systems will be accountable to the communities in which they operate [4] Technology solutions have been proposed in various forms, such as intelligent homes, telehealth and care robots. The main aim of introducing these technologies is to decrease the need for people to present to the GP and local hospitals through prevention and management. To be viable as a solution, all such technology must be capable of providing good care, as defined by the values of the recipients of care, and their support networks. Such care values for the elderly have been shown to include autonomy, security, respect, trust, privacy, social wellbeing, and more [5–9]. The values literature defines the foregoing as ‘towards’ values, but it also describes ‘away from’ values [10–12], and one example is that elderly people want to engage in positive ageing, and get away from the vulnerabilities which are frequently associated with ageing and the challenges of managing chronic illness in regional and rural areas. To achieve good care is ideally a mixture of human and intelligent technologies, which together reduce as far as possible the vulnerability of elderly people living in communities. Due to the important role of the family, particularly in community care, interventions are needed that support caregivers who may not be local. Individualised, personal care is possible with emerging technologies. However, being individually-based, these applications of smart technologies have limited integration in the whole life of an older person. For instance, the Australian Productivity Commission [13] identified a range of care needs for older people when accessing aged care and support. These include gateway needs, such as information, advice, referrals, assessment and care co-ordination and management; a range of health services; housing and residential care options; disability services and community services such as transport, social and wellness activities and carer support services. Moreover, studies have identified that the aged care system is complex, and difficult to navigate to obtain appropriate services [8, 14–16]. These care needs are going to be seen globally as policies increasingly favour ageing in-situ for as long as feasible. Thus, it is critical that whole-of-life, community-based options and technology options are developed to meet the care needs of the ageing population. Such technology options offer the potential for health and aged care professionals and systems to continue or increase the social accountability of service provision, assuming that the technologies themselves are designed and implemented with social accountability in mind. Designing socially accountable technology for the purpose of care is complex, requiring that design incorporate care needs and values of individuals and their carers, as well as broader societal needs and expectations [4]. In this paper, we explore some of these complexities through the experience of a telehealth pilot project. We look at this project through a vulnerabilities framework, as a means for addressing multiple ‘social accountability’ demands.

With the emergence of technologies such as the Internet of Things (IoT), Machine Learning (ML), Artificial Intelligence (AI) and integration of data analytics platforms, intelligent aged care solutions can be envisioned to not only assist in providing quality care but to also help reduce the overall cost associated with care facilities [17]. These technological advancements can open the doors for many enhancements within the aged care domain such as clinical decision support systems to improve quality of care and operational intelligence. However a key obstacle in utilising these technological solutions in the health care industry is to overcome the challenges associated with the privacy and integrity of data [18]. A secure framework is required to protect the overall design of such an intelligent health care infrastructure in order to protect against unauthorized access to sensitive personal information.

Intelligent augmentation of human care can address many existing challenges. For example, despite persuasive evidence of our need for connection, and the clear demonstration of the influence of connection on our physiology, there is today, according to Cacioppo [19] ‘a worldwide epidemic of disconnection’. Loneliness and social isolation is far more than a social misfortune. It is a significant problem of health and happiness which is distinct from, but contributes to, the likelihood of depression, functional decline, early entry to hospitalisation and care, and higher levels of dependency. Over time if it is not addressed, loneliness and social isolation can contribute to generalized morbidity and mortality [20]. According to Dury [21] ‘older people are more vulnerable to loneliness and social isolation, and are more at risk of a range of health and social issues which can be directly linked to loneliness’. Various schemes have been put in place to try to reduce the effects of loneliness and social isolation, and there is some existing evidence to support their expansion; however, Dury [21] notes that ‘more research would provide clarity regarding their effectiveness’. Some intelligent technologies are already making significant impact in this area, including telepresence, robotic care, and others [5, 15, 22]. Although this paper reports on a simple technology-based health intervention, some of the learnings are critical as we seek to design and implement more complex or unfamiliar technologies.

The prevalence and impact of social isolation and loneliness in regional and rural areas may be more dramatic than in metropolitan areas [14, 23–25]. Strategies and structures that enable the delivery of earlier intervention in the progression of social isolation, social disconnection, depression, and morbidity/mortality is likely to improve health outcomes and quality of life for aged care clients, and provide greater social and community connectedness.

This article begins with the theoretical framework, which is about the vulnerability experienced by some of the ageing population, before then describing a study involving rural and regional seniors in Australia.

2 Theoretical Framework: Vulnerability

This work is situated within a 'vulnerabilities' approach. From an ethical perspective, this approach is located in the area of social accountability. That is, there is a moral obligation on society to care for its elderly population. As such, societal structures need to empower, rather than limit older people, such that their age-related illnesses are cared for, which the person is assisted to enjoy as full a life as possible. The theoretical framework is based on [26] framework of vulnerability, in which vulnerability is defined in terms of exposure to risk, relative capacity or resources to counter risk and meet one's needs, undermining agency and/or exacerbating powerlessness. This framework proposes that there are three main sources of vulnerability: inherent, situational and pathogenic. Inherent vulnerability captures a range of factors which are attributes of all humans, as finite, fallible beings, subject to fragility [26]. Whereas the literature focus is on recovery from temporary vulnerabilities, some research shows that ageing itself is a cause of inherent vulnerabilities and therefore older people will increase in vulnerability as they age [27]. As can be seen in Fig. 1, age and health status are both inherent factors; particularly when an individual's health status includes complex and chronic illness. Situational vulnerability refers to factors that are external to the individual, embedded in the broader social context in which the person or people group is situated. In the Australian context, living in regional or remote locations compounds age- and health-related vulnerabilities because of restricted access to care compared to urban areas. The final category, pathogenic vulnerability, is related to social and interpersonal relationships. This type of vulnerability stems from dysfunctional social interactions that create or maintain power imbalances and marginalise an individual or people group. Social factors are of interest in the context of older people experiencing chronic illness, insofar as those factors make ageing and chronic illness more difficult to manage, or disempower people to make informed choices about their care.

Additionally, vulnerability may also be "assumed" or "imposed" [27]. That is, vulnerability may be imposed by the deliberate actions or neglect of others (for example, through government policy or deliberately inflicted interpersonal harm), whereas "assumed" vulnerability acknowledges that some vulnerabilities are intrinsic to situations into which people willingly enter, for instance, trust-based relationships. In both cases, vulnerabilities are avoidable [27].

Thus, under this framework, older people with chronic health concerns may experience multiple sources of vulnerability. This framework also allows for the variety of experiences of ageing and health status which exist. Not all older people are vulnerable, and those who are will experience vulnerability differently, and to different degrees. The framework, depicted in Fig. 1, shows how different factors overlap. Therefore, appropriate approaches to mitigating vulnerability and increasing agency requires tailored, value-sensitive strategies to address overlapping sources of vulnerability. These include inherent factors (particularly where social supports are lacking and/or health status is in decline), and situational factors including personal, social, economic and other circumstances. Solutions to the growing health care needs of the ageing population need to be designed and implemented with care, so as to take

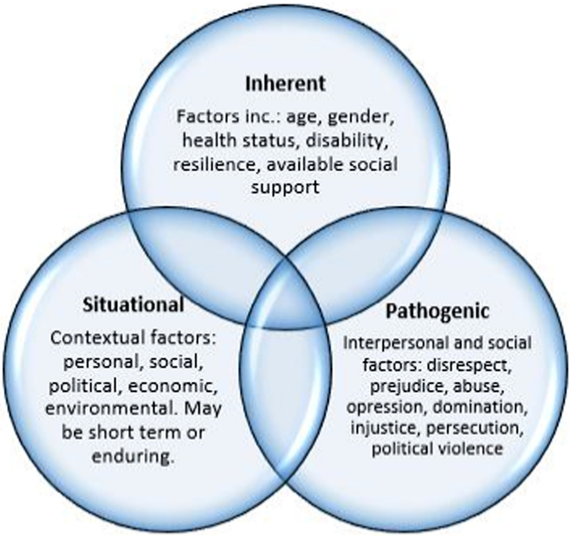

Inherent
Factors inc.: age, gender, health status, disability, resilience, available social support
Situational
Contextual factors: personal, social, political, economic, environmental. May be short term or enduring.
Pathogenic
Interpersonal and social factors: disrespect, prejudice, abuse, opression, domination, injustice, persecution, political violence

Fig. 1. Vulnerability framework, based on [26]

into account the values and needs of older people, and enhancing their capacities [3] – that is, taking care not to contribute to vulnerability. Using this framework, we suggest, is a helpful approach to designing socially accountable technology to support the delivery of socially accountable health care to older people.

3 Background to the Telehealth Project

Although the overall results of the telehealth project were previously reported [5], here we revisit specific findings to explore the implications for social accountability. In particular, this project highlighted that ‘social accountability’ has multiple dimensions when technology is being applied in the context of caring for vulnerable, older people with chronic illness, whose data is being captured and monitored. In our data driven world there is an increasing emphasis on intelligent monitoring, in which there is promise of more fulfilling lives, whilst at the same time personal liberties, such as socialising face to face, privacy and other considerations are being eroded.

It is well established that if older people maintain strong social and community connections they also maintain higher levels of wellness and functionality [30, 31]. Older people face barriers in retaining their social and community networks because:

- Relatively small changes in health and function (e.g. deterioration in their vision or hearing, deterioration in continence or balance) can lead to discontinuing social activities, and
- Physical, attitudinal and social barriers in the community challenge the capacity and motivation of older people to maintain community activities and/or taking up new opportunities [32].

The aim of the telehealth project was to (1) evaluate the use of Telehealth hardware and software in the homes of older people with chronic illnesses living in a regional community (Orange, NSW, Australia) and to (2) review the social and economic impact of the use of this equipment. In order to achieve this interviews were conducted prior to the installation of the equipment and at the completion of the trial. The first aim could only be met through an analysis of the final interview data, whereas contrasting and comparing the data for participants from pre and post interviews led to the achievement of the second aim.

Tunstall Telehealth monitoring equipment was installed for at least two months in the homes of clients who consented to participate in the study. The monitoring equipment assessed a core set of measurements (such as blood pressure, heart rate, and weight) and obtained custom measurements depending on each client's health condition, e.g., heart failure, chronic obstructive pulmonary disorder, hypertension and diabetes. The clinical/triage team from LiveBetter Services Ltd (LiveBetter) ascertained the custom measurements for each client. Clients who consented to installation of the telehealth monitoring equipment were further given the choice to participate in the research.

3.1 Research Design

The research questions were:

- What is the impact of the Tunstall telehealth systems on user perception of well-being and social functioning?
- What is the economic impact of the use of these systems?

These questions were answered in a mixed methods approach, involving pre and post interviews, observations, and through the use of a national standard instrument, the Depression Anxiety Stress Scales (DASS), though the DASS results are not reported here. Approval to conduct the study was granted by the University's Human Research Ethics Committee.

The client organisation, LiveBetter, identified and recruited 18 participants for the pre-phase and 11 for the post-phase. Pre and post data reported here only relates to 11 of the 18 participants in common to both phases. Seven pre-participants did not participate in the post-evaluation for various reasons. At the time of the study the majority lived in Orange and the surrounding region and the furthest participant lived 120 km outside Orange.

The transcribed interviews were analysed using thematic analysis with QSR NVivo, a software package for managing data.

4 Findings

The full thematic findings have been previously reported [5]. Here the focus is on social accountability and thus the economic and data driven aspects affecting vulnerable older people are focused on. Three themes emerged, 'service delivery', 'social impact', and 'technology'. Within the first was a category of 'economic impact' and within the last were categories that included ‘equipment’, ‘hardware’, and ‘interface’. Exemplary quotations related to these are presented next, and then discussed in the following section.

4.1 Economic Impact

People on retirement pensions can find even small costs, such as the cost of batteries, difficult to manage.

> The oximeter apparently takes a lot of battery power to connect with all of the other things that I use. So it always used to be the first one that I did in the morning, which I was very glad about that because having heart problems your fingers aren’t hardly warm enough in the mornings … then they rang me the other week and said “It was using too much batteries that way” so she said “Don’t worry. We have reversed the order so it’s actually the last test.” And I said “I just needed that” I thought to myself. So – and they were trying to save on the batteries so they’re obviously trying to run it economically.

Pernsioners often go without what most people consider basic necessities. This is important to note because it has significant implications on the affordability of telehealth systems.

> Oh the girls can take me shopping if I want to go shopping. So they had to take me … I don’t feel safe enough to walk up the – get out of the car and walk up the street myself … you don’t get enough money to go shopping on the pension, by the time I pay everything out, and it’s just enough – more or less just to last you a fortnight … this month’s been a really bad month for me because I had to pay $408 out for my car green slip. And then I had a lot of other expenses on top like the rent, the money from phone, the money for my insurance on the furniture … by the time I got the pension that week I had nothing left, probably had about one week – one fortnight there in this month I had about 5c to last until Thursday.

In some instances the cost of help is prohibitive, even though it would improve the quality of life.

> I don’t have much in the way of friends or social activities, or anything like that. I feel too bloody useless like I can’t do anything because I can’t walk properly, I can’t use my hands the way I used to because they shake all the time. I want to try and do some walking up and down in the pool, but that’s a bit expensive.

4.2 Equipment

There are phyiscal barriers to using some telehealth systems.

> I’m only doing … the temperature and sugar. I mean, the weight, I can’t do it because I have a wheelie and I can’t see the scales.

The ability to self-monitor gives reassurance to some people.

> I would never know me blood pressure was up without it, because you can’t go up the doctor every day. I hadn’t been up the street for about probably three years. So I find that very handy – very good.

4.3 Hardware

In addition to the battery example, above, another hardware limitation is was identified by the following participant concerned how best to self-monitor when the equipment was not reliable.

> The two things that I did find, is there are some IT issues with it. We had to reboot it, the machine, once. And secondly, a week ago, or something, I tried to use the screen to make a phone call to the people and it said this page cannot be sourced, or some such thing, and then it just froze. Had to turn the whole thing off. And then for a while it stopped talking to the scales, and all this sort of thing.

4.4 Interface

Human computer interaction, as seen in the hardward example, above, was also challenging in regards to interpreting interface messages.

> It said "Please take you ECG reading now" you take it and you press the button, and then you go onto the next one, and you finish the four. And then when that's all finished you just press the finish button, and then you just watch for the – I always watch for the signal little round piece – red piece that goes round and round to make sure it's gone to Brisbane.

> There's the blood pressure, then weight, and then ask me about salt, and how I feel today. But see it's only got better, worse, or something on the thing. Well it doesn't have in-between, so every time I press it I've got to press – same, same is the word. I just put the same because some days I feel in between it and I can't put worse down, but I could say not as well or something like that.

> Then the questions come up … asking about have you diarrhoea or vomiting … probably diarrhoea about three times … I'm not sure of the relevance.

4.5 Further Findings

In addition to the above quotations from the categories most relevant to this article, the following ones also illustrate the positive nature of the outcomes of this monitoring, and thus what the potential is for intelligent monitoring in the future to augment human care, as will be seen in the discussion section which follows.

In the initial interviews, several participants reported a relatively low level of awareness about self-management or interpretation of basic measures such as blood pressure, pulse, and blood-oxygen. When feeling unwell participants would consult a health professional. Understanding of medication, and its impact, was at the lower end of health literacy. This changed with the follow-up interviews. Several participants had been recording in notepads, or sheets of paper, their vital signs and had become comfortable in linking the recording of changes to how they felt. They were not using any recordings of their measurements but a technology, pen and paper, which they were comfortable with. Despite not starting with a clear level of comfort with the proposed technology at the initial interview, most were converts and wished to keep the equipment in the follow-up interview. Exceptions were those participants with diabetes who were used to having regular self-testing with their own instruments, and who were more comfortable in self-managing at the commencement of the study and reported little change as a consequence of the use of the tele-health equipment. Examples of self-assisted monitoring:

> It good service because daily, ten o'clock in the morning, eleven o'clock - between ten and eleven I got a chance to check my data because sometime it is different.

> When I first had the equipment I thought, … well it won't be much use to me. But in the last few months I've found it's been excellent, because I've been able to monitor a bit more and (nurse clinician) said "If you can take your blood pressure twice a day." But I've been taking it three times now, if I didn't have that equipment I wouldn't be able to do that, and they wouldn't be able to compare whether my blood pressure was dropping, or whether it was too high …

Telehealth monitoring also provided social connectivity, with 46% of the participants living alone. The nurse clinician called all the participants at least once a fortnight, including participants with stable vital signs and well-managed conditions. Participants expressed their appreciation of these conversations with the nurse clinician. These phone calls provided social connection and reassurance of remote monitoring for older participants.

Through the observations journaled during the project the following outcomes were also noted. The participants expressed to the nurse clinician that they valued their partnership with the nurse, who helped set health goals and provided advice for health-related decisions. Most importantly, this relationship provided access to a trusted nurse clinician. The participants would initiate conversations with the nurse clinician to clarify health information they had received from medical specialists and seek help to solve health-related issues. One client with heart failure stated that "the telehealth nurse has become one of the pillars in my health along with my General Practitioner (GP) and my Cardiac Nurse".

Telehealth played an important role in chronic disease management by facilitating interdisciplinary care. The nurse clinician shared information with and collaborated with GP's, Nurse Practitioners, Registered Nurses, Pharmacists and home-care workers. Data obtained from Telehealth monitoring influenced medical management decisions related to medications and identified the need for further investigations. In one case, data from the ECG Telehealth peripheral contributed to investigations which led to a more invasive procedure, significantly improving the participant's quality of life.

Finally, preventable admissions/reduced emergent medical and health consultations were also reported by participants. Several participants reported a stabilization of their vital signs, such as blood pressure, with improved compliance with medications and the ability to interpret "good" from "poor" health days. When blood pressure and/or weight was up adjustments to activity ensued, with increased activity and socialisation on good health days, and adjusted activity levels on poor health days. Participants generally reported better awareness about how they felt based on the measurements, and that meant overall they felt that they were doing more. Participants reported taking ad hoc measurements, outside of designated reporting times, just to "check" how they were measuring up against their perceptions. This suggests an improved health literacy and an ability to self-monitor and better manage their daily living. Participants were more likely to discuss trends and/or examples of good/poor days and what they felt were different about them. Medications had been reviewed and adjusted and several participants indicated that ad hoc medical and health consultations were not as frequent. They generally expressed a greater degree of comfort with their ongoing management of their condition than the health professional directed management previously. To what extent preventable admissions have been avoided could not be ascertained, and nor could frequency of contact with health professionals be verified without access to personal health records, but participants were more confident in reduced reliance on direct visits to monitor their health status than previously.

5 Discussion

In this discussion section we are applying the model discussed in Sect. 2, and findings from the telehealth project to how intelligent assistive technologies might augment human care to support the healthcare needs of the ageing population. In particular, we highlight that such technologies have the potential to address or reduce vulnerabilities or add to or compound vulnerabilities. That is, designing and implementing such technologies are a matter of social accountability.

Utilising technology to support positive ageing is not in itself a new concept. Intelligent architectures have been investigated for the potential solutions they offer for older people experiencing disabilities and chronic illnesses [7, 8, 34, 35]; and being geographically isolated from health services [15]. Moreover, numerous studies have suggested that such technologies contribute to the wellbeing and self-concept of older people [36]. These findings were consistent with what was found in the telehealth project.

5.1 Inherent Vulnerabilities

Factors identified in the framework in Sect. 2 as examples of inherent vulnerabilities include declining health status, disabilities and available social support. Assistive technology, such as that used in this telehealth study assumes a one-size-fits-all approach. The human client has to adapt themselves to the technology. As seen, this has some success. Better would be intelligent technologies that are self-adaptive, and can be tailored to each individual client. To successfully integrate technical solutions to address inherent vulnerabilities, some steps that need to be taken with careful consideration are:

- Increase familiarity and skills with technology through facilitating regular use, either in-home or at a senior-friendly tech hub, including familiarizing older people with security options such as password managers or thumb print access.
- Measure the impact and perceptions of technologies among older users, with particular focus on their experience of wellbeing and enhancing capacities.
- Design educational programs for older people and their care-givers to empower consent and choice.

5.2 Situational Vulnerabilities

Several challenges emerged from the project. The chief of these were the cost of the service and the security of data. The benefits of the telehealth service are undermined if the cost is prohibitive. That is, if people are unable to access the service due to economic factors, or will forgo other things (such as regular meals or adequate heating) in order to afford the service, it will increase vulnerabilities for some sections of the community.

Similarly, if the personal data captured in real time, across multiple networks is not secure, this will increase the vulnerability of older people using the service. This security may be an issue in the design of the software or hardware, but may also result from users being unfamiliar with securing devices in their own home. These challenges can be addressed from a social accountability perspective. Public resourcing could be used to ensure equitable access to such services, particularly if it is reducing the economic and capacity constraints on the health system. There are also reasonably simple solutions to in-home security, such as biometric access options.

5.3 Pathogenic Vulnerabilities

As the use of assistive technologies is on increase, their implications on the lives of our most vulnerable people is debatable, particularly in addressing dysfunctional social structures. Even if we set aside the issues of human rights and their legal obligations, the social accountability perspective is worth consideration. The concept of social care is related to trust, respect, dignity, privacy and security, and assistive technologies pose huge risks in each of these aspects. For instance, use of artificial intelligence is related to losing of trust. When rural elderly people interact with a machine that has human characteristics and treats that machine as if it were a care giver, do their expectations change? And, is there a level of deception involved that makes the use of such machines unethical [35]. Similarly, there is concern, for instance, that using robots for elder care could end in increased social isolation, and could involve deception and loss of dignity [36]. How far does the concept of smart home invade ones privacy [36] and how secure are our elderly? More needs to be learnt about how rural elderly people feel about their confidential information in terms of cybersecurity threats in this highly connected world of cloud, IoTs, and mobile devices. All such implications are debatable in different perspectives, viewpoints and ever-changing technological landscapes.

6 Conclusion

There are lessons to be drawn from this project for how we can design and implement intelligent technologies in a data-driven society which is both human-centric and socially accountable. Human care will always be needed, but given the increasing percentage of older people, compared to the overall population, the cost of such care requires intelligent technological augmentation. It is critical that such technologies be designed to address multi-faceted social accountabilities, to older people and the community, so that health and aged care service provision can, likewise, meet the numerous and complex needs and expectations of older people, their carers, and the communities in which they live.

References

1. Swerissen, H., Duckett, S., Moran, G.: Mapping primary care in Australia. Grattan Institute (2018)
2. Australian Bureau of Statistics (ABS) Patient Experiences in Australia: Summary of Findings, 2016-17. Cat no. 4839.0 (2017)
3. World Health Organisation (WHO). Ageing and Health. https://www.who.int/en/news-room/fact-sheets/detail/ageing-and-health. Accessed 04 Sept 2020
4. Fleet, L.J., Kirby, F., Cutler, S., Dunikowski, L., Nasmith, L., Shaughnessy, R.: Continuing professional development and social accountability: a review of the literature. J. Interprof. Care **22**(sup1), 15–29 (2008)
5. Burmeister, O.K., Ritchie, D., Devitt, A., Chia, E., Dresser, G., Roberts, R.: The impact of telehealth technology on user perception of wellbeing and social functioning, and the implications for service providers. Aust. J. Inf. Syst. **23** (2019)
6. Teipel, S., Babiloni, C., Hoey, J., Kaye, J., Kirste, T., Burmeister, O.K.: Information and communication technology solutions for outdoor navigation in dementia. Alzheimer's Dement. J. Alzheimer's Assoc. **12**(6), 695–707 (2016)
7. Burmeister, O.K.: The development of assistive dementia technology that accounts for the values of those affected by its use. Ethics Inf. Technol. **18**(3), 185–198 (2016). https://doi.org/10.1007/s10676-016-9404-2
8. Schikhof, Y., Mulder, I., Choenni, S.: Who will watch (over) me? Humane monitoring in dementia care. Int. J. Hum Comput Stud. **68**(6), 410–422 (2010)
9. Burmeister, O.K., Weckert, J., Williamson, K.: Seniors extend understanding of what constitutes universal values. J. Inf. Commun. Ethics Soc. **9**(4), 238–252 (2011)
10. Burmeister, O.K.: What seniors value about online community. J. Commun. Inform. **8**(1), 1–12 (2012)
11. Burmeister, O.K.: Websites for seniors: cognitive accessibility. Int. J. Emerg. Technol. Soc. **8**(2), 99–113 (2010)
12. Zimmerman, M.J.: Intrinsic vs. extrinsic value. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy. Stanford University, Stanford (2012)
13. Productivity Commission, Caring for Older Australians, Final Inquiry Report. Canberra (2011)
14. Bernoth, M., Burmeister, O.K., Morrison, M., Islam, M.Z., Onslow, F., Cleary, M.: The impact of a participatory care model on work satisfaction of care workers and the functionality, connectedness and mental health of community-dwelling older people. Issues Ment. Health Nurs. **37**(6), 429–435 (2016)
15. Pakrasi, S., Burmeister, O.K., Coppola, J.F., McCallum, T.J., Loeb, G.: Ethical telehealth design for users with dementia. Gerontechnology **13**(4), 383–387 (2015)
16. Shaw, R., Greig, J., Bone, Z., Morrison, M.: Mapping the funding and communication practices of aged care services in a regional Australian community. Rural Soc. **21**(1), 74–80 (2011)
17. Dermody, G., Fritz, R.: A conceptual framework for clinicians working with artificial intelligence and health-assistive smart homes. Nurs. Inq. **26**(1), e12267 (2019)
18. Jaigirdar, F.T., Rudolph, C., Bain, C.: Can i trust the data i see?: A physician's concern on medical data in IoT health architectures. In: Proceedings of the Australasian Computer Science Week Multiconference, pp. 1–10. ACM, Sydney (2019)
19. Cacioppo, J.T.: Epidemic of Loneliness. Psychology Today blog. https://www.psychologytoday.com/au/blog/connections/200905/epidemic-loneliness. Accessed 4 Sept 2020

20. Petigrew, S.: Reducing the experience of loneliness among older consumers. J. Res. Consum. **12**, 1–4 (2007)
21. Dury, R.: Social isolation and loneliness in the elderly: an exploration of some of the issues. Br. J. Commun. Nurs. **19**(3), 125–128 (2014)
22. van Wynsberghe, A.: Healthcare Robots: Ethics, Design and Implementation. Ashgate Publishing, Farnham (2015)
23. Walker, J., et al.: Insights and principles for supporting social engagement in rural older people. Ageing Soc. **33**(6), 938–963 (2013)
24. Burmeister, O.K., Bernoth, M., Dietsch, E., Cleary, M.: Enhancing connectedness through peer training for community-dwelling older people: a person centred approach. Issues Ment. Health Nurs. **37**(6), 406–411 (2016)
25. Burmeister, O.K., Islam, M.Z., Dayhew, M., Crichton, M.: Enhancing client welfare through better communication of private mental health data between rural service providers. Aust. J. Inf. Syst. **19**, 1–14 (2015)
26. Rogers, W., Mackenzie, C., Dodds, S.: Why bioethics needs a concept of vulnerability. Int. J. Feminist Approach. Bioeth. **5**(2), 11–38 (2012)
27. Martin, A., Hurst, S.: On vulnerability—analysis and applications of a many-faceted concept: Introduction. In Les ateliers de l'éthique/The Ethics Forum. Centre de recherche en éthique de l'Université de Montréal, Montréal (2017)
28. Bernoth, M., Dietsch, E., Burmeister, O.K., Schwartz, M.: Information management in aged care: cases of confidentiality and elder abuse. J. Bus. Ethics **122**(3), 453–460 (2013). https://doi.org/10.1007/s10551-013-1770-7
29. Lotz, M.: Vulnerability and resilience: a critical nexus. Theor. Med. Bioeth. **37**(1), 45–59 (2016). https://doi.org/10.1007/s11017-016-9355-y
30. Pantell, M., Rehkopf, D., Jutte, D., Syme, S.L., Balmes, J., Adler, N.: Social isolation: a predictor of mortality comparable to traditional clinical risk factors. Am. J. Public Health **103**(11), 2056–2062 (2013)
31. Steptoe, A., Shankar, A., Demakakos, P., Wardle, J.: Social isolation, loneliness, and all-cause mortality in older men and women. Proc. Natl. Acad. Sci. **110**(15), 5797–5801 (2013)
32. Hatfield, J., Hirsch, J., Lyness, J.: Functional impairment, illness burden, and depressive symptoms in older adults: does type of social relationship matter? Int. J. Geriatr. Psychiatry **28**(2), 190–198 (2013)
33. Crawford, J., Cayley, C., Lovibond, P.F., Wilson, P.H., Hartley, C.: Percentile norms and accompanying interval estimates from an Australian general adult population sample for self-report mood scales (BAI, BDI, CRSD, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). Aust. Psychol. **46**(1), 3–14 (2011)
34. Zwijsen, S.A., Niemeijer, A.R., Hertogh, C.M.P.M.: Ethics of using assistive technology in the care for community-dwelling elderly people: an overview of the literature. Aging Ment. Health **15**(4), 419–427 (2011)
35. Alzheimer Europe: Alzheimer Europe Report: The ethical issues linked to the use of assistive technology in dementia care. Alzheimer Europe, Luxembourg (2010)
36. Lê, Q., Nguyen, H.B., Barnett, T.: Smart homes for older people: positive aging in a digital world. Future Internet **4**(2), 607–617 (2012)

The Relevance of Humans and Structure: Managerial and Organizational Challenges in Smart Factories

Dennis Grenda(✉) and Anne-Marie Tuikka

School of Economics, University of Turku, Rehtorinpellonkatu 3,
20500 Turku, Finland
{degren, anne-marie. tuikka}@utu. fi

Abstract. Rapid technological change that permeates all areas of life characterizes the Industry 4.0. This forces companies to develop digital strategies and transform their businesses into the so-called "smart factory". However, many managers still understand digitalization as simply automation of production processes and therefore disregard the complexity and challenges it brings for their companies. This paper identifies managerial and organizational challenges from two dimensions: human and structural. From the human dimension, organizations must offer intelligent training concepts, as future employees will increasingly collaborate with machines and therefore need a holistic view of production facilities and comprehensive knowledge. From the structural perspective, organizations need to decentralize the classic management task and create the environment for self-organizations, so that the company is more agile and can quickly respond to changes within the digital era.

Keywords: Digitalization · Industry 4.0 · Smart factory · Organizational challenges

1 Introduction

The convergence of industrial production and information and communication technologies, called Industry 4.0, is currently one of the most frequently discussed topics among practitioners and academics [1]. It originated in 2011 as an approach to strengthening the competitiveness of the German manufacturing industry [2]. Promoters of this idea expect Industry 4.0 to deliver "fundamental improvements to the industrial processes involved in manufacturing, engineering, material usage and supply chain and life cycle management" [2]. Enabled through the communication between people, machines, and resources, Industry 4.0 is characterized by a paradigm shift from centrally controlled to decentralized production processes. A key element in Industry 4.0 is the so-called smart factory, also termed digital or intelligent factory [3]. A smart factory is a digitized factory that is equipped more efficiently and flexibly through the networking of the components involved in value creation [4]. The goal is a production environment that organizes itself by being completely digitally networked. All subareas of a factory are included, from the manufacturing systems to the logistics

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 171–180, 2020.
https://doi.org/10.1007/978-3-030-62803-1_14

systems. The basis for the smart factory are cyber-physical systems (CPS) and the Internet-of-Things (IoT). CPS and IoT establish the connection between real (physical) and virtual (cyber) elements via networks and information technology [5]. This link enables machines and products as well as entire warehouses and production facilities to communicate with each other. In the ideal smart factory, people no longer have to intervene in the actual production process, but merely monitor the processes.

Even though a smart factory represents “a future state of a fully connected manufacturing system”, it is no longer a vision, as multiple cases show [4]. The German automobile manufacturer Audi has introduced the modular assembly concept to replace the traditional rigid assembly line. It is based on small, separate workstations between which driverless transportation systems move the vehicles and the components [6]. Besides, the electronics producer Siemens has established a smart factory in Amberg. In this plant, products can communicate with production machines. Product codes tell production machines what requirements they have and which production steps must be taken next. Furthermore, products and machines determine which items on which production lines should be completed when to meet delivery deadlines [7]. The advantages of a smart factory compared to a conventional factory are individualized production processes and greater flexibility [8]. However, while new technologies automate production processes increasingly, it can be observed that managers have not yet recognized that smart factories go much further than “automation”. This can be seen from statements such as “We have been represented on the internet for many years and have achieved a high degree of automation in our business processes through consistent use of IT” [9]. As correct and important as such statements may be, they do not reflect a central aspect of smart factories, namely the fact that their technological possibilities enable completely new business models that go far beyond automation and optimization of business processes [10]. The rapid technological change of digitalization permeates all areas of life and consequently forces companies to comprehensive transformation processes. That means new management concepts may be needed – towards the agility of an organization that uses new methods and promotes self-organization within Industry 4.0.

This conceptual paper studies the managerial and organizational challenges in smart factories. The objective is to identify the challenges that companies face when they aim to transform their manufacturing to smart factories through reviewing academic articles searched from Business Source Complete (EBSCO), ScienceDirect (Elsevier), Scopus and Google Scholar with a publishing period between 2009 and 2019. Apart from journal articles, we also included some conference papers, books, consulting reports and websites related to industry 4.0” and “smart factories”. Review of the prior literature provides a focused assessment of the current state of the art as regards to our research question: *What type of managerial and organizational challenges emerge when traditional factories are transformed to smart factories?* We address the challenges related to transforming manufacturing companies to smart factories from two perspectives: human and structural. From a human perspective, we shed light on the role of people and their relevance in a smart factory. From a structural perspective, we provide insights how the organization should be designed and structured to meet the requirements of digitalization.

2 Smart Factories

A smart factory is characterized by a socio-technical interaction of all in the production participating actors and resources. At the centre are sensor-based and spatial distributed production resources (e.g. production machines, robots, storage systems, operating resources) that are networked, self-directed and self-configuring [2]. Consistent engineering of both the production and the product being manufactured allows the digital and physical world to mesh seamlessly. The basis for a smart factory are cyber-physical systems (CPS). These systems detect physical data by means of sensors and act by means of actuators on physical processes, store data, evaluate it and interact with the physical and digital world. Furthermore, they have human-machine interfaces and thus provide communication and control options. Over the Internet of Things (IoT), CPS communicate and cooperate with each other and humans in real time [11]. The CPS makes it possible for the first time to handle and manage the complexity arising from this interconnectedness [12]. Due to the networking ability of CPS, they are sometimes referred to as socio-cyber-physical systems [13]. Another important feature of CPS is that they have their own intelligence and, based on real-time data, they can make independent decisions about how the further process steps - for example in the production of a product - should proceed [12]. This intelligence is made possible by so-called embedded systems. These are embedded in the CPS and consist of hardware and software components for the realization of system-specific functional features [14]. Through this interconnectedness, the digital and the real-world merge to form the smart factories.

The fully developed smart factory will create a completely new production logic. In a smart factory, the products know their production history, their current and target state, and actively steer themselves through the production process by instructing machines to perform the required manufacturing tasks and ordering conveyors for transportation to the next production stage [15]. All sensors and actuators in the smart factory provide their data as semantically described services that can be specifically requested by the resulting products. Semantic machine-to-machine communication with active digital production memories makes the product an information carrier, an observer and an actor [16]. Characteristic of this view of the factory is that the systems of the factory are no longer centrally managed but decentralized in cloud-based systems that connect to the described CPS. This also defuses the question of highly integrated systems, since integration is replaced by communication. Furthermore, licensing costs are eliminated and replaced by pay-per-use models. Everything becomes a service, and everything is only paid for when it’s used, and the customization costs for software will go down accordingly, as more granularity based on those apps will allow more flexible user customization [17]. Thus, the customer is actively involved in the value-added chain in order to carry out the desired requirements or requested services himself. To run a smart factory, you need smart products as well as smart services. Smart products know all of their properties, which are required, for example, to manufacture a product or how they need to be assembled together with the required product components [16]. Information about the product, its production parameters or the necessary configurations of plants are in the right place at the right time and can be further processed digitally. In addition, the product history such as the continuous process steps or the actually manufactured features are stored directly on the product. In order to generate smart services, it is necessary to evaluate the digital data from the smart products. This will be achieved by allowing manufacturers to access the usage data from networked products in the future. As soon as the smart products leave the factory, they connect to the Internet and are then reachable digitally. Sophisticated data analytics enable manufacturers to filter out disparate patterns from usage data to develop new business models [17].

3 The Role of Humans in Smart Factories

Even before smart factories existed, there was already an approach to penetrate the production technically. This approach was called "Computer Integrated Manufacturing" (CIM) and ran under the metaphor "The deserted factory". CIM was technology-automation-driven and saw the function of humans only in the task of controlling and monitoring, similar to the monitoring personnel in a nuclear power plant. Due to its disregard for human values, CIM completely disappeared in the 1990s [18].

Will the situation be similar for Industry 4.0, especially for smart factories? There are opposing views in research regarding this issue. Frey and Osbourne [19] argue that human labor will become outdated in the industrial sector. Machines and robots will fulfill the tasks that were originally performed by humans, leading to massive job losses. Similarly, Balliester and Elsheikhi [20] point out that especially low and middle-skilled workers are at a high risk of losing their jobs.

However, the majority of authors believes that people will still be needed, nor will the deserted factory be aimed at in the same way as CIM, since, according to the current state of technology, it is not feasible either. According to Shrouf, Ordieres & Miragliotta [16], one reason for this is that the smart factory will not be deserted, since humans with their day-to-day intelligence are also superior to the best expert software in exceptional situations. The authors add that if fully automatically, we will not be able to cope with the flexibility requirements of volatile markets for the near future, and people are still in demand here with their skills. These statements are supported by Cantoni & Mangia [21] to the effect that the tactile abilities of humans are very difficult to realize by sensors and that the human's ability to associate is superior to artificial intelligence solutions. Furthermore, people can be trained relatively quickly for new tasks and can quickly adapt to situations that change at short notice. In contrast, a machine can only handle and react to what it was designed for, but then very quickly and with high repeat accuracy.

Even if, according to today's assessment, humans are still relevant, there will be a lot of changes for them in the context of smart factories. For instance, humans will interact and collaborate with robots in their daily work. This requires new skills and competences for the workers. They will be faced with increased complexity, abstraction and problem-solving requirements. Nevertheless, a high degree of self-directed acting, communication skills and self-organization skills will still be needed [22]. In spite of the digital penetration of the self-organizing value-added chain, humans are given a central role in this, up to the statement that in the future production will follow the tact of humans [21]. However, to what extent this can be achieved or even be effective from the customer's point of view remains an open question, even if it is assumed that the deployment of staff will be much more flexible than is often the case today.

Furthermore, the flexibility offered by CPS contribute to work organization models that better meet the needs of workers regarding work-life balance. For instance, Kagermann et al. [2]. argue that machines release workers from doing routine tasks and enable them to focus on more creative activities. This would lead to physical relief and especially allow older workers to remain productive for longer.

Remarkably many authors emphasize that technological progress fosters the unique abilities of humans and that they will also be able to assert themselves as strategic decision-makers and flexible problem solvers in the overall CPS. However, what it means for employees with "ordinary" skills has not yet been answered. It is clear that there was hardly a worse time for these employees to become redundant as digitalization progressed [22].

4 The Organizational Structure for Smart Factories

In order to implement smart factories, a new level of organization and governance is required [23]. The question is how the organization should be designed, if automation and real-time-oriented control systems take on more and more tasks along the value chain. An important element is decentralization, which means that decisions that were previously made centrally are now delegated to the respective areas. In contrast to the CIM approach (the deserted factory) of the 1980s, people are included in smart factories and are regarded as a key variable, which - despite increasing automation and digitization - makes a smart factory a socio-technical system [24]. This poses questions as regards the organization of smart factories. Respectively, it is postulated by a majority of scientists and expert committees that an adequate and employee-friendly organization is necessary, so that digitalization can also be realized [17]. However, this goes without specifying in detail the elements included in the organizational structure. This suggests that technical conceptualization is much more advanced than the basic concepts of organization required for successful implementation.

In theory, the smart factory manages the value-added process, the intelligent product recognizes the production process, its' processing status, as well as possible deviation steps, in order to be able to adjust them, and also triggers the logistics process, as it also knows its customers [16]. It is clear that organizations must build new business models to achieve customer value from smart factories. However, the question arises as to how an organization must be built and designed to be supportive of all this (technological) intelligence. And the question that follows - what is left of "non-intelligence", which also falls to the organization and must be collected organisationally The BCG model provides first clues to answer these questions [25]. This model was developed based on the following considerations: With the increasing decentralization of autonomous value chains due to digitalization, the centrally organized processes, structures and resources of planning and control must be broken up in order to make decentralized and agile decisions. As a result, centrally organized management functions can be decentralized to achieve shorter reaction time and self-organization. As stated by Kagermann et al. [2] self-organization is required when dealing with flexible working hours, locations and tasks in a smart factory. Managers must be able to delegate, cooperate, and share their existing power with employees. Odważny, Szymańska and Cyplik [26] even suggest organizations to create a digital culture. In this culture, employees not only possess the digital competences, but are also initiators of change. The ability to change must become daily business in every organization. This is critical, as an organization must remain proactive and constantly ask what digitalization means for it.

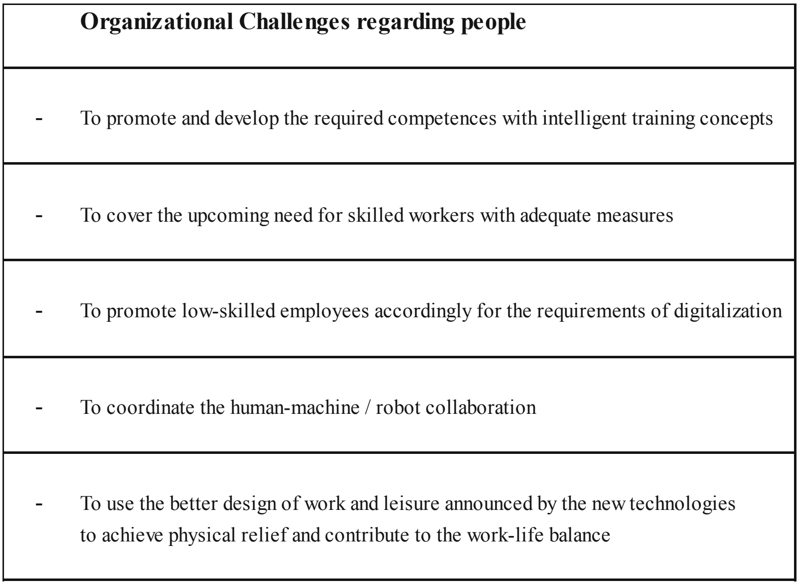
Thus, rules must be defined for the management of digital technologies, which are capable of intercepting increasing volatility throughout the system. This means that organizations need a high degree of adaptability and thus the technological, structural and organizational suitable environment for self-organization. For the transformation of technology, organization and processes of companies, the classic management tasks (leadership, planning, implementation, monitoring) must be broken down for the levels of change to digitalization. Successively, the management of a company will have to deal with these tasks in order to ultimately transform the value-added activities. This extensive change requires the adaptation of planning and control structures and processes, the selection and introduction of new (autonomous) technical systems and the monitoring and control of the economic viability of this change.

5 Discussion

Our review of prior literature on smart factories suggests that the role of humans will change to some extent as manufacturing companies are transformed to smart factories. It can be assumed that the activities in factories become more demanding due to their diverse digital networking. A continuous qualification build-up takes place on the basis of the digital competence model. This includes beside the technical topics and the overall understanding of the digital enterprise a maximum of personal responsibility and motivation. Subject-specific topics include the obvious use of the widely available online functions for quality and order management, the mastery of communication methods and tools, the derivation of concrete measures from current information and real-time decision-making in their own responsibility [27]. In order to retain these employees and their know-how and to be able to use them in other areas, the employees have to upgrade their qualifications and expand their knowledge. Training is necessary to offer employees these options. This requires well-trained people who are both technically and personally suitable for carrying out training. Such training staff have to bring with them a great deal of new knowledge in dealing with business processes, new technologies and the cooperation between humans and machines in order to provide their trainees with the knowledge that is important for the future area of activity [28].

Overall, the employee in the smart factory will determine the superior production strategy, monitor the implementation of this strategy and if necessary, intervene in the cyber-physical production system (CPPS). As part of a cyber-physical system, the employee will assume a greater degree of responsibility overall and complete his tasks with the support of various human-technology solutions. Challenges related to these changes are entangled to the role of humans in smart factories, which were presented in Sect. 3. The ones that managers and organizations have to address are summarized in Table 1.

Table 1. Managerial and organizational challenges regarding people

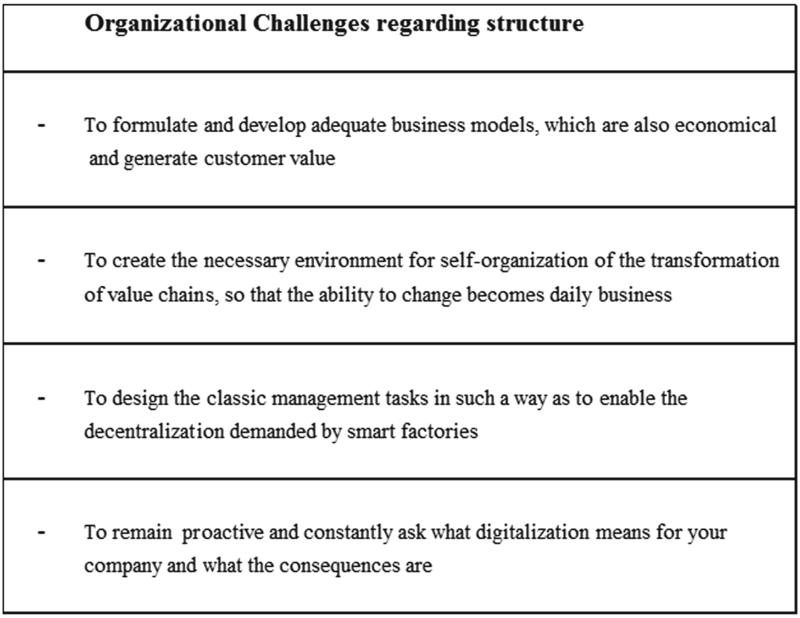


	Organizational Challenges regarding people
-	To promote and develop the required competences with intelligent training concepts
-	To cover the upcoming need for skilled workers with adequate measures
-	To promote low-skilled employees accordingly for the requirements of digitalization
-	To coordinate the human-machine / robot collaboration
-	To use the better design of work and leisure announced by the new technologies to achieve physical relief and contribute to the work-life balance

On the one hand, implementing a smart factory requires the development of suitable and appropriate methods and concepts for the necessary step “away from central management systems”. This is a task of the normative level to actively shape the framework conditions and the environment for the required degree of mutability and self-organization and control [29]. On the other hand, a socio-technical design perspective is needed in which employees and technology are coordinated with one another in order to create the necessary framework conditions and to be able to realize the expected potential [29]. Smart factories will need completely new economic and organizational structures to tap their potential [17]. Therefore, in order to achieve this, the organizational structure and successful implementation of smart factories requires agreement with the employees [21]. Managerial and organisational challenges regarding the structure in smart factories were discussed in Sect. 4 and are summarized in Table 2.

While this paper has discussed challenges related to the role of humans and to organizational structure separately, they can be seen to interact. For example, creating environment for self-organization may foster employees’ abilities to develop their competences. Analyzing such interconnections further could be an intriguing topic for future research in the area of smart factories.

Table 2. Managerial and organizational challenges regarding structure



	Organizational Challenges regarding structure
-	To formulate and develop adequate business models, which are also economical and generate customer value
-	To create the necessary environment for self-organization of the transformation of value chains, so that the ability to change becomes daily business
-	To design the classic management tasks in such a way as to enable the decentralization demanded by smart factories
-	To remain proactive and constantly ask what digitalization means for your company and what the consequences are

6 Conclusion

This paper has developed an understanding of smart factories by presenting their technologies, and their effects on the role of humans and the structure of organizations. From these two perspectives, we have identified the challenges that organizations and their management face and that they must address.

The endeavor to create a networked, intelligent world in real time, in which the separation of physical and virtual world is removed, becomes a driving force in the Industry 4.0. Value chains are broken, and humans are placed in a one-to-one relationship in the production of their goods or for the maintenance of their services. They control the smart factory in the cloud for the purchase of smart products and smart services via the Internet of Things. The role of humans in the smart factory will be to network automatic subsystems. Factory workers' required abilities will probably shift from motor skills toward associative and sensory skills. With the networking of production systems and the increasing automation, skilled workers with a holistic view of production facilities and comprehensive knowledge are needed. Organizations must therefore decentralize the classical management tasks to allow a shorter reaction time and to enable self-organization. From this point of view, it legitimizes to speak of an (upcoming) paradigm shift, and this against the background that it is not yet clear what the consequences of this technological change will be, both economically and socially. However, they will be particularly serious if the scenarios presented in this article will realize. Hence, the manager's role as a change agent in the organization aiming to transform traditional factories to smart factories is an important topic for further research.

Current technological changes and their identified potential in transforming manufacturing practices raise the question of what still includes factory work in the future, if everything that can be digitized is digitized. The impacts of smart factories to factory workers and to society pose interesting avenues for future research.

References

1. Drath, R., Horch, A.: Industrie 4.0: hit or hype? IEEE Ind. Electron. Mag. **8**(2), 56–58 (2014)
2. Kagermann, H., Helbig, J., Hellinger, A., Wahlster, W.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group. Forschungsunion (2013)
3. Stock, T., Seliger, G.: Opportunities of sustainable manufacturing in Industry 4.0. Procedia Cirp **40**, 536–541 (2016)
4. Osterrieder, P., Budde, L., Friedli, T.: The smart factory as a key construct of Industry 4.0: a systematic literature review. Int. J. Prod. Econ. **221**, 107476 (2019)
5. Fatorachian, H., Kazemi, H.: A critical investigation of Industry 4.0 in manufacturing: theoretical operationalisation framework. Prod. Plan. Control **29**(8), 633–644 (2018)
6. Ilika, D.: Here's how Audi plans to scrap the assembly line (2017). https://www.autoguide.com/auto-news/2017/07/here-s-how-audi-plans-to-scrap-the-assembly-line.html
7. Li, B.-h., Hou, B.-c., Yu, W.-t., Lu, X.-b., Yang, C.-w.: Applications of artificial intelligence in intelligent manufacturing: a review. Front. Inf. Technol. Electr. Eng. **18**(1), 86–96 (2017). https://doi.org/10.1631/FITEE.1601885
8. Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., Yin, B.: Smart factory of Industry 4.0: key technologies, application case, and challenges. IEEE Access **6**, 6505–6519 (2017)
9. Barley, S.R.: Why the internet makes buying a car less loathsome: how technologies change role relations. Acad. Manag. Discov. **1**(1), 5–35 (2015)
10. Kusiak, A.: Smart manufacturing. Int. J. Prod. Res. **56**(1–2), 508–517 (2018)
11. Hermann, M., Pentek, T., Otto, B.: Design principles for Industrie 4.0 scenarios: a literature review. Technische Universität Dortmund, Dortmund (2015)
12. Frazzon, E.M., Hartmann, J., Makuschewitz, T., Scholz-Reiter, B.: Towards socio-cyber-physical systems in production networks. Procedia Cirp **7**(2013), 49–54 (2013)
13. Wan, J., Yan, H., Liu, Q., Zhou, K., Lu, R., Li, D.: Enabling cyber–physical systems with machine–to–machine technologies. Int. J. Ad Hoc Ubiquitous Comput. **13**(3–4), 187–196 (2013)
14. Bartodziej, C.J.: Empirical study. In: Bartodziej, C.J. (ed.) The Concept Industry 4.0. BestMasters B, pp. 79–88. Springer, Wiesbaden (2017). https://doi.org/10.1007/978-3-658-16502-4_5
15. Kagermann, H.: Change through digitization—value creation in the age of Industry 4.0. In: Albach, H., Meffert, H., Pinkwart, A., Reichwald, R. (eds.) Management of Permanent Change, pp. 23–45. Springer, Wiesbaden (2015). https://doi.org/10.1007/978-3-658-05014-6_2

16. Shrouf, F., Ordieres, J., Miragliotta, G.: Smart factories in Industry 4.0: a review of the concept and of energy management approached in production based on the Internet of Things paradigm. In: 2014 IEEE International Conference on Industrial Engineering and Engineering Management, pp. 697–701. IEEE (2014)
17. Christensen, J.: Digital Economics: The Digital Transformation of Global Business. BoD–Books on Demand (2016)
18. Chryssolouris, G., Mavrikios, D., Papakostas, N., Mourtzis, D., Michalos, G., Georgoulias, K.: Digital manufacturing: history, perspectives, and outlook. Proc. Inst. Mech. Eng. Part B: J. Eng. Manuf. **223**(5), 451–462 (2009)
19. Frey, C.B., Osborne, M.A.: The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Chang. **114**, 254–280 (2017)
20. Balliester, T., Elsheikhi, A.: The future of work: a literature review. ILO Research Department Working Paper, p. 29 (2018)
21. Cantoni, F., Mangia, G. (eds.): Human Resource Management and Digitalization. Routledge, Abingdon (2018)
22. Brynjolfsson, E., McAfee, A.: The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. WW Norton & Company, New York (2014)
23. Kohnke, O.: It's not just about technology: the people side of digitization. In: Oswald, G., Kleinemeier, M. (eds.) Shaping the Digital Enterprise, pp. 69–91. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-40967-2_3
24. Urbach, N., Röglinger, M.: Introduction to digitalization cases: how organizations rethink their business for the digital age. In: Urbach, N., Röglinger, M. (eds.) Digitalization Cases. MP, pp. 1–12. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-95273-4_1
25. Novacek, G., Rashi, A., Hoo, S., Maaseide, S., Rehberg, B., Stutts, L.: Organizing for a digital future (2017). https://www.bcg.com/publications/2017/technology-organizing-for-digital-future.aspx
26. Odważny, F., Szymańska, O., Cyplik, P.: Smart Factory: the requirements for implementation of the Industry 4.0 solutions in FMCG environment-case study. LogForum **14**(2), 257–267 (2018)
27. Murawski, M., Bick, M.: Digital competences of the workforce–a research topic? Bus. Process Manag. J. **23**(3), 721–734 (2017)
28. Lorenz, M., Rüßmann, M., Strack, R., Lueth, K.L., Bolle, M.: Man and machine in Industry 4.0: how will technology transform the industrial workforce through 2025. The Boston Consulting Group, vol. 2 (2015)
29. Norta, A.: Creation of smart-contracting collaborations for decentralized autonomous organizations. In: Matulevičius, R., Dumas, M. (eds.) BIR 2015. LNBIP, vol. 229, pp. 3–17. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21915-8_1

Automated Testing of Refreshable Braille Display

Shivam Kumar Singh, Sujit Kumar Chakrabarti, and Dinesh Babu Jayagopi

IIIT-B, Bangalore, India
shivamkumar.singh@iiitb.org,
{sujitkc,jdinesh}@iiitb.ac.in

Abstract. A majority of visually impaired population of India and other developing economies live in poverty. Accessibility without affordability has little meaning to this population. Assistive technology has great potential to make education accessible to this population, e.g. through refreshable Braille display devices. However, most existing solutions in this space remain out of reach for these users due to high cost. Innovation in data science and software engineering can play an important role in making assistive technological solutions affordable and accessible. In this paper, we present a machine-learning based automated testing approach that has played an important role in enabling us to design one of the most affordable refreshable Braille display devices of the world. The key component of our approach is a visual inspection module (VIM) created using Convolutional Neural Networks (CNNs). In our experiment, our model was able to detect malfunction of a Refreshable Braille display with 97.3% accuracy. Our model is small enough to be run on a battery-powered computer in real- time. Such accurate automatic testing methods have the potential to significantly reduce the cost of RBDs.

Keywords: Automated testing · RBD · CNN · OpenCv · ORB · DCT · KNN · VIM

1 Introduction

Visual impairment is a global health issue. it is estimated that globally, there are 441 million visually impaired people encompassing range of impairment from mild levels to blindness. Over 90% of these live in developing countries like India [5]. It is estimated that India has more than 62 million people with some form of visual impairment out of which more than 8 million suffer from permanent blindness [25]. Education and integration of visually impaired people is a fundamental challenge that we need to solve. In this quest, we have taken help of many assistive technologies, one such technology is Refreshable braille display.

RBD is an electromechanical device that allows a visually impaired person to read contents of a text file using a refreshable tactile feedback reader. One of the most important components of this device is the actuator which powers a retractable pin either up or down based on electrical voltage given by the control unit. Eight of these

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 181–192, 2020.
https://doi.org/10.1007/978-3-030-62803-1_15

retractable pins combine to make a cell and multiple cells combine to make a complete display. As discussed, every cell having eight mechanical components means that they are prone to degradation and failure at the time of manufacturing and after heavy use. While manual testing is possible, it is very expensive and time-consuming, and also possible only much later in the production cycle – after the product integration. In the production stage, manual testing becomes a significant bottleneck in scaling up the production. Therefore, it is very important that the testing process be automated if we want to scale up the production process and make the product economically viable.

This paper proposes a new method to test an RBD using a digital image processing technique based on deep learning. In this method, we capture an image of the cell being tested. After processing this image, we feed this to our convolutional neural network [22] (CNN) which predicts a value. We compare this value to input we gave to the cell. If they both match with each other, we declare the cell to be error-free.

We have tested both traditional feature engineering and modern feature learning approach as explained in Sect. 4.1 and we have also experimented with multiple Neural Network architecture and hyper-parameters as explained in Sect. 4.2. After performing these experiment we have chosen feature learning approach with an architecture with excellent accuracy and acceptable inference time.

As a first step, we created a dataset by capturing images of different configurations of a cell. In all, 4096 photographs were taken for 256 different cell configurations with 16 different illumination conditions. The dataset was further augmented to over 12000 images using standard techniques. About two third of this dataset, labelled with the cell configurations they corresponded to, were used to train a CNN model to identify the label for any given image. One third of the set was used as validation set. We tested the model on 768 images. Our model identified the input correctly with over 97.3% accuracy. Each identification took close to 0.06 s, and a scheme with multiple photographs of the same RBD image input, say 50 such images will take 3 s, which is well within the acceptable limits for real-time deployment within a production line. We also benchmark our results against traditional feature engineering based approaches and show the efficacy of using a feature learning based approach (i.e. CNN in our case).

The paper is structured as follows; in Sect. 2, we relate our work with current literature. In Sect. 3, we explain our approach. In particular, we discuss the test architecture in Sect. 3.1 and the data set creation process we followed in Sect. 3.2. In Sect. 4, we discuss the experiments conducted to validate our approach and our CNN model and its architecture. In Sect. 5, we conclude the paper with a summary and a discussion on future work.

2 Related Works

Data science has come to the forefront in combating the challenges related to disability and inclusiveness through assistive technologies. A team at MIT in collaboration with NUS developed self driving wheelchair, and design was improved by researchers from College of Engineering and Computer Science, California State University at Northridge. [1, 17]. Navigation assistance is another area that has seen a significant boost in last 5 years because data from multiple sources like GPS, accelerometers, gyroscopes, and cameras is helping us solve this issue. Efforts have also gone in improving RBDs using Data Science like integrating optical character recognition (OCR) into the machine so that it can recognize and display any text [9].

Reliability of the cells is major concern in RBDs. Innovations in the direction of designing reliable RBD cells has a long history going back as far back as 1950 s [7]. There have been made many more attempts in the line of design improvement [4, 16, 26, 28, 31–33, 35].

In the quality assurance of any product a necessary compliment to design is testing. Better, faster, cheaper and effective testing is central to early discovery of faults thus preventing client site failures. This paper focuses on automated testing of RBD cells.

Automation of testing using visual inspection is done in various industries like automobile, lumber, bottling, textile etc. with great results since 1980s [10]. Visual based inspection of PCB is also prevalent [18]. In automobile sector, inspection of parts like brake cylinder, camshaft, cylinder bore etc. are being done using visual inspection [23]. Automated Visual Inspection (AVI) is also used for analysis of radiographic images like X-ray from as early as 70s and 80s [36]. Inspection for the glass and ceramic industry and the inspection for the food and packaging industry is also vision based [10]. With the advent of deep learning, testing based on visual inspections have become ubiquitous as neural networks have become particularly good at feature extraction and pattern recognition. Many industries have recently adopted this type of testing. One of the major areas that benefited from this is transportation sector. Visual inspections are used to find crack and anomaly in bridges, tunnels and railway tracks [15]. AVI has also found a fundamental place in healthcare. It is also used by doctors to complement them while verifying different medical reports. AVI makes it easier to detect diseases like Skin Cancer [13] and Parkinsons Syndrome [24]. AVI is used to automate the testing process of printed control boards. It has been found that use of this technology has proved to be highly efficient [34]. Classification of solder joints have also been done using visual inspection by the help of neural networks [20]. Visual-based automated testing is becoming prevalent now. Testing of Printed Circuits boards (PCB) shares a lot of similarity with testing of an RBD therefore, we have decided to take this approach over other approaches for our case.

In this paper, we propose a visual-based automatic inspection method for RBDs. We systematically compare, the traditional feature engineering versus modern feature learning based methods. After exploring the architecture and the hyperparameter space, we suggest an architecture with an excellent accuracy and acceptable inference time.

3 Our Approach

In this section, we describe the test architecture of our proposed solution for visual-based automatic inspection of RBDs. We describe the dataset curation process towards building the model to predict the bit pattern from the visual input.

3.1 Test Architecture

Figure 1 shows the overall architecture of our test setup. The test setup has three main components. The first component is the RBD that is to be tested. The second is the *visual inspection module* (VIM). The third is the *comparator (X).*

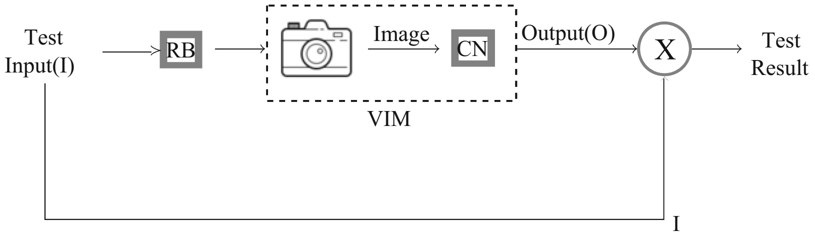


Fig. 1. Test architecture

At the time of testing, an input (i.e. a byte corresponding to the character to be displayed) is given to the cell. The Braille character displayed by the cell as a result is the actual output of the system. Due to mechanical faults, there is a probability that this output may deviate from the input. It is the purpose of the testing system to find these deviations and notify when and where they happen. The approach is explained below:

1. For each input, one photograph of RBD module (consisting of two cells as shown in Fig. 2) is taken by the digital camera mounted on the system.
2. The pre-trained CNN unit analyses each image and predicts the output that is displayed by the cell. This is the actual output O.
3. The comparator compares O with I. If they match, the cell is considered to be working as expected.

Our main objective is to design a reliable VIM that runs in real- time. To design this VIM, we created a image dataset first. We trained a CNN model using this dataset. After getting the desired accuracy, we use this model as our multiclass classifier to predict if an RBD is displaying the correct value or not. Our comparator compares this predicted value and original input (ground truth) to test if the RBD is working correctly or not. In Sects. 3.2 to 3.4 we will explain the complete process of creating the dataset and in Sect. 4 we will explain the our experiments with different CNN architectures.

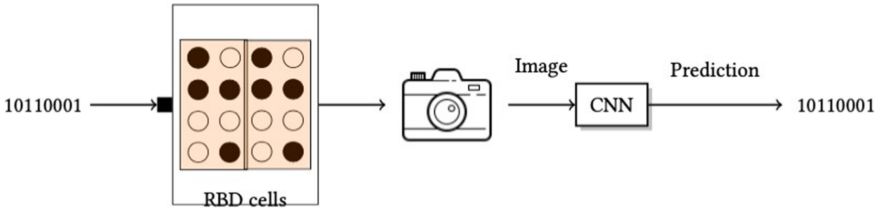


Fig. 2. Prediction by CNN

3.2 Creating the Dataset

A comprehensive dataset is needed in order to train a neural net- work. However, a dataset for RBDs does not exist, which meant that we had to create our own dataset. To create a dataset we need to capture pictures of all the combinations of a braille cell in different lighting conditions. One thing worth noting here is that in our design, two cells combined form one module i.e. each module has 16 pins. Even if one cell of the module malfunctions, we have to replace the whole module. Therefore, both cells are given the same input and tested at the same time. We came up with a comprehensive plan to efficiently capture all the images.

3.3 Capturing Images

There are 8 pins on one cell of an RBD which means that total 256 i.e. 2^8 configurations are possible. We need to capture images in various lighting conditions hence we will use 4 different colored LEDs in four top corners of the box. With 4 different LEDs we will get 16 i.e. 2^4 lighting conditions. The design of the data creation enclosure can be seen in Fig. 3. 2^4 images for every configuration means that we will get 2^{12} i.e. 4096 images. However, in practice it was found that 4096 training examples were not enough to train a CNN as the parameter space is huge, therefore, we needed to synthesize artificial data too.

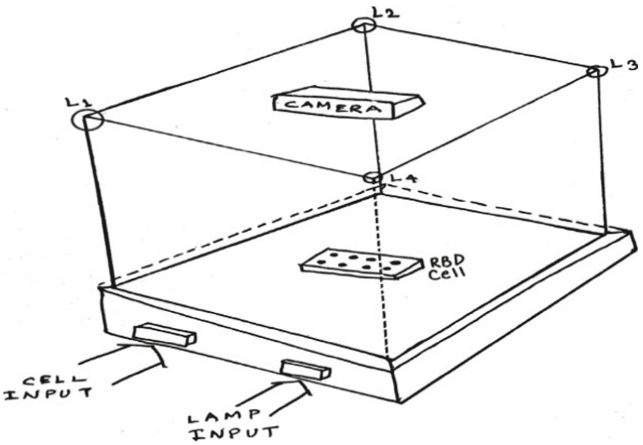


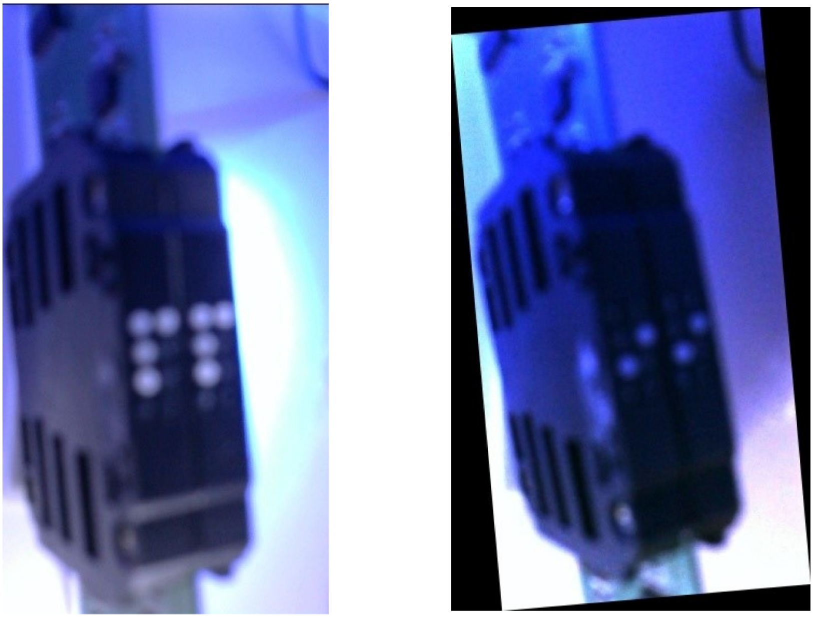
Fig. 3. Data creation enclosure

3.4 Data Augmentation

To create additional synthetic data we rotated and changed the colour patterns of all the images. Each image was rotated by 5 and −5° and slight variations were introduced in the colour pattern. These variations ensures that our dataset has diverse range of illumination conditions. Rotation ensures that model will perform well even if images are captured from different angles because it will be an invariant model [14]. After doing this we got 48 images for each configuration bringing total no. of images to 12288. In Fig. 4 and we can see a captured image and a synthesised image.

3.5 Data Pre-processing

In order to feed the data to a deep learning model, we need to make it as useful as possible as the real-world data is often incomplete, inconsistent, and lacking in certain behaviours or trends, and is likely to contain many errors [11]. However, in our case, the data was created in a controlled environment, therefore, it did not need extensive pre-processing. However, a series of steps were taken to make it more refined and useful. All the images were converted to gray-scale because when we tried the same experiment with RGB scheme, results were marginally better but significantly more computationally intensive. This also helped as it boosted the prediction time significantly. A Gaussian blur was also applied to make the images a little bit smoother [6]. Each image when captured was of size 640 × 480, it was reduced to size 100 × 100. 100 × 100 image contains enough features needed by a network and it is significantly faster to process than a 640 × 480 image.



(a) Captured image **(b)** Synthesised Image

Fig. 4. Data creation enclosure

4 Experiments and Results

Our task is a multiclass classification problem with 100 × 100 image as the input. The output classes correspond to the 256 possible pin configurations.

4.1 Feature Engineering Vs Feature Learning

We compared feature engineering approach versus feature learning based approach. Our engineered features were extracted as follows: we used Oriented FAST and Rotated BRIEF (ORB) [30] feature detector (which is an alternative to SIFT and SURF as they are patented). ORB uses FAST keypoint detector to determine the key points [29]. Then a Harris corner detector is applied to find top points. After finding top points, BRIEF descriptor is used [8]. ORB can detect both corners and blobs and it is rotation invariant and resistant to noise. Finally, we tried two classical ML multi-class classification models i.e. logistic regression (LR) and decision tree classifier (DT). As regards, feature learning, we used a CNN on the same data, the exact details of the architecture is explained in the subsequent section. Overall, results suggest feature learning is better. Our CNN model performs much better than engineered features along with shallow ML models. Details of the performance can be found in Table 1.

Table 1. Performance: Feature engineering vs Feature learning

Model	Logistic regression	Decision tree	CNN
Accuracy (%)	88	72	99.6

4.2 Experiments with CNN Architecture

We tried multiple CNN architecture combinations before deciding on the final architecture. Hyper-parameters are the parameters that affect other parameters of the model. In our model, hyperparameters are: number of layers, size of each layer, kernel size of a convolution layer, number of filters in a convolution layer, size of Maxpooling, activation functions, batch size and number of epochs. We experimented with models with one and two convolution layers. Kernel sizes were picked from $\{3, 5\}$, number of filters was picked from $\{32, 64\}$. We also experimented with ReLu [12] and sigmoid [12] activation functions. However, sigmoid suffers from vanishing gradient problems. So we used ReLu. We found out that model with two convolution layers with Kernel size as 3, number of features as 64 and activation function as ReLu performs better than other models.

We trained our network on different combinations of hyperparameters and see what works best for us. Different hyper-parameters were tested and we monitored the loss function on both Validation set and Training set on Tensorboard [2]. Table 2 shows results from one such experiment. Model 1 is the chosen model, Model 2 had only 32 filters in convolution layer, Model 3 had only one convolution layer and Model 4 has 32 filters with kernel size of (5 × 5). Model 1 performed the best.

Table 2. Comparison of experimented models

Model	Train accuracy	Validation accuracy
Model 1	99.6%	99%
Model 2	98.1%	97.2%
Model 3	96.7%	95.4%
Model 4	98.8%	98.1%

Table 3. Summary of the model

Layer type	Output shape	Kernel size
Convolution	98 × 98 × 64	3 × 3
ReLu	98 × 98 × 64	–
Max Pooling	49 × 49 × 64	3 × 3
Convolution	47 × 47 × 64	3 × 3
ReLu	47 × 47 × 64	–
Max pooling	23 × 23 × 64	3 × 3
Flattened	33856	–
Dense	64	–
ReLu	64	–
Dense	256	–
Softmax	256	–

In our final model (Model 1) the first and second layers are convolution layers with ReLu activation function and Maxpooling [12]. Third and fourth layers are fully connected, after flattening. We have used Adam [21] as our optimiser with batch size of 32. Figure 5 shows architecture of the model. After trying out many combinations, we found the best balance between speed of prediction and accuracy with these hyper-parameters (Table 3):

- No. of layers = 4
- Kernel size of a convolution layer = (3, 3)
- No. of filters in a convolution layer = 64
- Size of Maxpool = (2, 2)
- Activation functions = ReLu, ReLu, ReLu, Softmax
- Batch size = 32
- No. of epochs = 10

After training the final model on training set of 8068, cross validation set of 3258 and test set of 768 images we got the training set accuracy: 99.6%, validation set accuracy: 99% and test set accuracy: 97.3%. This shows that our algorithm has both low bias and low variance.

There are two important factors that we need to consider while prediction i.e. confidence in the prediction and time elapsed while prediction. While we have over 97% accuracy, we would still like more confidence and for that we can take multiple samples and test all of them all. We have to make sure that process is as quick as possible because this will help us test a cell at the assembly line itself. Our model takes 2.8 s to predict 50 images and compare if it matches with the test input or not. This does not take into account the time taken by the camera to capture 50 images as image capturing task is easily parallelizable.

5 Conclusion and Future Work

Affordability is at the heart of accessibility. There are many examples reported in literature where data science and computing are used in the design of accessibility features [1, 3, 17]. To the best of our knowledge, there is not enough reported work reporting how data science and computing can significantly bring down the manufacturing cost, thus adding affordability to accessibility. In this paper, we have presented a method for automated testing of RBD cells based on deep learning that has played an important role in bringing down the manufacturing cost of an RBD. In our experiments, our method performed with an accuracy of 97.3%. We believe that our model can be used on any type of Refreshable Braille Display. Also, note that in the production environment parameters like illumination and camera position/angle can be much more closely controlled. This makes it likely that the figures obtained in our experiments are more conservative in terms of accuracy and speed than those of a production environment.

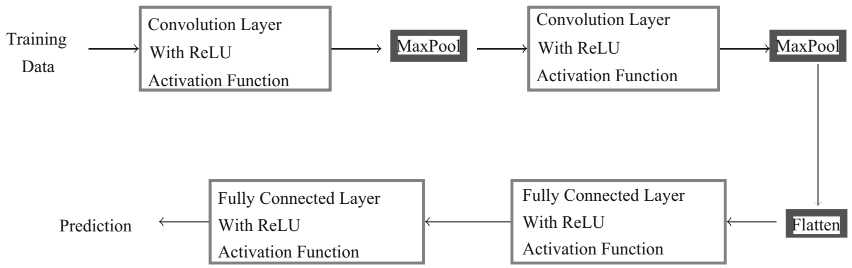

Training Data
Convolution Layer With ReLU Activation Function
MaxPool
Convolution Layer With ReLU Activation Function
MaxPool
Flatten
Fully Connected Layer With ReLU Activation Function
Fully Connected Layer With ReLU Activation Function
Prediction

Fig. 5. Model architecture

Currently, our method tests one module at a time. We can extend the method to test a complete RBD consisting of 7 such modules at one time, which will significantly enhance the testing speed. There are 2 different approaches to solve this problem: Semantic Segmentation and Sliding window Method. State of the art semantic segmentation models like Mask-RNN, Faster R-CNN [27] are too slow (5-8 FPS) for our use case as we need to test the modules on a moving conveyor belt. Sliding window method does not work well without borders around each module. Therefore, we need to slightly change the design of the module itself by putting contrasting color borders. To test the entire RBD display at a time, we also have to create a dataset to train our neural network for testing multiple cells at a time. However, this means that the number of input combinations will go to 2112 which well beyond feasible range. Test generation techniques like T-wise coverage [19] can be used to bring down the input space within feasible range.

A. Dataset. The entire dataset has been released in public domain and can be found at https://www.kaggle.com/shivam3376/refreshable-braille-display-cell.

B. Code. We are also releasing the code to train and delpoy the model. It can be found at https://github.com/shivamkumarsingh114/Automated-testing-of-RBDs.git.

Acknowledgement. This work was supported by Karnataka Innovation Technology Society, Dept. of IT, BT and ST, Govt. of Karnataka

References

1. Smart fm trials self-driving wheelchair (2017). https://smart.mit.edu/newsevents/smart-fm-trials-self-driving-wheelchair
2. Abadi, M., et al.: Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 2016), pp. 265–283 (2016)
3. Barbosa, J., Tavares, J., Cardoso, I., Alves, B., Martini, B.: Trailcare: An indoor and outdoor context-aware system to assist wheelchair users. Int. J. Hum. Comput. Stud. **116**, 1–14 (2018)
4. Blazie, D.: Refreshable braille now and in the years ahead. Braille Monit. **43**(1), 1–6 (2000)

5. Bourne, R.R., et al.: Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob. Health **5**(9), e888–e897 (2017)
6. Bradski, G.: The opencv library. Dr Dobb's J. Softw. Tools **25**, 120–125 (2000)
7. Bryce, J.W., Wheeler, J.N.: Reading apparatus (1950). US Patent 2,521,338
8. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_56
9. Chakraborty, P., Mallik, A.: An open source tesseract based tool for extracting text from images with application in braille translation for the visually impaired. International Journal of Computer Applications **68**(16) (2013)
10. Chin, R.T., Harlow, C.A.: Automated visual inspection: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 557–573 (1982)
11. Dharmarajan, R., Vijayasanthi, R.: An overview on data preprocessing methods in data mining. Int. J. Sci. Res. Dev. **3**(3), 3544–3546 (2015)
12. Ertam, F., Aydın, G.: Data classification with deep learning using tensorflow. In: 2017 International Conference on Computer Science and Engineering (UBMK), pp. 755–758. IEEE (2017)
13. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
14. Gandhi, A.: Data augmentation: how to use deep learning when you have limited data (2019). https://nanonets.com/blog/data-augmentation-how-to-usedeep-learning-when-you-have-limited-data-part-2/
15. Gibert, X., Patel, V.M., Chellappa, R.: Deep multitask learning for railway track inspection. IEEE Trans. Intell. Transp. Syst. **18**(1), 153–164 (2016)
16. Grunwald, A.P.: Reading and writing machine using raised patterns (1971). US Patent 3,624,772
17. Hartman, A., Nandikolla, V.K.: Human-machine interface for a smart wheelchair. J. Robot. **2019** (2019)
18. Jarvis, J.F.: A method for automating the visual inspection of printed wiring boards. IEEE Trans. Pattern Anal. Mach. Intell. **1**, 77–82 (1980)
19. Jorgensen, P.C.: Software Testing: A Craftsman's Approach. Auerbach Publications, Boca Raton (2013)
20. Kim, T.H., Cho, T.H., Moon, Y.S., Park, S.H.: Visual inspection system for the classification of solder joints. Pattern Recogn. **32**(4), 565–575 (1999)
21. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
22. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
23. McKeown, P., Cooke, P., Bailey, W.: The application of optics to the quality control of automotive components. In: Solving Quality Control and Reliability Problems with Optics, vol. 60, pp. 77–84. International Society for Optics and Photonics (1975)
24. Ortiz, A., Martínez-Murcia, F.J., García-Tarifa, M.J., Lozano, F., Górriz, J.M., Ramírez, J.: Automated diagnosis of parkinsonian syndromes by deep sparse filtering-based features. In: Chen, Y.-W., Tanaka, S., Howlett, R.J., Jain, L.C. (eds.) Innovation in Medicine and Healthcare 2016. SIST, vol. 60, pp. 249–258. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39687-3_24

25. Pascolini, D., Mariotti, S.P.: Global estimates of visual impairment: 2010. Br. J. Ophthalmol. **96**(5), 614–618 (2012)
26. Ren, K., Liu, S., Lin, M., Wang, Y., Zhang, Q.: A compact electroactive polymer actuator suitable for refreshable braille display. Sens. Actuators, A **143**(2), 335–342 (2008)
27. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
28. Roberts, J., Slattery, O., Kardos, D.: 49.2: rotating-wheel braille display for continuous refreshable braille. In: SID Symposium Digest of Technical Papers, vol. 31, pp. 1130–1133. Wiley Online Library (2000)
29. Rosten, E., Drummond, T.: Machine Learning for High-Speed Corner Detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_34
30. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision, pp. 2564–2571. IEEE (2011)
31. Runyan, N., Blazie, D.: EAP actuators aid the quest for the ‘holy braille’ of tactile displays. In: Electroactive Polymer Actuators and Devices (EAPAD) 2010, vol. 7642, p. 764207. International Society for Optics and Photonics (2010)
32. Runyan, N., Nassimbene, E.: Alphanumeric ‘displays’ for the blind: a technology search. Technical report, IBM Technical Report, IBM Corp., Los Gatos, CA (1974)
33. Tretiakoff, O., Tretiakoff, A.: Electromechanical transducer for relief display panel (1977). US Patent 4,044,350
34. Vitoriano, P.M., Amaral, T.G., Dias, O.P.: Automatic optical inspection for surface mounting devices with IPC-A-610D compliance. In: 2011 International Conference on Power Engineering, Energy and Electrical Drives, pp. 1–7. IEEE (2011)
35. Wallace, G.G., Teasdale, P.R., Spinks, G.M., Kane-Maguire, L.A.: Conductive Electroactive Polymers: Intelligent Materials Systems. CRC Press, Boco Raton (2002)
36. Wright, L.: Results of X-ray television inspection of electronic parts (1967)

Data Processing Automation for Bulk Water Supply Monitoring

Arno de Coning[1,2](✉) and Francois Mouton[3,4](✉)

[1] University of Pretoria, Pretoria, South Africa
arnodeconing@gmail.com
[2] North-West University, Potchefstroom, South Africa
[3] Noroff University College, Oslo, Norway
moutonf@gmail.com
[4] University of the Western Cape, Belville, South Africa

Abstract. Water as a resource is becoming more scarce with South Africa having several provinces being struck with droughts. Up to 30% of water is lost through leaks in water distribution networks. It is common practice to monitor water usage in large water distribution networks. These monitoring systems unfortunately lack the ability to alert on high flow rates and detect water leaks unless the data is reviewed manually. The paper will explore statistical and Artificial Intelligence approaches to test the viability to detect leaks. This will can then be used as an alerting team to improve operational efficiencies of small teams and reduce repair time of leaks and thus reduces water lost through leaks.

Keywords: Artificial Intelligence · Automation · Big data · Critical infrastructure · Data optimization · Leak detection · Water management

1 Introduction

Water as a resource has become more scarce with 40% of the world's population living in water stressed areas - Guppy and Anderson, 2017. Fresh water has reduced by 55% from the 1960's and the forecast is that it will increase by another 50% by 2030 [4]. Economic impact equates to US$ 500 billion per annum due to water insecurity - Guppy and Anderson, 2017. Sustainability Development Goal 6 (SDG6) has been developed due to this scarcity and projections by the United Nations (UN) to work towards water security to the world that is affordable to the masses [8].

South Africa as a region has been struck with droughts over several of its provinces [2, 10]. This in turn has forced introductions of water restrictions with the hope that water supply can be maintained to the communities. The unfortunate fact is that around 30% of water losses occur from leaks in distribution networks [4, 8, 9]. Reduction in these water losses can assist in alleviating water supply in all ready stressed water regions. An additional benefit can be realized on utilities bill savings if leaks is reduced on the client side.

This paper focuses on a specific client site, a University Campus, model development to make use of water monitoring data to detect water leaks and increase reaction time. Currently the site has a monitoring system in place, however, in its

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 193–202, 2020.
https://doi.org/10.1007/978-3-030-62803-1_16

current state it does not perform early leak detection without excessive manual work. The current monitoring system has been installed to collect usage data per hour and can be displayed on a web interface. This system has around 300 monitoring station reporting to a central server but the large amount of information requires manual intervention to view each site on the system to determine if there are irregularities in the water usage. This manual intervention of the data is required to the fact that it has no alerting mechanism installed.

It is unfortunate that current monitoring system still lacks the ability to intelligently alert on high consumption while technology and commonly available techniques can greatly improve the reaction time on water leaks. The authors have previously suggested methods of utilizing the monitoring data to trigger alerts on water leaks while minimizing false alarms. This data driven alerting approach is essential for small teams to manage large water distribution systems. Section 2 addresses an overview of the current system and the data that is available that can be used for the alerting system. Section 3 takes statistical approach to test the effectiveness to improve leak detection from the monitor- ing system. Section 4 covers an artificial intelligence (AI) implementation on the same dataset and performs a comparison of the results with the statistical approach. This section furthermore delves into a detailed discussion on the proposed improvement measures and future work that can be considered. Finally, Sect. 5 concludes this paper with a short summary and the direction of future work that is proposed.

2 System Overview

Ongoing repairs to water networks are essential requirement to ensure a continuous reduction and early detection of water leaks. Unfortunately, this is not the case in most of the implementations in industry. The focus of this paper is on the client side, the section of the water network on which the client can exert control over, of the water network. The client unfortunately only has control over their side of the water network and has to entrust the supplier to perform regular maintenance on the other side. The current system receives information from monitoring equipment installed on the water distribution network. Data is sent to a central server and trends is viewed through a web-based system with the minimum, maximum, and average flow information. Figure 1 provides an ex-ample of the system for a week's worth of data. This system has around 300 sites to be viewed manually to potentially detect leaks and this is where improvement is required as small teams cannot analyze this system constantly. The first step initially was to directly run queries on a daily basis to generate a report for nightly leak flows. This report analyses the usage of the sites between 00:00 and 05:00, during which time the site should be empty and the flows that are detected has a high likelihood of being leak flows in buildings or bulk water supply lines. Repairs can then be initiated on these sites and prioritize the actions taken by the severity of the leak. Additionally, the report includes the site name where the monitoring equipment is placed and the hourly kiloliter per hour (kl/h) usage. This is also converted into the South African Currency (ZAR) per day and the equivalent pipe size in millimeters (mm) that would cause such a leak. The reason for this conversion is due to the water site that was used as a case study for this paper. An example of the report is indicated below in Table 1 with the information available to initiate repairs on.

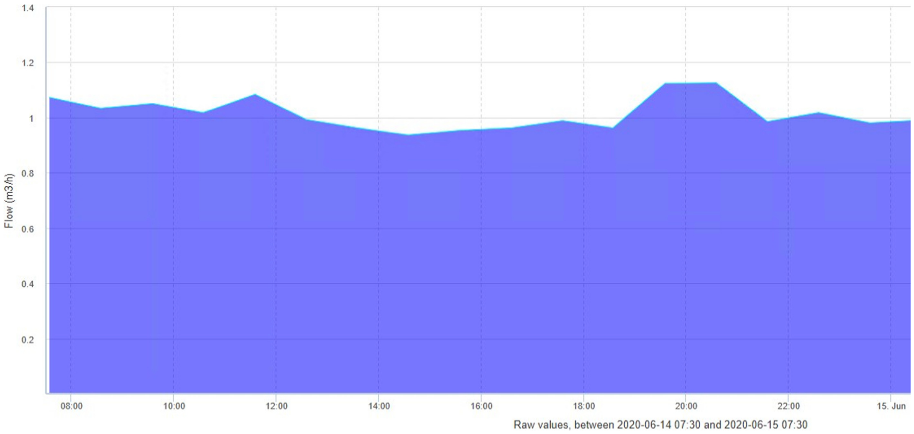


Fig. 1. Monitoring system overview

Table 1. Example of night leak flow report

Site number	Night leak flow (kl/h)	ZAR per day	Equivalent pipe size (mm)
Site 1	1.3148	993.36	13.64
Site 2	1.2695	959.13	13.4
Site 3	1.2533	946.89	13.32
Site 4	0.1199	90.59	4.12
Site 5	0.1011	76.38	3.78
Site 6	0.9863	745.17	11.81
Site 7	0.9626	727.26	11.67

The reported depicted in Table 1 does have a positive impact on the detection of leak flows and have an extremely positive impact to reduce the reaction time of addressing the detected leaks. The problem still remains though that at the current moment, it only takes a specific time period into account and leaks outside of this period is missed. Improvement is thus required to attempt to detect leaks and send alerts to decrease manual intervention required to react. An additional aspect to take into account is that water usage trends change during the day and even time of year. Trend changes are a common occurrence in several sectors and is known as seasonality. The first step is to take a statistical approach on the data in Sect. 3 to attempt to detect leaks. An Artificial intelligence (AI) approach is then implemented to detect anomalies in Sect. 4.

Development of these models require some insight into the data available from the system. A site has been selected to test the leak flow detection during the years worth of

Table 2. Data from sites example

Date & Time	Flow (kl/h)
2019/01/01 00:43	0.785
2019/01/01 01:43	0.79
2019/01/01 02:43	0.795
2019/01/01 03:43	0.835
2019/01/01 04:43	0.781
2019/01/01 05:43	2.607

data. This specific site has been selected due to a large leak that was detected with the nigh leak flow report. An example of the data can be seen in Table 2. The data for this study has been downloaded for the year of 2019 and should cover the seasonality aspects as well. A classification of the academic year can be split into six distinct sections that are used to address the seasonality of the data and the seventh to include public holidays. These current set of identified classification are:

- Class Weekday
- Class Weekend
- Exam Weekday
- Exam Weekend
- Recess Weekday
- Recess Weekend
- Public holiday

The system overview indicates that the monitoring system has useful information but requires intelligence to adapt the system to alert on leak flows. Section 3 investigates the statistical approach to determine from the data if a leak is present.

3 Statistical Model Development

The first method to test is to the average the flow rate of 2019's data. This can then be used as the threshold to test the statistical approach performance in Sect. 3.1 and is indicated as Year average in the scenario tests. Average over the dataset is 0.678 kl/h for this specific site. This average seems fairly low as time of the day is not taken into account. An average is thus calculated for each hour of the day with the result varying between 0.29 kl/h and 1.2 kl/h with the scenario indicated as Non classifier average. Results of these approaches can be seen in Fig. 2 and labeled as with the respective scenario names.

Both these approaches do not take the seasonality into account and the following statistical approach will be to determine the average flow rates for the specific time of day combined with the classification. The calendar for the academic year is used to determine the specific dates for this classification. An hourly average is then calculated for each of the seven-day classifiers and the trends can be seen in Fig. 2 combined with the Year average and Non classifier average trends. A comparison of the performance is

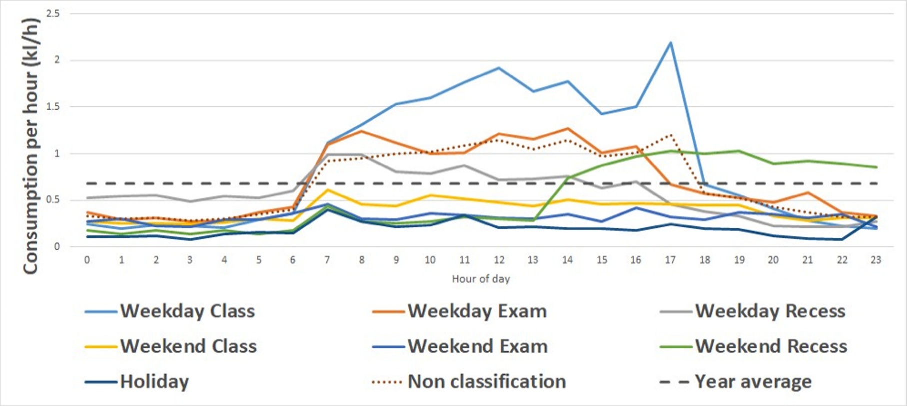


Fig. 2. Site flow statistical analysis

discussed in Sect. 3.1 after the AI implementation in Sect. 4. The accuracy of the models require testing to determine its accuracy and if false alarms will occur or leaks will not be detected.

3.1 Model Performance Testing

The performance of the different models is also tested to determine the impact that they individually have on the leak flow reporting and the subsequent impact on false alarms. During testing it is required to determine how each method will cause false alarms and report on positive leak flow results when comparing against the seasonality classifier hourly data. Each test result can visually be interpreted from Fig. 2 with the spaces between each dataset as the leaks that would have generated false alerts on or not alerted on. The scenario tests are as follows:

- Scenario 1: Year average where it triggered above threshold
- Scenario 2: Year average where it triggered above threshold and below classification hourly rate. This is then a false alarm in the test
- Scenario 3: Year average where it did not trigger threshold and above the classification hourly rate. This is then a false positive in the test
- Scenario 4: Year average where it triggered above threshold and above classification hourly rate. This will then be a positive result for leak flow.
- Scenario 5: Non classifier average where it triggered above threshold
- Scenario 6: Non classifier average where it triggered above threshold and below classification hourly rate. This is then a false alarm in the test
- Scenario 7: Non classifier average where it did not trigger threshold and above the classification hourly rate. This is then a false positive in the test.
- Scenario 8: Non classifier average where it triggered above threshold and above classification hourly rate. This will then be a positive result for leak flow.
- Scenario 9: Classifier where it triggered above threshold that should equate to positive results for leak flow

A total of 8090 data points was available for the specific site. Each of the scenarios are tested on the available 8090 data points and the results of the testing is shown in Table 3.

Table 3. Scenario performance results

Scenario number	Data points	Percentage of total data points
Scenario 1	2477	30.62%
Scenario 2	834	10.31%
Scenario 3	1169	14.45%
Scenario 4	1643	20.31%
Scenario 5	2900	35.85%
Scenario 6	726	8.97%
Scenario 7	641	7.92%
Scenario 8	2174	26.87%
Scenario 9	2815	34.80%

Scenario 9 indicated a total of 34.8% of the data as leak flows when compared to the statistical data and the models are compared to this approach as it include more classifier into the statistics averages. The year average approach would have incorrectly detected 10.31% as false positives and did not report on 14.45% of leaks above the classifier. This approach reported correctly on only 14.45% of the 34.8%. The non classifier approach had a decrease on the false positives with only 8.97% and a decrease on the amount it did not detect to 7.92%. This approach also increased the correct detection to 26.87% of the 34.8%.

The statistical approaches can improve the leak detection rate and the performance increased when taking the classifiers into account that gives a better reflection of the seasonality. Section 4 investigates the AI application on the same dataset to determine the potential improvement on the statistical models.

4 AI Implementation and Results

AI implementation has increased in the recent years. There are several use cases in industry and this section tests the accuracy when implementing AI methods on the flow data to detect leak flows. The training is based on supervised training techniques to tests the best performance. These models make use of the time series data as used in Sect. 3 with the day classification as the first classifier and the time of day as the second classifier data for the model input. The following approaches are implemented, and performance measured to test the viability of future implementation:

- Approach 1: Supervised training with only the classifiers to predict water usage for the specific hour. The alert trigger will then be if the current usage is above the predicted value.

- Approach 2: Supervised training with the statistics approach average usage per classifier as inputs and predicting the output value. The alert trigger will then be if the current usage is above the predicted value.
- Approach 3: Artificial Neural Network (ANN) classifier implementation if the values are above the statistic seasonality data it is classified as a leak and below not a leak.

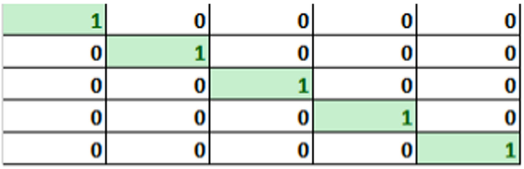
4.1 Data Preparation

Data preparation is an essential step prior to input into any AI model. All three of the approaches require the classifiers as input to ensure the seasonality is taken into account for the model performance. These columns are label encoded to take the 7 day classifiers and 24 h classifiers into a integer value to be used [3]. Output of this step results in a 2-column array with day classifiers values from 0 to 6 and the hours from 0 to 23 as indicated in Table 4.

Table 4. Label encoded result

Label of day classifier	Label of hour classifier
0	20
0	21
0	22
0	23
3	0
3	1
3	2

The values are then OneHotEncoded to split these classifier values into its own column and this output results is a 31 column array. This step assists in the model not adding a higher importance to the higher label encoder data value as each column can only be zero (0) or one (1) as the output [3]. Figure 3 indicates the hour values in each of its individual columns with the fist column value as hour 00:00 and then followed by hour 01:00 for the next index.



1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

Fig. 3. OneHotEncoded results

Approach 1 and 2 make use of the flow values as output that the model predicts. Approach 2 takes the statistics data as input with the classifier data. Approach 3 takes the same values as Approach 2 but instead of the flow data as output it has a list of zeros (0) and ones (1) where the zero (0) occurs if the flow value is below the statistic value and one (1) if it is above the value.

4.2 Supervised Training Implementation

The supervised training implementation is tested with several *sklearn* models on the dataset. Approach 1 and 2 makes use of regression models to predict the expected output and compare the current usage to determine if a leak is present. The models tested is linear regression, Gradient Boosting regression, Random Forest regression, KNeighbor Regression, Support Vector Regression (SVR) [3, 5, 7]. The regression models are tested for both approach 1 and 2 as they have different input datasets. Accuracy of the prediction is calculated by the r^2 test with 1 being the best accuracy. The results of the model testing can be seen in Table 5. The accuracy scoring is quite low on these implementations and is discussed in Sect. 4.4.

Table 5. Supervised regression results

Model	Approach 1 − r^2 score	Approach 2 − r^2 score
Linear regression	0.156	0.205
KNeighbor regression	0.021	0.021
Random Forest regression	0.178	0.167
Gradient Boosting regression	0.186	0.186
Support Vector regression	0.183	0.183

4.3 Classifier Implementation

The classifier implementation requires a output result set with specific false (no leak) or true (leak) values. A model is then trained to predict this output value where the previous regressors predicted the actual flow data for the specified day and hour classifiers. The data output change is split between the false and true output when the flow in the input dataset is above the statistical values in Sect. 3.1. Two different approaches are tested to compare the performance. The first approach is to implement a Support Vector Classification (SVC) and then an ANN implementation, as a second approach. This accuracy is then determined by generating a confusion matrix which indicates *True positives* (top left), *True negatives* (bottom right), *False positives* (bottom left), and *False negatives* (top right) [1, 5, 7]. The SVC implementation had an accuracy of 90.6% with the output confusion matrix in Eq. 1. The ANN implementation had a higher accuracy at 97.78% with the confusion matrix in Eq. 2.

$$\begin{bmatrix} 1046 & 12 \\ 140 & 420 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1041 & 17 \\ 19 & 541 \end{bmatrix} \tag{2}$$

4.4 Results Discussion and Future Work

The implementation of the classifier models had a theoretical improvement on the reaction time. The classifier led to a water leak detection accuracy of 97.78%. It is specifically stated as a theoretical improvement based on the fact that several improvements can be made to the model to give a true reflection of the leak flows. Implementation of the regressors has had very low accuracy and this can be attributed to multiple factors that also needs to done to further improve upon the classifier approach.

The statistical models with day and time classification has improvement over the fixed threshold alerting but the years worth of data could have leak con- stantly skewing the data. A process is thus required to log specific leaks and time span for a site to be used in conjunction with flow data to either exclude the data or reduce the flow by the leak amount to improve accuracy of the sta tistical approach. This will further benefit the AI approaches as the statistical models are used to predict the leak flows. An additional step can be taken to reduce the flow data by the night leak flow data to ensure better accuracy to predict the leak flow data. The regressor models have the low accuracy as it has to few input variables to predict the usage. Water usage is commonly attributed to the amount of people in a building and this can greatly assist in the predic tion process. Future work should thus be to introduce occupancy data as input to the model to improve accuracy of predictions and in turn leak detection. In addition, one should also have a look how social engineering attacks could have an impact on water monitoring systems [6].

5 Conclusion

Water as a resource has become more scarce with 40% of the world's population living in water stressed areas [4]. The unfortunate fact is that around 30% of water losses occur from leaks in distribution networks [8]. Current monitoring systems lack the ability to intelligently alert on leaks within the system without large amount of manual intervention to review the data. A data driven approach is thus proposed to analyze the data to detect leaks from the monitoring system and then to alert relevant personnel to take action to repair the leak. The automation of data analysis will assist in improved reaction times on water leaks correction by small management teams that would have required several man hours per day to detect.

The first approach that was performed was a statistical approach that deter- mines the average flow over a year dataset for a year. This was further adapted determining the average flow per hour of the day as the trends change during the day. Finally, the statistical analysis approach was adapted to take the hour of the day and the time of year classifier into account as seasonality also has severe impact on the usage. The second approach was to test AI models to firstly predict the usage for the hour to determine if a leak is present. Regressor implementation was used to predict the usage based on the flow data from the monitoring system with hour of the day and time of year information. This was then further adapted to test the implementation of an ANN classifier model to determine if a leak is present.

The model that performed the best with current testing was the ANN model classifier with 97.78% accuracy when combining the statistical data that includes the time of day and time of year classifier information. This model has room for improvement as the statistical model currently may include leak flows in that can potentially skew the results. The models can also benefit by the inclusion of additional input parameters such as building occupancy data. Expansion of data sets will assist in improving model performance while minimizing potential class imbalance.

It is proposed that the current water management policy should be enforced that would have assisted in accurate logs of leak flows as they are detected with the duration and severity. This will then assist in model training while this leak can then be removed the data to improve the input data to the model as well. The authors are planning to conduct a further study on the impact of non-compliance on current water management policy.

References

1. Alwis, R.: Introduction to confusion matrix [classification modeling]. https://medium.com/tech-vision/introduction-to-confusion-matrix-classification-modeling-54d867169906
2. Baker, A.: What it's like to live through cape town's massive water crisis. https://time.com/cape-town-south-africa-water-crisis/
3. Boschetti, A., Massaron, L.: Python Data Science Essentials. Packt Publishing, Birmingham (2015)
4. Guppy, L., Anderson, K.: Global Water Crisis: The Facts. University Institute for Water, Environment and Health, pp. 1–16, September 2017
5. Joshi, P.: Artificial Intelligence with Python. Packt Publishing (2017). https://books.google.no/books?id=O1AoDwAAQBAJ
6. Mouton, F., Teixeira, M., Meyer, T.: Benchmarking a mobile implementation of the social engineering prevention training tool. In: Information Security for South Africa (ISSA), pp. 106–116, August 2017. https://doi.org/10.1109/ISSA.2017.8251782 (2017)
7. Raschka, S.: Python Machine Learning. Packt Publishing (2015)
8. The United Nations: Goal 6: Ensure Access to Water and Sanitation for All. https://www.un.org/sustainabledevelopment/water-and-sanitation/
9. The Water Project: Water in Crisis - South Africa. https://thewaterproject.org/water-crisis/water-in-crisis-south-africa
10. Welch, C.: Why Cape Town is Running Out of Water, and Who's Next. https://www.nationalgeographic.com/news/2018/02/cape-town-running-out-of-water-drought-taps-shutoff-other-cities/

Peace and War

ICT in Peace, War, Safety and Security

Brett van Niekerk

University of KwaZulu-Natal, Durban, South Africa
vanniekerkb@ukzn.ac.za

Information and communication technologies impact on peace, war, safety, and security, in particular with the increasing prevalence of cybersecurity, state-backed cyber operations, and advanced criminal organisations using technology to augment and support their operations. The growth in online disinformation and 'influence operations' through the use of social media, instant messaging, and online news media is an emerging form of soft power for national and sub-national actors to provide local, regional, and global influence.

This has a significant impact on law enforcement and the military, who need to deal with threats in both the physical and virtual domains (and a hybridisation of both) as well as rapid and open civilian communications relating to their operations. This necessitates capacity building and culture shifts at national levels and within the branches of law enforcement and the military. Within IFIP Technical Committee 9, Working group 9.10, the focus is on the areas of ICT uses and Peace and War. In HCC 14, four papers related to these themes were accepted for publication.

Louise Leenen, Joey Jansen van Vuuren, and Anna-Marie Jansen van Vuuren focus on the cybersecurity culture challenges unique to law enforcement agencies in Africa in their paper, "Cybersecurity and Cybercrime Combatting Culture for African Police Services." They contend that law enforcement agencies have a different culture to most organisations, and therefore previously developed cybersecurity culture frameworks may not be a good fit for law enforcement. Not only do they need to behave securely, they need to actively combat and investigate cyber-crimes. The authors propose a three-phase framework with 16 activity steps to promote an inclusive cyber-security culture in law enforcement.

"State's Capacity Building for Cybersecurity: an IR Approach" by Seiko Watanabe focuses on the Realism perspective of international relations through an analogy of nuclear deterrence with cybersecurity as the current situation in international affairs. He provides a discussion on deterrence theory related to alliances and cybersecurity. Using Grand Theory, areas of capacity building are discussed, and the paper finds an interdependence in the theories of Realism, Liberalism, and Constructivism when considering cybersecurity capacity building.

"An Analysis of Twitter during the 2017 Zimbabwean Military Intervention" by Brett van Niekerk, Martina Jennifer Zucule De Barros, and Trishana Ramluckan use the concept of social information warfare as a lens to view the online activity associated with the military intervention and subsequent anti-government protests. The research illustrates the challenges that face military movements in the age of social media, but also the fact that online tools are still powerful even in areas of low Internet penetration. However, in this case, the use of social media for command and control (as in other global events) was not present.

"Using Cyber Application towards Positive Psychology Interventions in Africa" by Carien van't Wout, Joey Jansen van Vuuren, and Anna-Marie Jansen van Vuuren illustrates the growing Internet availability in Africa, and in particular mobile phones, has provided improved communication of 'positive psychology' initiatives, ranging from coping strategies, sharing of agricultural information, to providing aid to children affected by conflicts on the continent. These initiatives show promise in supporting economic empowerment and overcoming the challenges experienced by many in accessing knowledge and positive psychology in particular.

Lastly, Tomoko Nagasako's "A Consideration of the Case Study of Disinformation and its Legal Problems," focuses on global cyber attacks that not only threaten industries and infrastructures, but also constitute a form of information warfare, affecting election results and democratic processes, and posing a threat to democracy.

Using Cyber Applications Towards Positive Psychology Interventions in Africa

Carien Van 't Wout[1](✉), Joey Jansen van Vuuren[2](✉), and Anna-Marie Jansen van Vuuren[2](✉)

[1] Council of Scientific and Industrial Research, Pretoria, South Africa
[2] Tshwane University of Technology, Pretoria, South Africa
jansenvanvuurenjc@tut.ac.za

Abstract. This paper describes "Positive Psychology" and its origins as well as its current status within the field of Psychology. Though the field is gaining rapid support in the rest of the world, it has not made the same progress in Africa, although its approach seems to form part of the continent's indigenous knowledge systems and practices. In order to comment on possible future developments in this field, Africa's fast increasing exposure to the cyber domain is illuminated, followed by examples of cyber domain interventions that have improved the well-being of Africans. Access to information and interactive communication seems to facilitate an increase in positive experiences, development of positive individual traits and positive communities, especially in the field of agriculture. This suggests that mobile applications and the cyber domain can be used as a platform for positive psychology interventions.

Keywords: Positive Psychology · ICT · Agriculture · Africa · Development

1 Background

We are currently living in a data-driven society. Data is regarded as a valuable resource in the digital age [1]. Still, data can be a potential source of weakness in the cyber domain where one is constantly under threat of cybercrime [2]. The United States Military Joint Publication 3–12, Cyberspace Operations, (2013), define the cyber domain as:

> *"A global domain within the information environment consisting of the interdependent networks of information technology infrastructures and resident data, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers"* [3].

The Information Society is another term used within the cyber domain referring to a society in which the creation, distribution, and manipulation of information has become the most significant economic and cultural activity: The information society is not only affecting human interaction, but additionally it is requiring the traditional organisational structures to be more decentralised, flexible and participatory.

The Fourth Industrial Revolution is contrasted with the Agrarian or Industrial society in that people use information more intensively in their decision-making

D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 207–221, 2020.
https://doi.org/10.1007/978-3-030-62803-1_17

activities, to communicate and compare, and to take greater control over their own lives. The Fourth Industrial Revolution also has a strong focus on the interactive use of Information Communication Technologies (ICT) by humans. The past two decades has seen substantial growth in access to ICT, particularly on broadband networks, mobile phones and other smart devices allowing access to the Internet. Furthermore, within the last five years the availability and use of broadband networks grew extensively, as illustrated in Fig. 1.

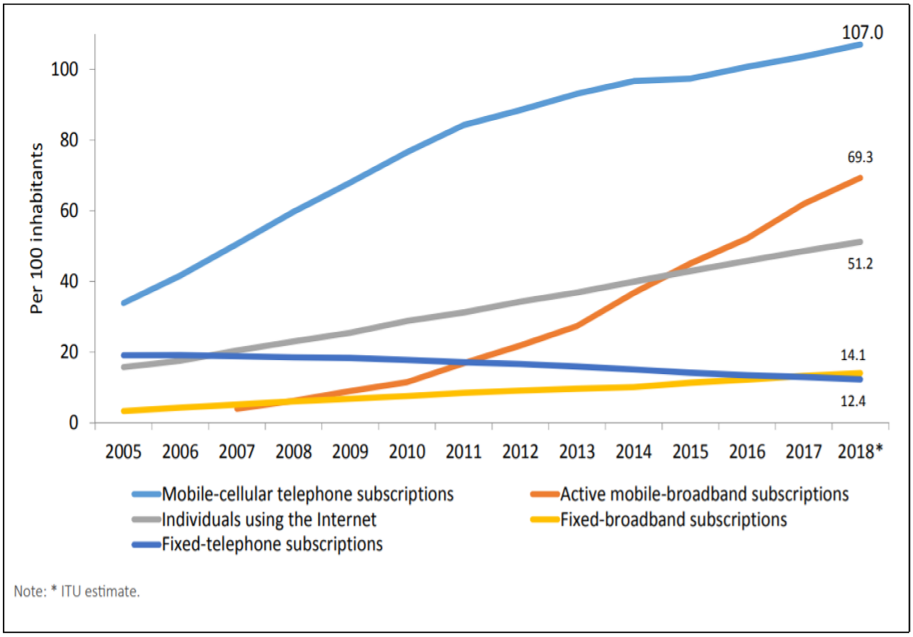


Fig. 1. Global ICT growth statistics from measuring the information society report, (2018) [4]

Between 2015 and 2016, a type of digital revolution took place on the African continent. This was the result of lower smartphone prices that allowed users, who would not have been able to afford computers, access to the internet. Thus, SmartPhone use doubled in Africa in the last two years,” [5]. Although the cost of data remains high in most African countries, service providers and developers are using this revolution by leveraging the power of mobile networks to transform services in the health, agriculture, education, energy and water management sectors [5]. Figure 2 provides the statistical support to illustrate this.

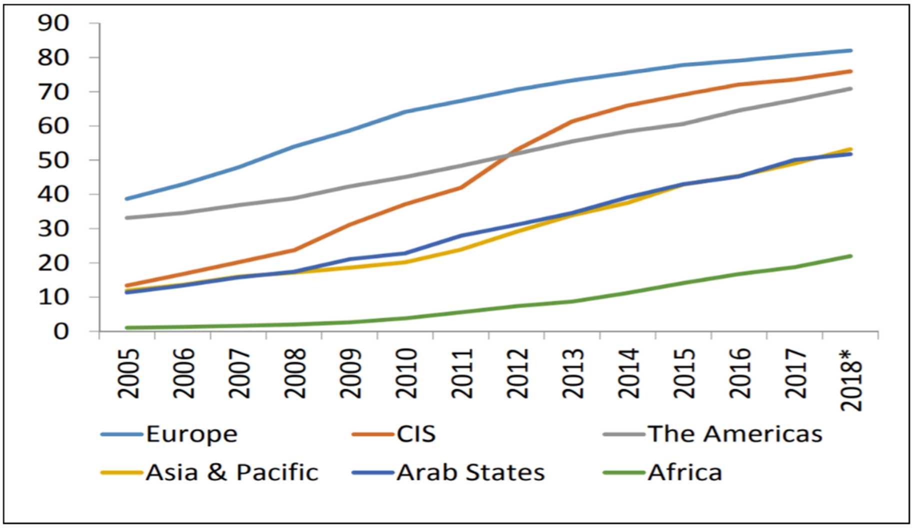


Fig. 2. Proportion of Internet users by region [4]

Significant technological advancement in mobile technology, has not only connected African users to their international counterparts via social media, but also granted them better access to information – this includes the news, awareness of local and international trends, as well as government and community support. This implies that the cyber domain facilitates the development of positive psychology constructs, e.g. self-determination, relatedness, belongingness, need for competence and autonomy. The next section will elaborate on this statement by describing positive psychology and its constructs.

2 Positive Psychology: An Introduction

> *At the meta-psychological level, [positive psychology] aims to redress the imbalance in psychological research and practice by calling attention to the positive aspects of human experience. At the pragmatic level, it is about understanding the wellsprings, processes and mechanisms that lead to desirable outcomes" [6].*

In the year 2000, the American Psychological Association president, Martin E. P. Seligman and co-author Mihaly Csikszentmihalyi, wrote an article in the *American Psychologist Journal* to introduce sixteen articles about a "new" psychology that concentrated on positive subjective experiences and positive individual traits as well as positive communities and institutions [7]. They argued that the aim of psychologists using this approach is to improve quality of life and to prevent the pathologies such as stress, anxiety and depression, rather than to focus primarily on treatment. Thus, positive psychology aims to increase the qualities and factors that enable individuals and communities to flourish, rather than to just endure and survive. The main positive psychology constructs discussed in their article are summarised in Table 1.

Table 1. Positive psychology constructs

Positive subjective experiences	Positive personality traits	The social context
Subjective well-being Contentment Satisfaction Optimal experience Hope Optimism Flow Happiness	Capacity for love Courage Self-determination Physical health Self-organising Self-directedness Interpersonal skills Adaptiveness Exceptional performance Creativity & talent Aesthetic sensibility Perseverance Forgiveness Originality Future Mindedness Spirituality Wisdom	Positive institutions that move people towards better citizenship: Responsibility Nurturance Altruism Civility Moderation Tolerance Work Ethic

It is generally accepted that psychology as a science originated in 1879, with Wilhelm Wundt’s establishment of the first psychology laboratory by in Leipzig, Germany. The aim was threefold: “curing mental illness”, “making people’s lives more productive and fulfilling”, and “identifying and nurturing high talent” [7]. Although positive psychology was only defined two centuries later as a scientific field, examples of earlier positive psychology studies included Watson’s work on effective parenting (1928), Jung’s search for the meaning of life (1933), and Terman’s research on marital happiness (1939). After the Second World War, thousands of psychologists found jobs treating mental illness at the Veterans Administration and academics only obtained grants for their research if it centered on pathology. This changed the focus within the psychology field. Although this period delivered positive outcomes in terms of understanding and providing therapy for mental illness (with its focus on assessing and curing individual suffering), the other two parts of the psychology profession’s original aims were sorely neglected. Seligman & Csikszentmihalyi’s article was a thus call for the profession to widen its scope to include positive human functioning for increased scientific understanding and effective interventions towards thriving individuals and communities.

Since then the work of these post-2000 researchers was consolidated, and these authors formed the Positive Psychology Steering Committee. Many other researchers need mention due to their considerable contributions to the field; these include Snyder and Lopez who edited the *Handbook of Positive Psychology* (2002) [8], followed by

Positive Psychology: Scientific and Practical Explorations of Human Strengths [9], sixteen special journal issues, as well as the establishment of *The Journal of Positive Psychology* in 2006 [10]. Studies of what is 'right' and 'positive' about individuals, communities and institutions yielded new findings and resources that enabled psychologists to improve their patients' quality of life and assist educators in nurturing high talent and learner growth. Major developments include the psychometric instruments measuring positive psychology constructs, such as the "Clifton Strengths Finder" and the "Clifton Youth Strengths Explorer" (the latter is focused on 10 to 14-year-olds). These measures have enabled strengths development programs in learning institutions and the workplace. In addition, social scientists have learned more about how people respond to emotional experiences in productive ways. Knowledge increased on making the most of positive emotions and curbing the effects of negative emotions, and subsequently about how these practices lead to positive life outcomes [9]. Rigorous investigations into the constructs of *resilience, self-efficacy, hope* and *optimism* also took place in the field of positive psychology [11].

In an opening article in the *Journal of Positive Psychology* (2006), Linley, Joseph, Harrington and Wood review the progress and possible future direction of the field. They attempt to redefine positive psychology and poste the question of where the field is now. One purpose of this paper is answering these questions, specifically in terms of the African contribution towards "the study of optimal human functioning" and to propose that the cyber domain can add value to this field.

3 Positive Psychology in Africa

"Positive psychology" may have been practiced in Africa before being termed as such. Many factors influence the practice of psychology on the continent, of which the most prominent seem to be culture, economy or development status.

3.1 Cultural Perspective

As North African countries are predominantly considered to be Muslim countries, one need to consider psychology's history from an Islamic perspective. Haque comments that Western psychology has not lived up to its professional goals, i.e. to help people understand themselves, the purpose of life, and how to live in a balanced and constructive manner. He argues that from a Muslim perspective, modern psychology assumptions are erroneous in proposing that human behaviour is observable only by the senses and therefore subject to quantification and measurement, while at the same time ignoring transcendental aspects. Hague comments that if psychology studies human behaviour and cognitive processes, it should include beliefs, attitudes, norms, customs and religious influences based on transcendental experiences and value systems [12]. These comments correspond to psychology's objectives.

In sub-Saharan Africa, "positive psychology" has often been compared to the philosophical concept of "Ubuntu". Scholars and traditional healers have argued from the African perspective that health cannot be achieved without achieving a balance of life with others and with the environment [13]. This worldview emphasizes the spiritual dimension (transcendental) which generates traditional knowledge as a valuable resource to be shared by all. Solidarity, caring and sharing are important in times of hardship, grief, illness as well as in good times (health, celebrations). Importance is placed on interpersonal relationships as well as the relationship between the physical and meta-physical. Lesonang (2016) explains that traditional healers follow a holistic approach to address the human experience of disease whilst modern medical practitioners attempt to heal only the affected part. They also have a dichotomous role of promoting the well-being of an individual and maintaining continuity in the way society functions. Traditional healers believe that it is their duty not only to alleviate suffering, but also to develop life in all forms [13]. It therefore seems that the African perspective of psychology is more oriented towards positive psychology principles in their assessment (scientific observation and measurement), diagnosis and treatment of mental illness.

3.2 Status of Positive Psychology in Specific Countries

In *Uganda*, plenty of opportunities exist to apply positive psychology. Because of civil wars, the HIV pandemic and frequent natural disasters, most people need some form of psychosocial rehabilitation. Physical disabilities are common, resulting from the rebel attacks, landmines and intentional mutilations. According to the CIA’s World Factbook, an estimated 1.8 million people were internally displaced in Uganda due to these conflicts and challenges. At the time of writing, the country had approximately 2.2 million orphans – mostly due to the HIV pandemic – as well as another 8 million vulnerable children [14]. The opportunity or need for positive psychology interventions are thus extensive; however, no university psychology programme or course currently exist that holistically centers on positive psychology [12]. However, despite professional psychologists’ central focus on mental illness, there are numerous activities demonstrating the presence of and application of it in Uganda. Educational, clinical, organisational and counselling psychologists, as well as medical doctors, nurses, social workers, counsellors, academics, traditional healers and religious or other leaders are involved in the practice of positive psychology. Examples of positive psychology applications include the following [15].

- HIV/AIDS counselling that emphasises positive living. Organisations such as Traditional Healers and Modern Practitioners Together Against AIDS (THETA) and The AIDS Support Organisation (TASO) give hope to those infected, promote good health and the development of positive attitudes towards life.
- Community-based rehabilitation (CBR) aimed at improving the well-being of persons with disabilities. Tertiary diplomas and degrees and Special Needs Education are offered in CBR. The focus is on empowerment of individuals and lifelong community participation.
- School guidance and counselling teaches children coping skills to deal with their problems independently and effectively.
- Teachers’ training includes a focus on their role to guide and/or counsel children to develop self-esteem, confidence, a sense of direction, interpersonal and problem-solving skills.

- Street children are trained in life skills with particular attention to independent living, coping with emotions, stress and problem solving in order to enable living a productive life away from the street.
- Life skills education is incorporated in primary school curriculums.
- Psychosocial rehabilitation to children affected by the war.
- A Community Resilience and Dialogue Project promotes well-being by promoting dialogue and small business development training, advanced business training and other economic development assistance tools.

Many of these positive psychology practitioners may, however, lack training and expertise. The emphasis is usually on community support, team building, life skills development, coping and forgiveness [15].

The challenges for the practice of positive psychology in Uganda are similar to those of other African countries, in that very few people voluntarily seek psychologists' help. Psychology is not very popular among citizens– which make it an unpopular study or career choice. This in turn hampers the development of culturally appropriate frameworks and methodologies. Positive psychology has not been developed formally and the term seems abstract to those that hear it. Most of those that apply its principles are unaware of the scientific study of it [15].

In the *Democratic Republic of Congo (DRC)*, having a mental disorder is a shame that extends to the family, because each individual is an integral part of a larger community. The family, clan or tribe also suffers and therefore treatment includes all these parties as well. Unfortunately, people with a mental disorder are often accused of witchcraft and therefore consultation is rather with a traditional healer than with a psychologist. Visiting a psychologist or psychiatrist for behavioral health concerns, or for improvement in individual positive psychology traits, is very rare. Limited mental health care facilities exist in comparison with the size of the country and its population. The DRC has only six mental health hospitals with 500 beds, no day treatment facilities and almost no mental health care practitioners [16].

Psychology is not a well-established discipline in DRC universities. Only two clinical psychology programmes exist and psychology remains more theoretical than practical. The focus is primarily on understanding the stages of human development for the training of primary and secondary school teachers. Job opportunities for psychologists are scarce to non-existent and leads to limited enrolment for these programmes at university level. The Congolese culture, like many other traditional African cultures, is described as one where the individual is not viewed as existing alone, but as existing corporately; the community (the whole) is more important than the individual (the parts). This creates a challenge for so-called Western psychology where the client/patient as an individual is paramount. The field positive psychology is still struggling to grow as a science as well as towards providing adequate mental health services to Congolese people. Thus, a need exists for a culturally appropriate therapeutic approach to mental health [16].

In *Kenya*, no universities were found offering Positive Psychology courses except for the Africana College of Professionals that offers a 3-day Certificate course [17].

In South Africa, Professor Irma Eloff, with research associates from Saudi Arabia, Zimbabwe and Nigeria, explores the state of Positive Psychology on the continent by asking participants in the profession to respond to the status and prospects of positive psychology in their respective countries. The six countries included are Lesotho, Malawi, Nigeria, Uganda, Zimbabwe and Algeria. The following is a summary of their quantitative and qualitative findings [18].

- Positive psychology is an emerging and limited field.
- When positive psychology is practiced, it is often implicit rather than explicit.
- The theoretical and conceptual framework is uncertain.
- Positive psychology is often linked to indigenous knowledge systems.
- Respondents are hopeful of a future for positive psychology in Africa.

The cyber domain provides numerous opportunities for improving the well-being of African societies – and by implication, for the enhancement of positive psychology constructs. Some case studies will be discussed in the next section.

4 Improving the Well-Being of Africans via the Cyber Domain

The scope and length limitations for this paper allows for the mentioning of a few examples, which is by no means an exhaustive list.

4.1 Enhanced Social Connectedness and Well-Being via Social Networking

According to The Oxford Handbook of Mobile Communication and Society, smartphones enabled opportunities and possibilities for communication, participation, creativity, learning, expressing identity and belonging, improving health and well-being and enjoying entertainment and games. Children with smartphones and internet access spend a large amount of time online, primarily for communication and entertainment purposes, but it also develops digital skills at an early age [19]. The positive outcomes of this includes an increased sense of relatedness, belongingness, need for competence and sense of autonomy

4.2 Economic Empowerment in Rural Agriculture

In Africa, the agriculture sector is a major source of employment. Therefore, if one wants to employ cyber applications to uplift a society, and thus practice positive psychology, one will need to incorporate agriculture. During an information communications technology conference in Kigali, Rwanda, ten of the best agriculture applications (apps) were identified. The table below provides a summary of some of these apps and their objectives as well as relevance to positive psychology [20].

Table 2. Agriculture apps developed for rural Africa

Application	Description	Objective
iCow	Dairy agricultural products accessed and subscribed to via a menu after dialling a short code. The system then sends messages to users at intervals, depending on the product choice	To increase farmer productivity through access to knowledge and experts; to encourage the development of a younger farming generation
Rural e-Market	Multilingual, easy and affordable app for smartphone, tablets or computers	Communicating market information; improve transparency and access to market information
mFisheries	A suite of open-source mobile and web applications for small scale fisheries. Comprises a virtual marketplace application which displays market prices using open data sources. Includes "GotFishNeedFish" with navigational tools such as compass and GPS, logging and retrieval application. It has training companions such as abbreviated first aid courses	Connects agents in the fisheries value chain
Esoko	Most popular African Agriculture platform. A customisable comprehensive platform to transform and manage information needs – with an all-in-one user-friendly interface, backed up with a deployment team to help users in all locations	Tracking and sharing of market intelligence. Links farmers to markets with automatic market prices and offers from buyers. Disseminate personalised messages based on crop & location. Manages extension officers and lead farmers with messaging
FarmerConnect	Cloud-based and mobile enabled platform that delivers personalised agricultural extension services and text/audio information in local languages to smallholders and farmers	Helps to stay connected with information and aiding agencies on a daily basis. Increase yields and income
M-Shamba	Interactive platform that provides information to farmers using mobile phone features, including cross-platform applications accessible in both smart and low-end phones and SMS. Currently used by 4000 farmers in Kenya to help them adapt to new technologies in rice farming	To provide information on production, harvesting, credit, marketing, weather and climate/ Customised information based on farmer location and crop/animal preference. Farmers can also share information on various topics with each other

These applications for smallholders and farmers (containing agricultural advice for better production results as well as information on current rates for specific commodities) has improved the lives of many Africans. In the past, such farmers or small business owners were exploited by dubious intermediaries that ‘made up’ the current prices for commodities to their own liking. Farmers could not verify the current buying and selling prices (the market) of their specific commodity. These computer applications have therefore not only mitigated exploitation, but have also improved productivity and subsequently the livelihood of these farmers [21]. In terms of positive psychology, well-being has been improved by reducing stressors such as hunger, poverty and under-nutrition and in turn increasing subjective experiences such as satisfaction, contentment, optimism and hope. The objectives of these apps may also lead to an increase in positive personality traits such as self-determination, self-organizing and positive emotions leading to physical health improvements.

4.3 Improving Literacy in Tanzania

Tanzania has one of the world’s largest youth populations and its youth are at the heart of Tanzania’s aspiration to become a middle-income country by 2025. The country’s economic and social progress as well as human development, however, depends on empowering and educating the youth with the skills needed to take this nationwide goal forward. The well-being of both individuals and their communities can be improved by quality education. It would lift families out of poverty and increase the country’s economic growth. Completing secondary education has been shown to strongly benefit individuals’ health, employment, and earnings throughout their lives. Yet millions of Tanzanian children and adolescents do not gain a secondary education or vocational training.

It is estimated that 5.1 million children aged 7 to 17 are not in school. Education ends for many children after primary school: only three out of five Tanzanian adolescents, or 52% of the eligible school population, are enrolled in lower-secondary education and even less complete their secondary education. Thus, secondary education can empower the youth with skills needed for sustainable development. This must include technical and vocational training, as well as education on citizenship and human rights. Access to essential information must be ensured to protect health and well-being.

Safe and equal enrolment for girls in secondary education can act as a powerful equalizer, ensuring all girls and boys have access to the same subjects, activities, and career choices. Formal vocational training is not available to all Tanzanian children who want it. Many children are rather taken into child labour - mostly to look after the family herds - to supplement their family’s income. Often these children work in exploitative, abusive, or hazardous conditions, in violation of Tanzanian law. Girls face many challenges on account of their gender; many marry before the age of 18 and thousands of adolescent girls drop out of school because of pregnancy [22].

The cyber domain can provide a number of solutions to address the challenges in Tanzania. One such cyber domain intervention to support Tanzania’s economic growth by improving the literacy of its youth deserves mention. The United Nations High Commissioner for Refugees (UNHCR) partnered with San Francisco-based non-profit enterprise *Worldreader* in an experimental project to provide four secondary schools in Tanzania with e-readers containing books in Swahili and English. The project provided some 2,300 students across two settlements with access to educational reading. The residents of both settlements and the surrounding communities can use the Worldreader

Mobile service. This provides them with access to a library of African and international books using a data-enabled mobile feature phone. Worldreader currently has e-reader programmes across nine African countries, including Rwanda, Ghana and Malawi [23]. Joyce Mends-Cole, UNHCR Representative in Tanzania, predicted that the impact of the initiative would be immeasurable; that it would increase literacy and expand mental and physical horizons - helping to provide quality education to large numbers of students. Mends-Cole argued that the access to this type of reading material will allow the youth to dream about greater possibilities, and for girls and women, the impact would be exceptionally good. The Co-Founder of Worldreader reported that their organization demonstrated that access to digital books can significantly improve early grade literacy outcomes. A video insert in the report about the e-Reader project, shows children using the e-Reader and commenting on how this device positively affected their lives. They read daily and want to keep on learning to follow a professional career one day. It is therefore evident that these devices have impacted positively on the individual traits of the users' subjective well-being [23]. The next phase towards youth development and education in Tanzania is to provide access to the internet. Camfed.org (2018), provides feedback on the positive outcome of this initiative; stating how a combination of e-reader technology and mentoring from young women has helped students succeed in their primary and secondary education. The literacy project using cyber tools, implemented by Camfed and Worldreader between 2015 and 2017 has driven up learning outcomes for marginalized children in Iringa district. Students can now access a wide variety of English language books and use a vocabulary function to look up unfamiliar words. The devices have encouraged a more collaborative way of learning and facilitates whole school literacy activities including debate competitions [24]. A Secondary student in Tanzania commented that e-readers have gave her a technological solution to simplify her learning [24].

4.4 Enabling a Financial Infrastructure for Migrant Workers: MPesa

Many rural areas lack banking infrastructure. The normal banking procedures for sending money abroad requires steep banking charges. This makes the transfer of small amounts by migrant workers a challenge. In Kenya, most of the population subscribes to a mobile payment service, of which the most popular choice is M-Pesa.[1] This mobile app has made a dramatic impact in Africa. The system was first launched by Vodafone's Safaricom mobile operator in 2007 as a simple method of texting small payments between users. Currently, it boasts 30 million users in 10 countries and includes a range of services including international transfers, loans, and health provision [25]. MPesa improves the livelihoods of communities by providing an accessible source of income from relatives who are migrant workers. This positive application in turn may produce further positive outcomes as per the positive psychology paradigm, such as optimism, hope, self-determination and positive communities.

Table 3 provides more examples of cyber domain solutions for achieving positive psychology related objectives for improving the wellbeing of Africans identified from

[1] "Pesa" means "money" in Swahili.

the challenges mentioned in this section. The intent of these examples is to illustrate that the cyber domain can be utilised in a directed manner – other than acknowledging the mere usefulness of ICT - to develop and/or increase positive psychology constructs and improve overall mental health.

Table 3. Examples of cyber solutions for improving the wellbeing of Africans

Objectives Related to Positive Psychology	Examples of Cyber Related Solutions	Example Instances	References
Provide tertiary training opportunities focused on positive psychology	Online learning in positive psychology	coursera.org; udmy.corg	https://mindisthemaster.com/positive-psychology-certificate-online-practitioner/
HIV/AIDS counselling emphasising positive living	Mobile app for people living with HIV/AIDS	Life4me + , Care4Today	www.healthline.com/health/hiv-aids/top-iphone-android-apps#care4today
Increased well-being for people with disabilities	Mobile apps for the disabled	Be My Eyes, Dragon Dictation	https://access2mobility.com/top-8-mobile-apps-for-persons-with-disabilities/
Overall wellbeing	Apps for overall wellbeing	Moodfit	www.verywellmind.com/best-mental-health-apps-4692902
Learning coping skills	Apps for coping skills	MoodMission	www.verywellmind.com/best-mental-health-apps-4692902
Science-based activities and games that are meant to reduce stress, build resilience, and overcome negative thoughts	Mobile apps to improve happiness	Happify	https://www.verywellmind.com/best-mental-health-apps-4692902
Life skills training for school children and street children (and to assist teachers and trainers)	Apps to Teach Responsibility and Life Skills to children	Faces I Make, Toontastic	https://blog.bit-guardian.com/top-10-apps-for-kids-to-learn-responsibility-and-life-skills/
Community dialogue through online platforms	Online platforms	Dialogue-Africa blog,	www.dialogue-africa.comEyalablog
Economic and Community upliftment by means of improved agricultural production.	Mobile farming apps	See Agricultural apps in Table 2	
Literacy: provide basic and secondary education to those that do not have access to schooling	Access to educational reading by means of e-readers	Worldreader Mobile service	www.telegraph.co.uk/technology/news/10653277/UN-launches-e-reader-initiative-in-Tanzania.html

These are only a few examples of many to illustrate the positive impact that the cyber domain has had on the lives of Africans. It further suggests that mobile apps and the cyber domain can be used as platform for positive psychology interventions. Specifically seen in the light of the current lack of trained practitioners in Africa, such interventions could be conceptualised and designed by skilled practitioners elsewhere or by local cyber professionals and remotely disseminated in a coordinated manner to facilitate the positive psychology interventions. Psychologists and other practitioners can also be supported in carrying out positive psychology interventions. If they communicate their requirements to cyber professionals, tailor-made tools can be developed to serve specific needs. The cyber domain can furthermore enable education in the field of positive psychology e.g. access to electronic resources and online courses.

5 Conclusion

Positive Psychology made a dramatic return to the psychology stage in 2000. Since then it has made remarkable progress in the western world as a recognised field of practice. In Africa, it has had limited recognition (as referred to in the Western theoretical and conceptual framework) in tertiary education or formal application but seem to have been practiced as part of traditional, cultural or religious systems for decades. The application of positive psychology principles is evident in many activities on the continent – all aimed at improving the well-being and livelihoods of its inhabitants.

The growth of access to ICT, especially mobile phones, the internet and the expanding cyber domain in Africa has granted users more access to information, connectedness and community support as well as economic empowerment. People can use these platforms to communicate and compare, to learn, to manage their businesses more effectively and to take control of their lives.

Mobile apps can assist those with little or no access to education and information support or knowledge building. Access to information and the enablement of interactive communication seems to facilitate an increase in all the positive psychology construct categories as mentioned in Table 1, namely positive experiences, development of positive individual traits and positive communities. The examples discussed illustrated how positive psychology constructs could be developed. This suggests that mobile apps and the cyber domain can also be used in a deliberate approach, as a platform for positive psychology interventions. It can be developed towards improving the quality of life and productiveness of people; for nurturing of high talent (e.g. math and structural design games) and learner development (literacy apps). A myriad of strength development programmes can be developed for learning institutions and work environments. Interactive apps can be developed to engage people in dialogue and to provide customised training and support towards economic empowerment, which will improve and enhance individual and community well-being. Such tools can be developed by multi-disciplinary cooperations or talented individuals. It can be remotely developed also, thereby overcoming some of the current challenges experienced for the application and future expansion of positive psychology in Africa.

References

1. Komukai, T.: People first: one can hide the data but not the truth. https://www.ifipnews.org/human-centric-computing-data-driven-society/. Accessed 06 Jan 2020
2. Metcalf, A., Scott, D.: The cyber domain. Mar. Corps Gaz. Quantico Prof. J. US Marines **99** (2015)
3. Scaparrott, C.M.: Cyberspace Operations, Washington DC, 5 February 2013
4. ITU: Measuring the Information Society Report 2018, vol. 2. ITU Publications, Geneva (2018)
5. Dahir, A.L.: SmartPhone use has doubled in Africa in two years. Smart Future. https://qz.com/748354/smartphone-use-has-more-than-doubled-in-africa-in-two-years/. Accessed 25 Sept 2016
6. Linley, P., Joseph, S., Harrington, S., Wood, A.: Positive psychology: past, present, and (possible) future. J. Positive Psychol. **1**(1), 3–16 (2006). https://doi.org/10.1080/17439760500372796
7. Seligman, M., Csikszentmihalyi, M.: Positive psychology an introduction. Am. Psychol. **55** (1), 5–14 (2000). https://doi.org/10.1037/0003-066X.55.1.5
8. Snyder, C.R., Lopez, S.J.: Oxford Handbook of Positive Psychology. Oxford University Press, Oxford (2009)
9. Lopez, S.: Major Developments in Positive Psychology. Galllup, Washington, D.C (2006)
10. Coetzee, S., Viviers, R.: An overview of research on positive psychology in South Africa. S. Afr. J. Psychol. **37**(3), 470–490 (2007). https://doi.org/10.1177/008124630703700307
11. Magaletta, P., Oliver, J.: The hope construct, will, and ways: their relations with self-efficacy, optimism and general well-being. J. Clin. Psychol. **55**(5), 539–551 (1999). https://doi.org/10.1002/(SICI)1097-4679(199905)55:5%3C539:AID-JCLP2%3E3.0.CO;2-G
12. Haque, A.: Psychology from islamic perspective: contributions of early muslim scholars and challenges to contemporary muslim psychologists. J. Relig. Health **43**(4), 357–377 (2004). https://doi.org/10.1007/s10943-004-4302-z
13. Lesonang, N.: Traditional healing and mental health of military forces in Africa. In: Gaj, V. D. (ed.) Military Psychology for Africa. African SUN Media, Stellenbosch, South Africa, pp. 237–260 (2016)
14. Central Intelligence Agency: The World Fact Book. https://www.cia.gov/library/publications/the-world-factbook/geos/print_ug.html. Accessed 20 June 2020
15. Chireshe, R., Ojwang, P., Rutondoki, E., Byamugisha, G.: Positive psychology: the case of uganda. J. Psychol. Afr. **18**(1), 195–198 (2008). https://doi.org/10.1080/14330237.2008.10820186
16. Ikanga, J.: Psychology in the democratic republic of the congo: its struggles for birth and growth. Psychol. Int. **25**(4), 5–6 (2014)
17. Kenyaplex, Colleges and Universities offering certificate in positive Psychology in Kenya. https://www.kenyaplex.com/courses/6664-certificate-in-positive-psychology-positive-psychology.aspx. Accessed 19 Sept 2017
18. Eloff, I., Achoui, M., Chireshe, R., Mutepha, M., Ofovwe, C.: Views from Africa on positive psychology. J. Psychol. Afr. **18**(1), 189–194 (2008). http://dx.doi.org/10.1080/14330237.2008.10820185
19. Ling, R., Fortunati, L., Goggin, G., Lim, S.S., Li, Y.: The Oxford Handbook of Mobile Communication and Society. Oxford University Press, Oxford (2020)
20. Fripp, C.: Top 10 mobile agriculture applications. IT News Africa. http://www.itnewsafrica.com/2013/11/top-10-mobile-agriculture-applications/. Accessed 18 Sept 2013

21. Daily Nation, Mobile apps making farm work easier, ‘cooler’. http://www.nation.co.ke/business/seedsofgold/ICT-in-agriculture-farming/2301238-3332292-rx75s0/index.html. Accessed 6 Aug 2016
22. HRW.Org: “I Had a Dream to Finish School” Barriers to Secondary Education in Tanzania. https://www.hrw.org/report/2017/02/14/i-had-dream-finish-school/barriers-secondary-education-tanzania. Accessed 26 Sept 2017
23. Williams, R.: UN launches e-reader initiative in Tanzania. The Telegraph - Technology News. http://www.telegraph.co.uk/technology/news/10653277/UN-launches-e-reader-initiative-in-Tanzania.html. Accessed 18 Sept 2014
24. CAMFED: Reference: tech-assisted learning takes off in Tanzania. https://camfed.org/latest-news/ereaders-succeed-in-tanzania/. Accessed 10 July 2018
25. Monks, K.: M-Pesa: Kenya’s mobile money success story turns 10. CNN Marketplace Arfica. http://edition.cnn.com/2017/02/21/africa/mpesa-10th-anniversary/index.html. Accessed 26 Sept 2017

States' Capacity Building for Cybersecurity: An IR Approach

Seiko Watanabe(✉)

Yokohama National University, Yokohama, Kanagawa, Japan
seikowatanabynu@gmail.com

Abstract. This paper discusses the current circumstances of security in cyberspace, such as cyber armies and cyber intelligence. Cyber intelligence plays a vital role in the balance of power. Most importantly, this paper explores previous studies of the International Relations (IR) theory of Realism. Cybersecurity can be applied as the equivalent of a nuclear deterrent of Realism and is inspired by the sense of the threat that allied countries felt in regard to cybersecurity. Countries utilize capacity building for military affairs, economics, and administration for cyber deterrence. Even though the circumstances of cybersecurity are deeply affected by the deterrence theory of Realism, concepts of capacity building for cybersecurity are derived not only from Realism but also from Liberalism and Constructivism. In the end, through this paper, I found that there is an interdependence of Realism, Liberalism, and Constructivism.

Keywords: Cyberattack · Capacity building · International Relations

1 Introduction

In recent years, the number of cyberattacks has rapidly increased, and both developing and developed nations have been unable to manage them properly because of the rapid increase in their number. This is because technologies of developing countries are not matured, and governments have been unable to catch up with the latest technology for cyber security. The intention of this paper is to describe the current situation of cyberwars and intelligence and it suggests that a sense of threat is closely tied to Realist theories of International Relations (IR). In particular, the balance of power (a Realist concept) and deterrence theory create a system of capacity building internationally. This system is well-supported by these theories, and the paper presents an analysis of the future of cybersecurity.

Many previously believed that cyberspace was an illusory world. However, in the late 2000s, nations began to notice cyberattacks perpetrated by other nations. There are two main categories of cyberattacks: (1) attacks on a nation's decision-making capabilities (intervening in elections or stealing/defacing information gathered by the government), and (2) attacks on important infrastructure (disrupting the operations of banks, hospitals, or power plants).

There are two significant examples of cyberattacks. In 2007, Estonian ministries, banks, and media suffered a serious attack by cyber terrorists, and in 2010, Iranian nuclear facilities were destroyed by two sophisticated worms, Stuxnet and Flame,

D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 222–232, 2020.
https://doi.org/10.1007/978-3-030-62803-1_18

which targeted and destroyed 58% of all hardware. In fact, a number of developed countries have established branch offices in developing countries; therefore, if cyberattacks occur at these branch offices, there is a high possibility that information regarding developed nations can also be stolen or damaged by cyberterrorists. Hence, developed nations need to invest in infrastructures or administrative resources to reinforce the defenses of developing countries.

2 Previous Studies of Cyberspace

2.1 Internet Intelligence and Cyberwar

John Arquilla of the US Navy's Naval Postgraduate School and David Ronfeldt of the Rand Corporation have defined cyberwar as a military operation conducted according to information-related principles. According to Richard A. Clarke and Robert K. Knake, "cyberwar" is defined as one government attacking the computers or networks of another government in order to disrupt infrastructure or steal important national security information [1]. Often, cyberwars are waged through intelligence. According to Michael Herman, intelligence activities on the Internet play a vital role in current IR [2]. The Internet allows for remote intelligence operations and keeps costs reasonably low. On the other hand, intelligence agencies are a controversial part of the modern state even though international law allows covert activity. According to Scott and Silver, accumulating information by intelligence does not contravene customary law [3, 4]. In fact, Scott notes that international human rights laws do not address spying, and governments cannot apply the actual rules that were established in the Convention.

According to Jon Swartz, in 2003, hackers attacked the computer systems of NASA and Sandia National Laboratories [5], utilizing computer viruses and worms. The hackers directly accessed military and government computers and checked private and government documents [5]. In the Republic of Korea, the government, media corporations, and political parties were attacked by North Korea in June 2013. In order to prevent such attacks, intelligence communities provide information to commanders and gather, process, analyze, disseminate, and assess information. Intelligence plays a vital role, especially in times of war, as the infrastructure is heavily reliant on the capabilities of cyber forces and the infrastructure of the Internet.

Additionally, *Cyberspace and International Relations* describes how specialized military cyber forces can sneak into an opponent's central infrastructures [6]. An example of specialized military cyber forces sneaking into the infrastructures comes from Estonia in 2007. As a result of cyberattacks on the Estonian government and citizens, cash machines and online banking services were compromised, and telecommunications, transportation systems, and the power grid were destroyed [7]. Estonia was particularly vulnerable because the government and infrastructure relied heavily on the Internet. The country, an enemy of Estonia, is alleged to be behind the attack.

The most troubling aspect of the Internet is that it is an anarchic world [8]. Despite several efforts by the United Nations (UN) to bind UN member states to act reasonably and create information security [9][1], conflicts between US allies and former Soviet nations have persisted. Hence, there are no solid norms governing the Internet or holding accountable countries that act unethically.

Thus, developing nations are targeted by Internet terrorists because of their vulnerable online infrastructure [10], and developed countries often capitalize on capacity building for developing countries. The cyber-developed nations spend money on governments, infrastructure operators, cultivation of human resources, crime prevention, and technology in developing countries for the sake of international security. According to Eade and Williams, capacity building is "strengthening people's capacity to determine their own values and priorities and to organize themselves to act on these" [11]. According to Echebarria, Barrutia and Aguado [12], capacity building is derived from the UN's Local Agenda 21 and is a critically important tool for conducting a set of sustainability policies. Though often quite similar methodologies are available for preparing LA21, each municipality has its own characteristics and idiosyncrasies and must, therefore, establish its own means of acting [13]. Agenda 21 is an encompassing plan of action on the global, national, and local scales for the members of the UN and for governments and other leading groups such as non-governmental organizations (NGOs). In 1992, within the UN's Sustainable Development Goals, the Sustainable Development Knowledge Platform, Agenda 21, the Rio Declaration on Environment and Development, and the statement of principles for the Sustainable Management of Forests were agreed by more than 178 governments at the United Nations Conference on Environment and Development held in Rio de Janeiro, Brazil. The Commission on Sustainable Development was created in December 1992 to ensure effective follow-up of the conference's principles and to observe and report on the implementation of the agreements at the local, national, regional, and global levels. It was permitted that a five-year review of the progress of the Earth Summit would be conducted by a special session of the UN General Assembly in 1997. According to the Japan International Cooperation Agency, JICA, developed nations should not simply fill the gap in technology between developed and developing nations but encourage developing countries to acquire proper knowledge and foster decision making on cyber policy.

2.2 Cyber Intelligence

According to Rafal Rohozinski [14], the main objective of cyber operations is to influence or control computer networks and operate and defend friendly critical cyber infrastructure and key resources. In addition, cyber operations attack critical and hostile

[1] The United Nations Group of Governmental Experts (GGE) on "advancing responsible State behavior in cyberspace in the context of international security" (formerly: on Developments in the Field of Information and Telecommunications in the Context of International Security) is a UN-mandated working group in the field of information security. Six working groups have been established since 2004, including the GGE 2019-2021. The UN GGE can be credited with two major achievements outlining the global agenda and introducing the principle that international law applies to the digital space.

cyber assets in cyberspace, which is known as active defense. This paper proposes appropriate roles and responsibilities for cyber intelligence within cyber operations.

Cyber operations are a sequence of tactical maneuvers that have strategic cyber objectives. This includes pre-cyber task orders, such as the US Air Force's pre-air tasking order. The pre-air tasking order is defined as a procedure "used to task and disseminate to components, subordinate units, and command and control agencies projected sorties, capabilities and/or forces to targets and specific missions" for three days from the outbreak of war; it normally "provides specific instructions to include fighter call signs, targets, weapons, controlling agencies, etc., as well as general instructions." The pre-cyber task order "guides cyber-attack according to the assigned procedure at each phase in real-time" [10]. Especially, according to the book, the roles and responsibilities of cyber intelligence at respective phases of cyber operations. Cyber operations are practices related to defense, assurance, and attack [that] achieve objectives in or through cyberspace. While a cyber-operation is being conducted, cyber intelligence must properly support cyber commanders and units to ensure cyberspace intelligence superiority. Cyber intelligence is a cyber-discipline that utilizes accumulated information accumulations and analysis approaches to provide direction in decision making to cyber commanders and cyber operation units. This is a key role in cyberattack and defense. The information collected is requested by cyber commands and units and is disseminated to relevant departments. Cyber intelligence is a criticaly important factor in the cyber operations cycle.

2.3 Cyber Threats in International Relations (IR)

According to Damien McGuiness and USA Today [7, 15], cyber threats occur across academic disciplines, including the study of computers, media, literature, engineering, and policy. However, International Relations (IR) cannot catch up with the real situation and create appropriate theories, despite the present circumstance of cyberattacks being emergent. According to Eriksson and Giacomello, very few attempts have been made to apply IR theory in analyzing this development [16–18]. Research that has focused particularly on aspects of the creation of information-age security threats has not been much influenced by theory or is mostly outdated [19–22].

2.4 Cyber Wars

Thomas Rid is a prominent scholar of the risk of information technology during conflicts. His book, *Cyber War Will Not Take Place*, states that cyberwar never happened in the past, is not taking place now, and is unlikely to occur in the future. He summarized a cyberwar as comprising a potentially lethal, instrumental, and political act of force conducted through malicious code. Rid also provided nuanced terminology for cyberattacks. All politically motivated cyberattacks are merely refined versions of three activities: sabotage, espionage, and subversion [23].

Ryan Maness also mentioned what cyberattacks were. He believed that nations have a manner that is based on strategies [24, 25]. Cyberattacks can yield a high return for low costs for nations. Gia also claimed that the offensive nation in a cyberattack agitates people, creating disruptive social movements in defensive countries. There are three strategies for the nation on the offense: invalidating domains, reorganizing domains, and instigating political discord.[2]

3 Deterrence Theory and Alliances for Cybersecurity

Today, in the arena of cybersecurity, scholars have begun to consider whether the strategies used for nuclear weapon deterrence might apply to the present conflict in order to fill the gap between real cyber situations and academic research. This concept derives from the idea that any country that possesses a weapon, they are able to come under the control of an enemy, which ensures the security of the country, an idea born during the Cold War. Thomas C. Schelling stated that, at that time, both the Soviet Union and the United States had sufficient nuclear power to both threaten and deter one another [27]. This balance of power was called mutually assured destruction [27]. In deterrence theory, global political stability can be accomplished because countries know that the costs of using nuclear weapons are greater than the gains. In addition, the idea of mutually assured destruction serves as a base for the offense-defense theory, which is an essential Defensive Realist theory [28]. Defensive Realism argues that nations support the status quo to maximize the power of their political and military forces. For alliances, a nuclear umbrella is a safeguard against a non-nuclear allied state. The core idea of deterrence theory is that nations should prepare for threats and defend their allies. Despite the problem that the origin of a cyberattack cannot be known for certain because of the use of Tor, which is an anonymous modification application that disguises IP addresses, the core idea of nuclear deterrence remains the same for cyber deterrence. In both the theories of deterrence, mutually assured destruction serves as a base for the offense-defense theory, smaller allied countries can receive benefits by depending on big countries and all cooperate to ensure the safety of cyberspace: cyber-developed countries engage in capacity building for the sake of less developed countries. In the following sections, the types of capacity building are explained.

This desire to deter is a result of feeling threatened. Countries that foresee the possibility of the compromising or even destruction of their infrastructure or financial institutions or fear cyberwar ally with one another to prepare for such attacks.

[2] While these studies seem to be persuasive, the present attitude towards cyberattacks has changed since the Clinton administration. Richard Clarke, who worked with the Clinton Administration's national coordinators of international security, stated that cyberattacks from opponent nations destroy the Internet network for both citizens and government. Additionally, 9-11 was an epoch-making event for the United States and the rest of the world. The US government is concerned that major infrastructure like the Internet is a likely target for an attack [26].

4 Capacity Building

As the previous chapter indicated, capacity building is derived from the UN's Agenda 21, which aims to stabilize cyber conflict worldwide. There are three areas in which capacity building can be applied using the Grand Theory, which can entail a particular conceptualization of capacity building (Table 1).

Table 1. Capacity Building concepts

Types	Explains	Applicable IR theory
Capacity Building for military affairs	Conducting operations, conflict prevention, and doctrine enforcement	Realism
Capacity Building for Economics	Providing financial support to share the same Internet infrastructures	Liberalism
Capacity Building for administration/norms of law enforcement	Helping developing countries' systems of law enforcement	Constructivism

4.1 Military Capacity Building

According to David Guy, a researcher at the Australia Strategic Policy Institute, building military capacity aims to enforce armed exercises, prepare for cyber-attacks from enemies, conduct operations, prevent conflict, and enforce doctrine, which is same as traditional Realism. Martin Hall has written "Realist theories, have indeed been put to use by the armed powers" [31]. According to Prashanth Parameswaran in The Diplomat [31], Association of Southeast Asia Nations (ASEAN) states cooperate with Japan to boost their cyber capabilities. In fact, Singapore's new ASEAN Cyber Capacity Program was announced at the inaugural ASEAN Ministerial Conference on Cybersecurity in October 2016 [32]. The Philippines also announced the launch of a cybersecurity working committee within the ASEAN Defense Ministers Meeting Plus in 2016. Another key movement for cyberspace is the ten-million-USD ASEAN Cyber Capacity Programme (launched by Singapore in 2016). This program enhances cybersecurity expertise across the region. In order to pursue a peaceful cyber world, the program also launched the Singapore-ASEAN Cybersecurity Centre of Excellence in 2019. The center is based in Bangkok at the ASEAN-Japan Cybersecurity Capacity Building Centre, launched in September 2018, which seeks to prevent cyberattacks. Almost 700 of the cybersecurity personnel who work there are from Southeast Asia and have graduated from Japanese-designed programs that include instruction in cyber defense, digital forensics, and malware analysis [33].

4.2 Economic Capacity Building

Capacity building for economics is primarily digitalization support, which is believed to enhance a nation's economy. This idea is derived from Liberalism, an IR concept. In Liberalism, scholars believe that economic interdependence greatly influences nations' relations. According to Robert Gilpin, "liberalism assumes that a market arises spontaneously in order to satisfy human needs [31]. Capacity Building also tries to satisfy citizens demands by aiding cyber infrastructures from developing countries to developing nations. For example, the United Kingdom has contributed to the economy of the Commonwealth through cyberspace capacity building in accordance with the Sustainable Development Goals. The United Kingdom invests in cybersecurity in developing countries, assuming that they will be able to utilize the technology by themselves [34]. By promoting investment and trade of goods and services in the Commonwealth through democracy and fair competition not only have Commonwealth countries benefitted, but the UK has as well. The UK government enhances the importance of building resilient digital economies and needs of ITC investments to commonwealth countries [35]. Also, the UK government invest partnership in Africa, Asia, the Pacific and the Caribbean for improving digital economies [36].

The United States has helped Nigeria to prosecute an international cyber-fraud scheme. Ayofe and Oluwaseyifunmitan [37] suggested that the erection of a structure for the implementation of information assurance in critical sectors of the economy (such as public utilities, telecommunications, transport, tourism, financial services, the public sector, manufacturing, and agriculture) and developing a framework for managing information is necessary for cyber-developing countries. According to Niels Schia [38], cyber capacity building and digitalization of the economy are deeply connected.

4.3 Normative/Administrative Capacity Building

Capacity building for administration and norms aids developing countries' systems of law enforcement, which deeply connect to an idea of Constructivism, a concept of International Relations (IR). According to Nicholas Greenwood Onuf, the Constructivists believe that rules is important for society. Governed rules give people perspectives about how they should behave in a society [39].

In fact, developing countries want to improve law or rules of cyber security in developing countries through Capacity Building. For instance, the Estonian government, which is a cyber-developed country, emphasizes the importance of capacity building for fighting cybercrime, and it introduced the European Council Convention on Cybercrime (also known as the Budapest Convention) to developing countries [36]. Estonia and partner institutions from the United Kingdom and the Netherlands have been supporting the cyber development of countries in Africa and Asia. The Cyber Resilience for Development project will last until June 2021; the project appears in the cybersecurity yearbook published by the Estonian Information System Authority. According to the Estonian government, [40], "the activity has been launched in Mauritius, Sri Lanka, Ghana, and Botswana. The purpose of the mission is to increase awareness about cybersecurity, help develop cyber strategies and action plans, enhance the capability of the teams for handling cyber incidents, and share the experience with providers of vital services and institutions of the state" [41]. In addition, enhancement of computer security and incident response teams, protection of critical information infrastructure and regulation, risk management and crisis exercises, cyber-related laws and strategies, and cyber hygiene and awareness have been strategically established. Sri Lanka and Mauritius have demonstrated an interest in Estonia's experiences with dramatic growth through e-government. The Estonian government incorporates with the State Infocommunication Foundation (RIKS), and a consortium of private sector companies, including Cybernetica, Dell EMC, Ericsson, OpenNode, and Telia (Resource: e-governance, e-estonia,).

5 Future Implications: GGE and Capacity Building

In spite of the continuing policy of Realism in cyber security and the state of anarchy, there is a tendency towards cooperation among nations in the UN. A Group of Governmental Experts on advancing responsible state behavior in cyberspace in the context of international security was established in 2004. According to the UN website, the members of the GGE countries are Australia, Brazil, China, Estonia, France, Germany, India, Indonesia, Japan, Jordan, Kazakhstan, Kenya, Mauritius, Mexico, Morocco, Netherlands, Norway, Romania, Russian Federation, Singapore, South Africa, Switzerland, United Kingdom, United States, and Uruguay [42]. The GGE is planning to hand in its final report to the UN in 2021. Sash Jayawardane, Joris Larik, and Erin Jackson states that GGE is now trying to determine how existing international law can be applicable to cyberspace [43]. Especially, in the GGE, the above nations discuss how international law can be applied to cyberspace. Also, GGE countries support Capacity Building [45]. In the near future, there is a great possibility that nations conduct capacity building while appreciating benefits of rules of GGE.

6 Limitations

The intent of this paper is to analyze categorical translation into realism, liberalism, and constructivism, while applying capacity building to the aforementioned theories. However, in the field of IR, some important studies examining realism, liberalism, regime theory, and global governance already exist. This paper fails to specifically mention regime theory and global governance theory. Furthermore, research on regime theory considering capacity building in cyber spaces already exists, and some scholars have already analyzed this in terms of global governance such as the Paris Call, Commission on the Stability of Cyberspace report. It is necessary to employ these studies and have an independent discussion as well. Global governance is a critical philosophy that relates to the GGE and other Internet regimes. In a future academic thesis, I would like to include regime theory and globalization theories in a comprehensive theory for a better discussion of cyber security.

7 Conclusion

This paper introduces a general description of the types of cyberattacks and discusses previous studies of Internet intelligence, cyberwar, and theory for cyberspace. According to previous studies, nuclear deterrence can be applied to incidents of cybersecurity. Such deterrence still applies to the present situation even though attribution problems exist that make identifying the source of a cyberattack almost impossible. In response to these circumstances, nations cooperate in cybersecurity efforts through capacity building. To illustrate capacity building clearly, this paper analyzed military, economic, and normative concepts of capacity building through example cases by applying the IR Theories. The cybersecurity field is still quite new, but scholarly work in the field is needed to prevent conflict in cyberspace. Especially, capacity building is quite useful to the present situation. There are three definitions of capacity building for cyber threats: Realism, Liberalism and Constructivism, which are core ideas of IR. The most important finding in this article is that the IR ideas of Realism, Liberalism and Constructivism exist in capacity building of cyber.

References

1. Clarke, R.A.: Cyber War: The Next Threat to National Security and What to do About it. Tantor Media, Oxford (2014)
2. Herman, M.: Intelligence Power in Peace and War. Cambridge University Press, Cambridge (1996)
3. Scott, R.D.: Territorially intrusive intelligence collection and international law. AFL Rev. **46**, 217 (1999)
4. Silver, D.B.: Intelligence and counterintelligence. In: Moore, J.N., Turner, R.F. (eds.) National Security Law, 2nd edn, pp. 935–965. Carolina Academic Press, Durham (2005)
5. Swartz, J.: Chinese hackers seek U.S. access. USA Today Education, 12 March 2007. http://www.usatodayeducate.com/wp-content/uploads/chinese.pdf
6. Kremer, J.-F., Müller, B. (eds.): Cyberspace and International Relations: Theory, Prospects, and Challenges. Springer-Verlag, Berlin (2014). https://doi.org/10.1007/978-3-642-37481-4
7. McGuinness, D.: How a cyber attack transformed Estonia. BBC News, 17 April 2017. https://www.bbc.com/news/39655415
8. Ludlow, P.: Crypto anarchy, Cyberstates, and Pirate Utopias. MIT Press, Boston (2001)
9. Geneva Internet Form, UN GGE and OEWG. https://dig.watch/processes/un-gge
10. Eom, J.-H.: Roles and responsibilities of cyber intelligence for cyber operations in cyberspace. Int. J. Softw. Eng. Appl. **8**(9), 137–146 (2014)
11. Eade, D., Williams, S.: The Oxfam Handbook of Development and Relief. Oxfam, Oxford (1995)
12. Echebarria, C., Barrutia, J.M., Aguado, I.: Local agenda 21: progress in Spain. Eur. Urban Reg. Stud. **11**(3), 273–281 (2004)
13. Valentin and Spangenberg (2000)
14. Rohozinski, R.: The New Reality of Cyberwar: Prospects and Challenges. Taylor & Francis, Oxford (2012)
15. Cybersecurity, Is cybersecurity a key component of our nation's homeland security?: US efforts to secure the information age. USA Today Education. http://www.usatodayeducate.com/wp-content/uploads/chinese.pdf

16. Eriksson and Giacomello (2006)
17. Giacomello and Eriksson (2007)
18. Latham (2003)
19. Bendrath (2001)
20. Bendrath (2003)
21. Eriksson (2001b)
22. Bendrath et al. (2007)
23. Rid, T.: Cyberwar will not take place. J. Strateg. Stud. **35**(1), 5–32 (2012)
24. Ryan, C.M.: Cyber Strategy: the Evolving Character of Power and Coercion. Oxford University Press, Oxford (2018)
25. Herrera, G.L.: Cyberspace and sovereignty: thoughts on physical space and digital space. In: Paper Presented at the 1st International CISS/ETH Conference on "The Information Relations and the Changing Face of International Relations and Security," Lucerne, Switzerland, 23–25 May 2005), p. 30 (2005)
26. Min, K.-S., Chai, S.-W., Han, M.: An international comparative study on cybersecurity strategy. Int. J. Secur. Appl. **9**(2), 13–20 (2015)
27. Schelling, p. 207 (1960)
28. van Evera, p. 6 (1998)
29. Hinkle, R.C.: Developments in American Sociological Theory, 1915-1950. SUNY Press, Albany (1994)
30. Bridgman, P.W.: The Logic of Modern Physics. Macmillan, New York (1927)
31. Hall, M.: Constructing historical realism: international relations as comparative history. Lund University (1999)
32. Parameswaran, P.: Japan-ASEAN cyber cooperation in the spotlight. The Diplomat, 24 February 2017. https://thediplomat.com/2017/02/japan-asean-cyber-cooperation-in-the-spotlight/
33. Tanakasempipat, P.: Southeast Asian cybersecurity center opens in Thailand. Reuters, 14 September 2018. https://www.reuters.com/article/us-asean-cyber/southeast-asian-cyber-security-center-opens-in-thailand-idUSKCN1LU1G0
34. Commonwealth Cyber Declaration. https://thecommonwealth.org/commonwealth-cyber-declaration. Accessed 31 Jan 2020
35. UK Government.: UK programme supporting cyber security in the Commonwealth: call for expressions of interest. https://www.gov.uk/government/publications/uk-programme-supporting-cyber-security-in-the-commonwealth-call-for-expressions-of-interest. Accessed 19 Feb 2020
36. Global Cyber Security Capacity Centre: Global Impact Knowledge and Policy Contributions from the First Five Years. https://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/GCSCC%20booklet%20WEB.pdf#search=%27UK++partnership+in+Africa%2C+Asia%2C+the+Pacific+and+the+Caribbean+cyber%27
37. Ayofe, A.N., Oluwaseyifunmitan, O.: Approach to solving cybercrime and cybersecurity. Int. J. Comput. Sci. Inf. Secur. **3**(1) (2009)
38. Schia, N.: Cybersecurity capacity building, digitalization, and the Global South. Eur. Cybersecur. J. **2**, 82–94 (2016)
39. Onuf, N.G.: World of our Making: Rules and Rule in Social Theory and International Relations. Routledge, New York (2013)
40. Plantera, F.: Estonia takes on a major role in cyber diplomacy with a new department for international cooperation. E-estonia (October 2019). https://e-estonia.com/estonia-cyber-diplomacy-international-cooperation/

41. Baltic News Service. Estonia supports developing cybersecurity in 4 countries of Africa, Asia. Leta., 2 April 2019. https://leta.lv/eng/defence_matters_eng/defence_matters_eng/news/4D29207F-14E5-4490-8973-F830B28E5C37/
42. United Nations Office for Disarmament Affairs. Group of Governmental Experts. https://www.un.org/disarmament/group-of-governmental-experts/
43. Jayawardane, S., Larik, J., Jackson, E.: Cyber governance: challenges, solutions, and lessons for effective global governance. The Hague Institution for Global Justice (2015). https://www.thehagueinstituteforglobaljustice.org/wp-content/uploads/2015/12/PB17-Cyber-Governance.pdf

An Analysis of Twitter During the 2017 Zimbabwean Military Intervention

Brett van Niekerk[1](✉), Martina Jennifer Zucule De Barros[2], and Trishana Ramluckan[1]

[1] University of KwaZulu-Natal, Durban, South Africa
{vanniekerkb,ramluckant}@ukzn.ac.za
[2] Technical University of Dresden (Technische Universitaet Dresden), Dresden, Germany
martina_jennifer_zucule_de.barrosl@mailbox.tu-dresden.de

Abstract. With the rapid expansion of social media, most notable events globally are recorded on one or more of the platforms. During the events of the Arab Spring uprisings, social media was seen to play a role in the protests. The role of social media increased in the Ukrainian anti-government protests of 2014 to 2015, and in the Fees Must Fall movement in South Africa in 2015 and 2016. In 2017, amongst growing pressure for Zimbabwean President Robert Mugabe to step down, the Zimbabwe Defence Forces intervened. The military intervention was followed by the citizen's staging marches calling for the president to step down. During these events, social media was used as one of the communication tools narrating the events. This paper conducts an analysis of Twitter during the military intervention and the protests, where the tweets were captured using the Followthehashtag tool during the protests. The focus is on four predominate hashtags, and an analysis of both the metadata as well as the message contents is provided. Qualitative technique such as word frequency and word clouds are used to illustrate the common themes being distributed on Twitter.

Keywords: Popular uprising · Social information warfare · Social media

1 Introduction

The role of social media in popular uprisings is increasing, for example Arab Spring in Africa and the Middle East, the #Occupy movement, the protests in Ukraine, and the #FeesMustFall movement in South Africa. In particular, social media has played a crucial role in generating consciousness and awareness of several protests around the world [15]. In Africa, social media has "been harnessed to make political demands on human rights, accountability and good governance" [16]. The Arab Spring popular anti-government protests across North Africa and the 2015 #FeesMustFall student protests is South Africa inspired similar movements in other regions of the continent such as Malawi, South Africa and Zimbabwe. In Zimbabwe, the use of social media for political demands and good governance became popular in 2013. However, in 2016, Zimbabwe was grip by recurrent citizen protests of a diverse nature against the

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 233–247, 2020.
https://doi.org/10.1007/978-3-030-62803-1_19

president's (Robert Mugabe) regime [15]. These protests led to the emergence of the #ThisFlag and the #ThisGown movements. In 2017, amidst growing calls for Zimbabwean president Robert Mugabe to step down and increasing tensions between the president and the former vice-president, the Zimbabwean Defence Force intervened.

The concept of social information warfare is the use of techniques to affect the information environment to influence or disrupt the decision making process of an adversary. The use of social media during physical protests can be used to gain recognition and influence international or local audiences to support the movement, or to aid in the coordination and provide support services to the protests [24].

This paper analyses the metadata and content of four Twitter hashtags during this period immediately following the military intervention. Descriptive analysis of the metadata and content analysis of the tweets is conducted.

Section 2 provides a background to information warfare, information-based conflict in Africa, Zimbabwe, and the Zimbabwean crisis in particular. Section 3 describes the methodology for the paper. The analysis is presented in Sect. 4, with the discussion in Sect. 5. The paper is concluded in Sect. 6.

2 Background

This section provides a background to Information Warfare (Sect. 2.1), with specific examples related to Africa (Sect. 2.2). A historical background Zimbabwe (Sect. 2.3) and an overview of the events leading up to and during the crisis (Sect. 2.4) are provided.

2.1 Information Warfare

Information warfare is an umbrella term for a number of functional areas that affect or operate in the information sphere, such as command and control, cyber operations, psychological operations, intelligence gathering and dissemination, and electronic warfare. These can operate in, be enabled by, or affect, the physical, virtual and psychological domains [25]. Whilst this is a primarily military concept, it can be extended to other spheres, including social, corporate, economic, and personal spheres [26, 27].

Social information warfare first showed promise by the Mexican Zapatista movement in 1994 to 1996 [28]. Radio broadcasts and soft-power provided the ability for governments to influence each other and populations; however, mobile phones with text messaging proved to enhance the ability for anti-government protests to coordinate, including in the Philippines, Iran and Africa [12, 29, 30], with specific examples in Africa being highlighted in Sect. 2.2. With the advent of social media, this became a major enabling technology for movements and protestors engaging in social information warfare, such as the 'Twitter Revolution' in Moldova, the Arab Spring movements, student protests in South America and South Africa, and protest and state-backed activity in Ukraine [12, 13, 24, 31]. Initially social media formed primarily an information dissemination platform for limited coordination and gaining broader recognition; however, in the #FeesMustFall protests and Ukraine, it was also used to organize and provide support services such as medical and legal aid [13, 24]. From these example, social information warfare can be seen to include the networking of individuals and groups to form a coordinated mass to promote a cause; however, it is alleged governments are using similar techniques for mass influence operations, and cyber-attacks have been seen to be used in conjunction with these techniques [32].

2.2 Information-Based Conflict in Africa

Information has long been a component in conflict in Africa, be it ideological, the use of military intelligence and deception, or other means. Even though information technologies do not necessarily exhibit the high penetration of other regions, they still have been seen to play a role in these conflicts. Examples include the use of radio broadcasts to incite genocide in Rwanda; in the 2007 Kenyan elections, the radio broadcasts were supplemented with text messages to incite violence. In 2010, Mozambique also experienced protests with text messages being used for information distribution. Social media is reported to have played a role in the Arab Spring protests across Africa and the Middle East [12]. In South Africa, a movement for free higher education, #FeesMustFall, extensively used social media for information dissemination and coordination of protests in 2015 [13].

In Zimbabwe, there have been reports of government surveillance and cyber-censorship of communications as well as anti-government movements defacing websites. The websites of newspapers were also targeted by distributed denial of service attacks. In 2013 a number of websites were attacked by the hacktivist group Anonymous Africa in protest against perceived human rights abuses [12].

2.3 Background to Zimbabwe

Prior to the colonial period of Africa, it is estimated that the first Bantu-speaking farmers settled in 150 BCE, which was followed by migrations of Zhizo, Shona-speaking, and Ndebele peoples. There were a series of kingdoms prior to British control being achieved in 1985 as the colony of Rhodesia. Independence was declared without British consent, followed by a period of sanctions and increasing resistance to minority rule and eventual independence. In February 1980 ZANU won the elections and Robert Mugabe became prime minister and eventually president up until his resignation in 2017. During this time, there was conflict with opposition parties [1].

During the time of the crisis, Zimbabwe was reported to have a population of 14,236,745, a GDP of 22,813,010,116 US$, and a "Poverty headcount ratio at national poverty lines" of 70% of the population [2]. In 2017 Zimbabwe was reported to have a fixed telephone line subscriptions of 1.86 subscriptions per 100 inhabitants, fixed broadband subscriptions of 1.32 per 100 inhabitants, 98.99 mobile telephone subscriptions per 100 inhabitants, 27.1% of inhabitants had Internet access [3], and there were 435 secure internet servers [2]. The 2017 Information Society Report published by the International Telecommunication Union (ITU) analyzes the ICT development index (IDI) of worldwide countries. From this report Zimbabwe was ranked 136 out of 176 countries with an IDI level of 2.92. Regionally, Zimbabwe ranked 12 out of 38 countries [17]. In the 2017 World Press Freedom Index, Zimbabwe was ranked 128 out of 180 countries. Since the new government, this has improved slightly to 127 of 180 in 2019 [4]. Given concerns around press freedom and the prevalence of mobile communication over fixed-line, the use of social media to distribute anti-government messages, despite there being reports of government attempts of cyber-censorship [4]. In 2016, during the Harare protests such as #ThisFlag, #Thisgown and #ZimShutDown, the government deployed several tactics to curtail the social media uprising. Several people were arrested, and social networks such as Facebook, Twitter and WhatsApp were blocked. In addition, the government started to formulate the Cyber Security Act to monitor and control online activism [15].

2.4 Brief Timeline of Events

In October 2017 there was growing tension between the first lady, Grace Mugabe, and the vice-president, Emmerson Mnangagwa over the succession to the President, which lead Mnangagwa fleeing to Mozambique and South Africa after his sacking in November. Mnangagwa had the support of some senior generals; the army chief General Chiwenga was overseas at the time and was warned of his impending arrest when returning to Zimbabwe [5–8].

On the 13 November Chiwenga, called a press conference to denounce the in-fighting and indicated the military will intervene. On the 14 November military convoys entered the capital Harare and seized the headquarters of the Zimbabwe Broadcasting Corporation (ZBC) and began raiding the homes of ministers aligned to Robert and Grace Mugabe, which continued into the morning of the 15th. A general made an announcement indicating that the military action was not a coup; this announcement was played in conjunction with normal programming on the 15th by the ZBC. On the 18th protests were held calling for Robert Mugabe to resign; on the 19th the ruling ZANU-PF party remove Robert Mugabe as their head and expelled Grace Mugabe and her supporters. Mugabe still refused to resign until the evening of the 21st, amidst impeachment proceedings being prepared [8–11].

3 Methodology

Tweet metadata and content were captured using the online tool Followthehashtag.com, broadly following the methodology of [13], which focused on the #FeesMustFall movement in South Africa. Four hashtags were focused on: #zimbabwe, #zimbabwecoup, #freshstart, and #mugabemustgo. Data collection began on the 15th November 2017 (the day after the military intervention) until the 24th November 2017. There are limitations on the Twitter API, where a maximum number of 1500 tweets can be extracted, however multiple queries were made to maximize the sample. According to Saunders, Lewis, and Thornhill, a sample of 384 is required for a population of 10 million at a 5% margin of error [14]. This indicates the number of tweets extracted is a sufficient sample for a 5% margin of error.

Descriptive analysis on the tweet metadata for the four hashtags was conducted, to illustrate the strength, gender demographics and potential reach of each hashtag.

Geolocation and trend analysis was conducted. Content analysis of the tweets was conducted using word frequency, visualized as a word cloud.

4 Analysis

This section provides the analysis of the tweets captured based on the methodology described in Sect. 3. An analysis of the metadata from the tweets provides information on the trends, demographics, and potential reach or influence of the tweets; this is presented in Sect. 4.1. Analysis of the message content is provided in Sect. 4.2.

4.1 Twitter Metadata Analysis

Table 1 presents a summary of the four hashtags assessed. The #zimbabwe hashtag is a general hashtag and was active previously. The #zimbabwecoup hashtag was from the time of the military action, whereas the #freshstart and #mugabemustgo hashtags emerged during the time of the protests. Whilst the two hashtags that were active for longer had an opportunity to gain more tweets, it can be seen that the #mugabemustgo hashtag had nearly double the number of tweets compared to the #freshstart hashtag. For the most part, the tweets were original and retweets, with limited replies. This indicated one-way dissemination of information, rather than dialogue. There are also a number of images and links, again indicating the dissemination of information.

Table 2 illustrates the demographics of the users tweeting. Gender detections ranged from 35% to 41%, however given the population size of the users this is adequate to generalize the gender proportions of the tweets. For all hashtags, there are more male users than female. The strongest female representation is for the #mugabemustgo hashtag, accounting for 44% of the users. These statistics imply male dominance, however a more unified resistance against the Robert Mugabe.

Table 3 presents the reach and possible influence of the various hashtags. The two longer running hashtags had a significantly greater potential audience. The #mugabemustgo hashtag had almost double the audience of #freshstart. These metrics also hold true for the potential impressions; however, the #zimbabwecoup has the lowest impressions per audience member, implying this was seen as more information dissemination. The #freshstart hashtag had the greatest ration of impressions to audience, implying a more emotional topic. The tweets per contributor appears fairly consistent at approximately 1.4; however, #freshstart again had a higher ratio than the other hashtags. The rate of the tweets appear consistent, except for #freshstart, where the rate is half that of #mugabemustgo. Given the metrics presented, #freshstart probably has the greatest dialogue, exhibiting more impressions and tweets amongst a smaller number of contributors compared to the other hashtags. The other hashtags, with fewer impressions per audience and fewer tweets per contributor can be considered to be more orientated for dissemination.

Table 1. Summary of the data from the four hashtags.

	#zimbabwe	#zimbabwecoup	#freshstart	#mugabemustgo
Measured from	2017-11-15	2017-11-15	2017-11-18	2017-11-18
Measured to	2017-11-24	2017-11-24	2017-11-24	2017-11-24
Total tweets	18323	13786	4321	8367
Total audience	147198145	42629048	8295909	17006991
Contributors	12905	9759	2727	6124
Original tweets	4794	5376	1800	2574
Replies	464	409	165	225
Retweets	13065	8001	2356	5568
Images and links	4199	4328	1165	1725

Table 2. Demographic information for tweets

	#zimbabwe	#zimbabwecoup	#freshstart	#mugabemustgo
Total users	17300	13786	4321	8367
Total gender detections	6338	4918	1612	3458
% of detections	36.64%	35.67%	37.31%	41.33%
Total male	4323	3462	967	1934
Total female	2015	1456	645	1524
% male	69%	71%	60%	56%
% female	31%	29%	40%	44%

Table 3. Reach of the hashtags

	#zimbabwe	#zimbabwecoup	#freshstart	#mugabemustgo
Total audience	137 793 162	42 629 048	8 295 909	17 006 991
Contributors	12330	9759	2727	6124
Total tweets	17300	13786	4321	8367
Total potential impressions	300 603 889	70 293 213	19 106 621	32 708 443
Measured from	2017-11-15	2017-11-15	2017-11-18	2017-11-18
Measured to	2017-11-24	2017-11-24	2017-11-24	2017-11-24
Tweets per contributor	1.40	1.41	1.58	1.37
Impressions/audience	2.18	1.65	2.30	1.92
Tweets per second	0.022430188	0.017704684	0.008282268	0.016185
Tweets per minute	1.34581126	1.062281035	0.496936078	0.971074
Tweets per hour	80.74867558	63.73686211	29.8161647	58.26444
Tweets per day	1937.968214	1529.684691	715.5879529	1398.347

Figures 1 and 2 present the trends for the number of tweets and potential impressions. In both cases, the trends are as expected, where they decline and peak during periods of increased activity in the physical world, such as the protests or the resignation. Despite remaining almost constant in terms of the number of tweets, major spikes can be seen correlating with these events for the #zimbabwe hashtag.

Figures 3, 4, 5 and 6 use PowerMaps in Microsoft Excel to visualize the geolocation data from the tweets. As is evident, the #zimbabwe and #zimbabwecoup hashtags were more prevalent in South African than in Zimbabwe. This is possibly due to the number of Zimbabwean ex-patriots living in South Africa, as well as the interest in political affairs between the neighboring countries. There could also be fear due to perceived oppression which kept the tweets internal to Zimbabwe low; once the political situation had swayed away from President Mugabe, the internal population may have felt more confident in voicing their opinion during the protests. As is state in Sect. 1, there were attempts to bock and retaliate against online activism, which may have initially prevented internal tweets until the population was confident that Robert Mugabe has limited control over the country and the security forces.

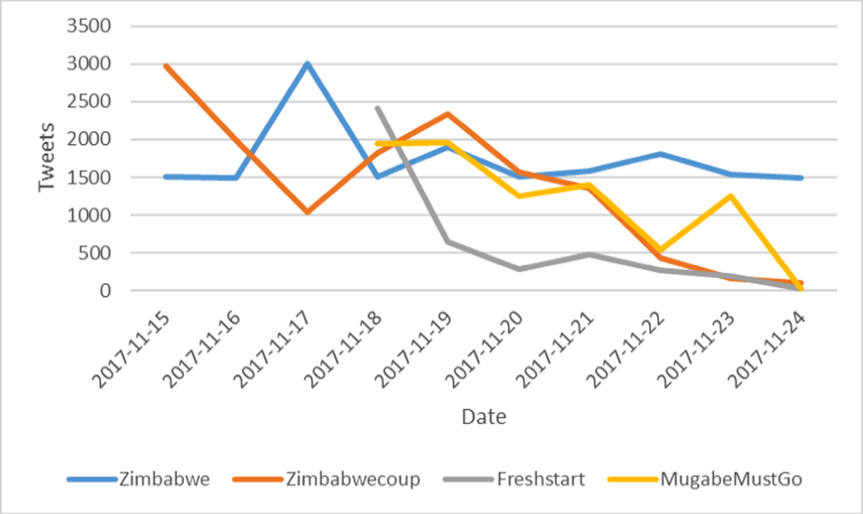


Fig. 1. Trends in the number of tweets

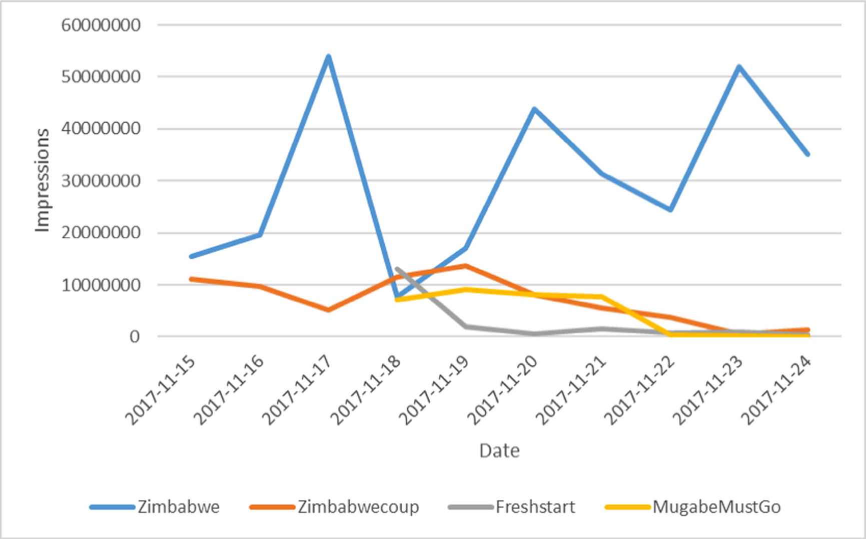


Fig. 2. Trends in potential impressions

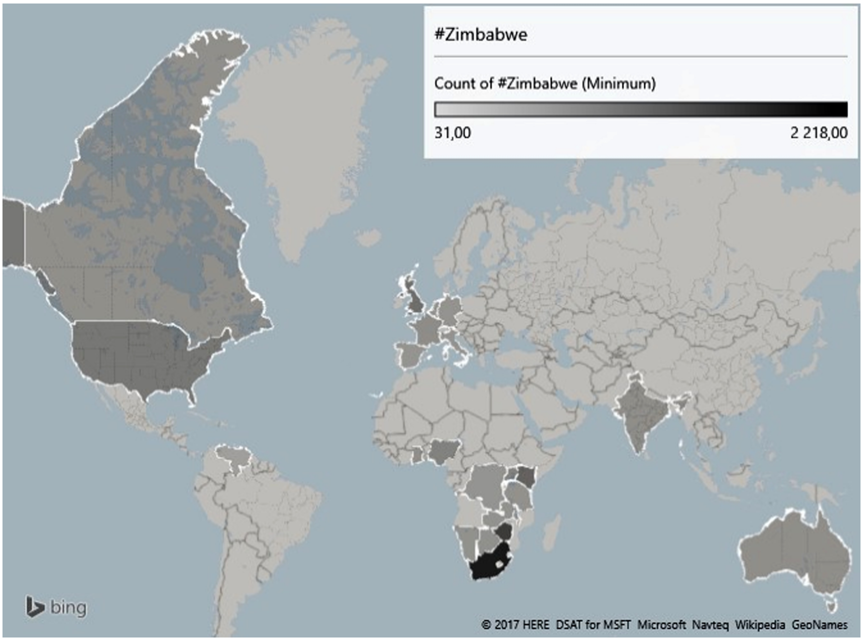

#Zimbabwe
Count of #Zimbabwe (Minimum)
31,00
2 218,00
bing
© 2017 HERE DSAT for MSFT Microsoft Navteq Wikipedia GeoNames

Fig. 3. Geolocation of the tweet origin for #Zimbabwe

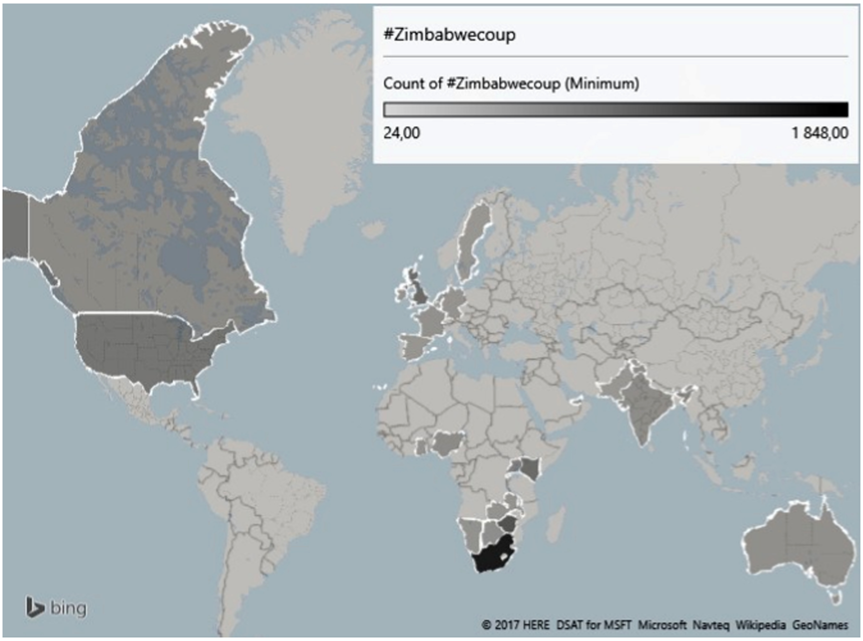

#Zimbabwecoup
Count of #Zimbabwecoup (Minimum)
24,00
1 848,00
bing
© 2017 HERE DSAT for MSFT Microsoft Navteq Wikipedia GeoNames

Fig. 4. Geolocation of the tweet origin for #Zimbabwecoup

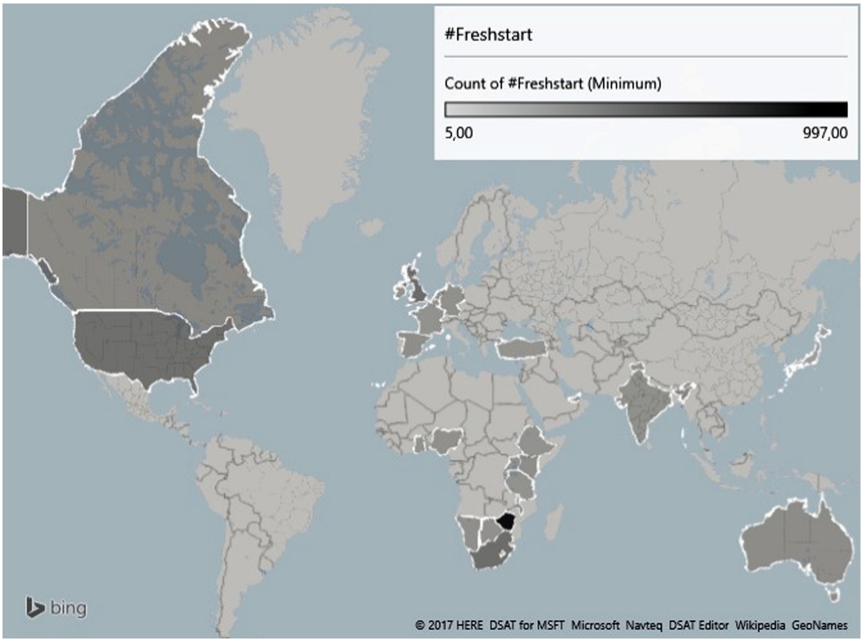

#Freshstart
Count of #Freshstart (Minimum)
5,00
997,00
bing
© 2017 HERE DSAT for MSFT Microsoft Navteq DSAT Editor Wikipedia GeoNames

Fig. 5. Geolocation of the tweet origin for #Freshstart

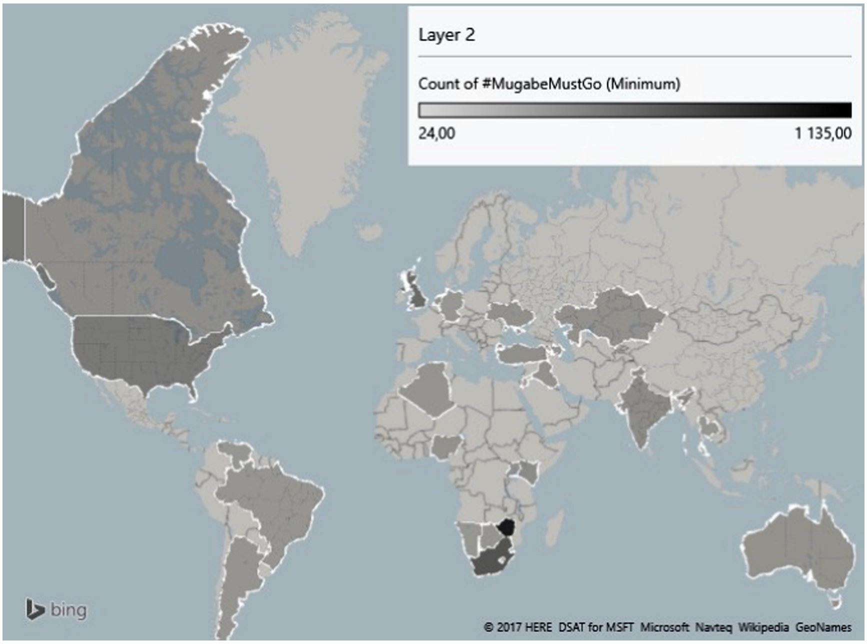

Layer 2
Count of #MugabeMustGo (Minimum)
24,00
1 135,00
bing
© 2017 HERE DSAT for MSFT Microsoft Navteq Wikipedia GeoNames

Fig. 6. Geolocation of the tweet origin for #MugabeMustGo

4.2 Twitter Message Analysis

This section presents more detail on the content of the tweets. A word cloud (Fig. 7) is used to visualize the word content, where the larger words are more prevalent. The most prevalent, as is expected, is the #zimbabwe hashtag. This is followed by the #Mugabe hashtag, illustrating the prevalence of the family in relation to the protests (either in opposition or support of them).

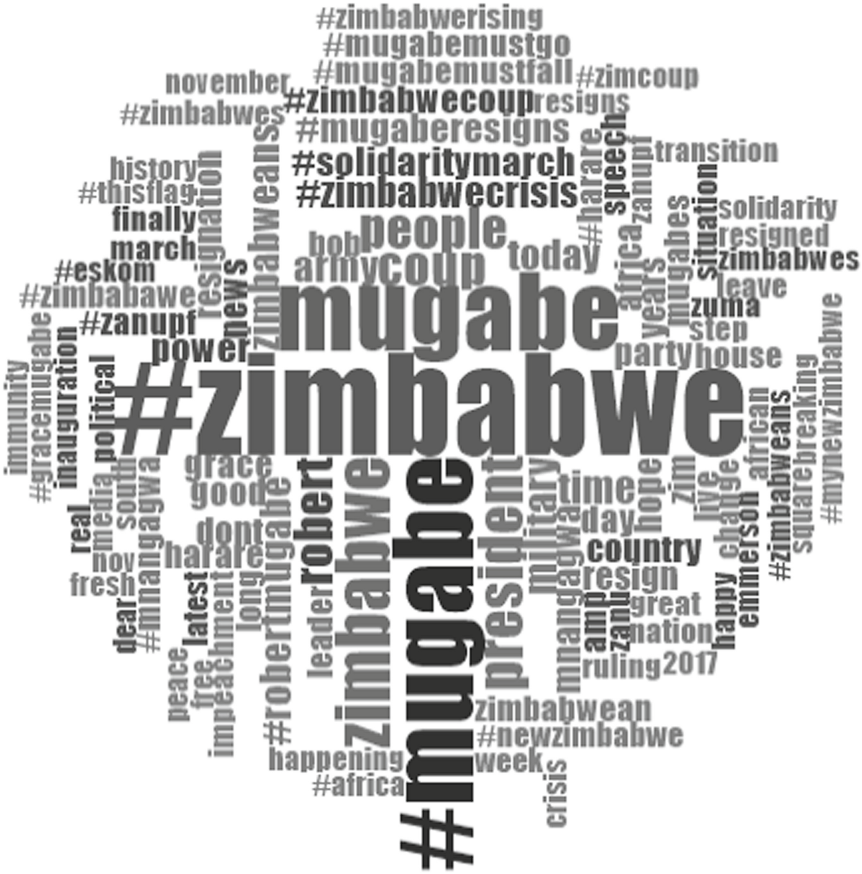
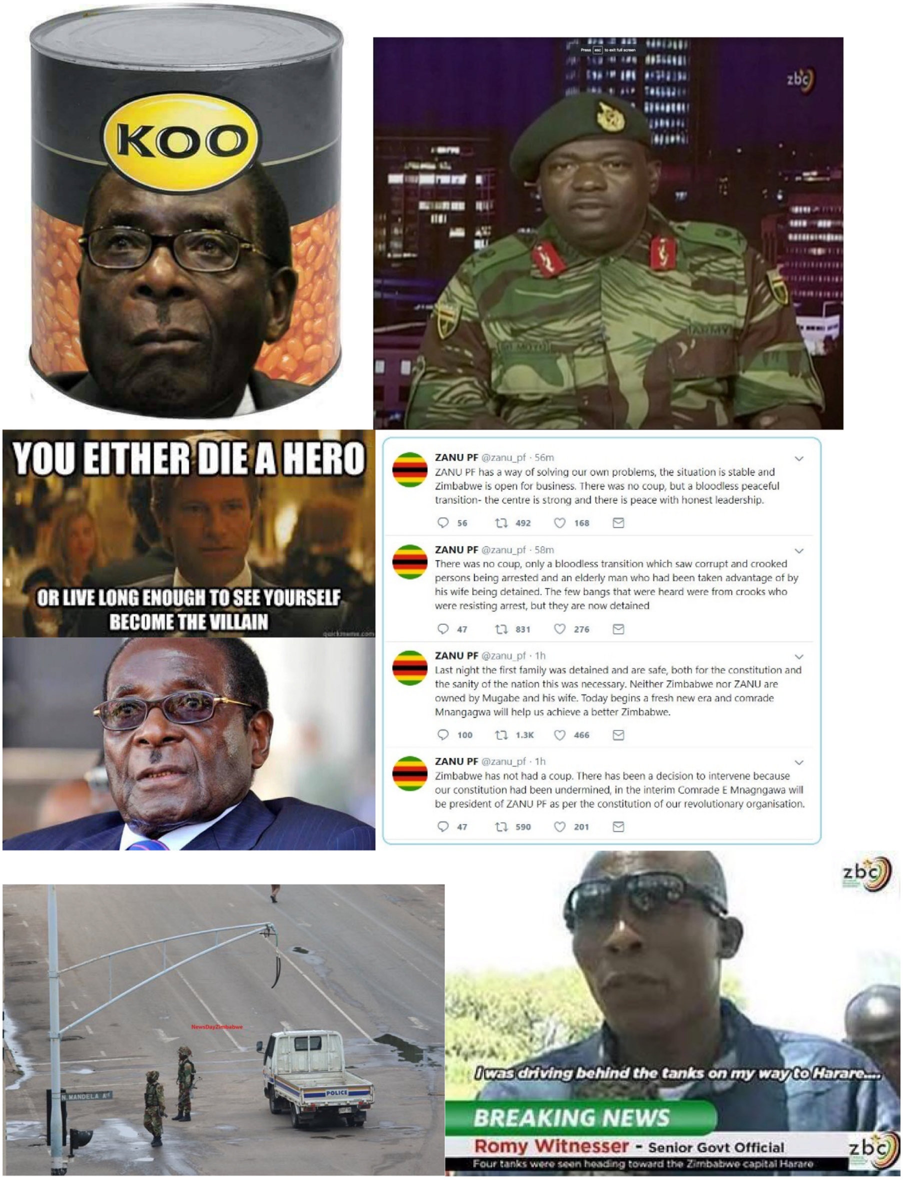


Fig. 7. Word cloud for tweet content

Images were also sent via Twitter; examples of these are provided in Fig. 8. Some appear light-hearted regarding the situation at the time, such as the image of a ‘KOO’ brand, playing on the word ‘coup’ [18], and a phrase from a movie [19]. Other images are clearly meant for the dissemination of information, including images of TV News broadcasts (an image of a general speaking on TV [20], and another image of a news report [21]) and the governing party’s official Twitter Feed [22]. Images also can document what the users can see, such as the troops in the street [23].


KOO
zbc
YOU EITHER DIE A HERO
OR LIVE LONG ENOUGH TO SEE YOURSELF BECOME THE VILLAIN
NewsDayZimbabwe
POLICE
N. MANDELA AV
zbc
I was driving behind the tanks on my way to Harare....
BREAKING NEWS
Romy Witnesser - Senior Govt Official
Four tanks were seen heading toward the Zimbabwe capital Harare
zbc

ZANU PF @zanu_pf · 56m
ZANU PF has a way of solving our own problems, the situation is stable and Zimbabwe is open for business. There was no coup, but a bloodless peaceful transition- the centre is strong and there is peace with honest leadership.
56 492 168

ZANU PF @zanu_pf · 58m
There was no coup, only a bloodless transition which saw corrupt and crooked persons being arrested and an elderly man who had been taken advantage of by his wife being detained. The few bangs that were heard were from crooks who were resisting arrest, but they are now detained
47 831 276

ZANU PF @zanu_pf · 1h
Last night the first family was detained and are safe, both for the constitution and the sanity of the nation this was necessary. Neither Zimbabwe nor ZANU are owned by Mugabe and his wife. Today begins a fresh new era and comrade Mnangagwa will help us achieve a better Zimbabwe.
100 1.3K 466

ZANU PF @zanu_pf · 1h
Zimbabwe has not had a coup. There has been a decision to intervene because our constitution had been undermined, in the interim Comrade E Mnagngawa will be president of ZANU PF as per the constitution of our revolutionary organisation.
47 590 201

Fig. 8. Images distributed by Twitter [18–23]

5 Discussion

The four hashtags investigated exhibit characteristics that indicate that they were primarily used for information dissemination rather than dialogue, except for the #freshstart hashtag which appears to have a higher extent of dialogue. This shows a lesser utilization of social media compared to other protests, such as the Maidan protests in Ukraine and the #feesmustfall protests in South Africa [13], where social media was also used to organize support in terms of legal and medical assistance; in Ukraine the social media was more akin to a command and control network [24]. However, the protests in Ukraine met strong government opposition [24], as did the #feesmustfall protests [13]. The protests in Zimbabwe following the military intervention was supporting the objectives of the intervention (the removal of Robert and Grace Mugabe from government), therefore there was not the need for command and control style communications. Given the military's commandeering of the national broadcasting and replaying of their own message, the initial strong geolocation data from South Africa indicates that Zimbabwean expatriates were disseminating information into Zimbabwe to keep the internal population informed of the broader global discussion compared to the 'one-sided' via available inside the country.

The use of social media to document the view of 'ordinary' citizens potentially provides for a more accurate view of history. However, this provides a challenge for those intending to cover up certain events. It also allows for real-time tracking of military or security force movements; this makes it more difficult to conduct a 'surprise attack', as also seen in the Ukraine [24]. This is again exhibited to a certain extent in the images distributed via Twitter during the military intervention in Zimbabwe.

Despite Zimbabwe's relatively low Internet penetration, the potential for social media as a tool for democracy and crisis communication was demonstrated during the military intervention. The previous attempts by the Zimbabwean government to curtail the use of social media indicates the concerns governments have over its use and potential to control a narrative, thereby transferring the soft power to the citizens and away from governments. The aspects of social information warfare were present in the social media use during the crisis, where information dissemination and political views and commentary were being distributed.

6 Conclusion

Growing dissatisfaction and political infighting led the Zimbabwean military to intervene and remove President Robert Mugabe from power in November 2017. As with most major modern events, there was commentary around the situation provided on social media. This paper provides an analysis of the Twitter metadata and content during the period of the military intervention to better understand how the virtual world reflected the physical occurrences. The Twitter use during this period was more for mass dissemination of information, potentially from outside the country into the country. The usage did not exhibit the command and control for protests as seen in other global uprisings; however, in the Zimbabwe case there was not necessarily the need for further usage. This case does illustrate the possible strength of social media to support democratic processes in terms of freedom of information. The transmission of images does however raise the security concern that movements of security forces or military can be tracked by most individuals with a basic cell phone.

References

1. South African History Online, Zimbabwe. https://www.sahistory.org.za/place/zimbabwe. Accessed 20 Jan 2020
2. World Bank: World Development Indicators. http://datatopics.worldbank.org/world-development-indicators/. Accessed 20 Jan 2020
3. International Telecommunications Union, Statistics. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx. Accessed 20 Jan 2020
4. Reporters Without Borders, Zimbabwe. https://rsf.org/en/zimbabwe. Accessed 20 Jan 2020
5. BBC.com: Zimbabwe succession row: Grace Mugabe warns of coup plot. https://www.bbc.com/news/world-africa-41530924. Accessed 21 Jan 2020
6. Kumbuka, D., Marawanyika, G., Latham, B.: Zimbabwe's ousted vice president flees after death threats. Bloomberg. https://www.bloomberg.com/news/articles/2017-11-08/mugabe-guts-zimbabwe-security-state-with-deputy-s-dismissal. Accessed 21 Jan 2020
7. Pindula News: ZRP Support Unit Attempted To Arrest General Chiwenga When He Returned From China. https://news.pindula.co.zw/2017/11/17/zrp-support-unit-attempted-to-arrest-general-chiwenga-when-he-returned-from-china/. Accessed 21 Jan 2020
8. Burke, J.: Military urges calm in Zimbabwe after it seizes key sites in capital. The Guardian https://www.theguardian.com/world/2017/nov/14/tensions-rise-in-zimbabwe-as-military-drives-through-outskirts-of-capital. Accessed 21 Jan 2020
9. BBC.com: Zimbabwe crowds rejoice as they demand end to Mugabe rule. https://www.bbc.com/news/world-africa-42035981. Accessed 21 Jan 2020
10. Moyo, J.: Robert Mugabe, in Speech to Zimbabwe, Refuses to Say if He Will Resign. The New York Times. https://www.nytimes.com/2017/11/19/world/africa/zimbabwe-robert-mugabe.html. Accessed 21 Jan 2020
11. BBC.com: Zimbabwe's Robert Mugabe resigns, ending 37-year rule. https://www.bbc.com/news/world-africa-42071488. Accessed 21 Jan 2020
12. Van Niekerk, B., Maharaj, M.: Information-based conflict in Africa. Sci. Militaria S. Afr. J. Mil. Stud. **41**(2), 24–41 (2013)
13. Ramluckan, T., Ally, S.E.S., van Niekerk, B.: Twitter use in student protests: the case of South Africa's #FeesMustFall campaign. In: Korstanje, M. (ed.) Threat Mitigation and Detection of Cyber Warfare and Terrorism Activities, pp. 220–253 (2017)
14. Saunders, M., Lewis, P., Thornhill, A.: Research Methods for Business Students, 4th edn. Pearson Education, Harlow (2007)
15. Mukuruva, B.: Social media as alternative political voice: a history of Harare 2016 mass protests. Honours dissertation, University of Venda, South Africa (2018)
16. Gukurume, S.: #ThisFlag and #ThisGown cyber protests in Zimbabwe: reclaiming political space. Afr. Journal. Stud. **38**(2), 49–70 (2017)
17. International Telecommunication Union (ITU): Measuring the information society report (2017). https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2017/MISR2017_Volume1.pdf. Accessed 17 Feb 2020
18. Twitter image. http://pbs.twimg.com/media/DOqiAoGXcAAjXtz.jpg. Accessed 18 Feb 2020
19. Twitter image, http://pbs.twimg.com/media/DOqbITBX4AAryPo.jpg. Accessed 18 Feb 2020
20. Twitter image. http://pbs.twimg.com/media/DOqFFdiXkAEmOM2.jpg. Accessed 18 Feb 2020
21. Twitter image. http://pbs.twimg.com/media/DOpzQjJW4AAA-2M.jpg. Accessed 18 Feb 2020

22. Twitter image. http://pbs.twimg.com/media/DOqhCtBWkAAzDCC.jpg. Accessed 18 Feb 2020
23. Twitter image. http://pbs.twimg.com/media/DOqY99pXkAAagXA.jpg. Accessed 18 Feb 2020
24. Van Niekerk, B.: Information warfare in the 2013–2014 Ukraine crisis. In: Richet, J. (ed.) Cybersecurity Policies and Strategies for Cyberwarfare Prevention, pp. 307–339. IGI, Hershey (2015)
25. Brazzoli, M.S.: Future prospects of information warfare and particularly psychological operations. In: le Roux, L. (ed.) South African Army Vision 2020, pp. 217–232. Institute for Security Studies (2017)
26. Cronin, B., Crawford, H.: Information warfare: its application in military and civilian contexts. Inf. Soc. **15**(4), 257–263 (1999)
27. Schwartau, W.: Information Warfare: Chaos on the Information Superhighway, 2nd edn. Thunder's Mouth Press, Emeryville (1996)
28. Ronfeldt, D., Arquilla, J., Fuller, G., Fuller, M.: The Zapatista "Social Netwar" in Mexico. RAND Corporation, Santa Monica (1998)
29. Faris, R., Heacock, R.: Cracking Diwn on digital communication and political organizing in Iran. OpenNet Initiative (2009). http://opennet.net/blog/2009/06/cracking-down-digital-communication-and-political-organizing-iran. Accessed 09 July 2009
30. Rigby, B.: Mobilising Generation 20: Technologies to Recruit. Organise and Engage Youth. Jossey-Bass, San Francisco (2008)
31. Hodge, N.: Inside Moldova's Twitter Revolution. Wired (2009). http://www.wired.com/dangerroom/2009/04/jailhouse-tech-sniffs-out-cell-phones/. Accessed 13 Aug 2009
32. van Niekerk, B.: Information warfare as a continuation of politics: an analysis of cyber incidents. In: 2018 Conference on Information Communications Technology and Society (ICTAS), pp. 1–6 (2018)

Cybersecurity and Cybercrime Combatting Culture for African Police Services

Louise Leenen[1](✉), Joey Jansen van Vuuren[2](✉), and Anna-Marie Jansen van Vuuren[2](✉)

[1] University of the Western Cape and CAIR, Cape Town, South Africa
lleenen@uwc.ac.za
[2] Tshwane University of Technology, Pretoria, South Africa
{jansenvanvuurenjc,Jansenvanvuurenal}@tut.ac.za

Abstract. Police forces are responsible to investigate cybercrimes and to protect their own assets from cybersecurity attacks. The majority of police forces find it difficult to fulfil their responsibilities in this regard in the face of constrained funding, a lack of awareness and training amongst law enforcement staff, the growing number of cybercrime incidences, and outdated or insufficient technology and infrastructure. Even if police forces are able to install technical controls to counter cyber threats, their staff members' cyber behaviour may be a weak link in the cybersecurity chain and will probably not have sufficient training. The cultivation of a cybersecurity culture has been shown to be the best approach to address human behaviour in the cyber domain. There are several frameworks and other resources available for an organisation to cultivate a cybersecurity culture but the organisational culture in law enforcement agencies is different than that in other organisations. The cyber behaviour, cybercrime investigation skills, training and education of police force members require customised strategies and research. African police forces find it particularly difficult to deal with these challenges due to a lack of funding and a shortage of cybersecurity capability and capacity. This paper presents guidelines for African police forces to formulate strategies and plans to train and educate their members and to foster an organisational cybersecurity and cybercrime combatting culture.

Keywords: Law enforcement · Cybersecurity culture · Cybercrime combatting culture · Police

1 Introduction

Most organisations have become aware of the risks and threats that accompany connectivity and digitization, and are implementing technical measures and policies to guide employees on ensuring cybersecure environments. One of the most important measures is cybersecurity awareness training. Influencing the cyber behavior of a workforce contributes to the establishment of a cybersecurity culture which will enhance cybersecurity in an organisation [1].

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 248–261, 2020.
https://doi.org/10.1007/978-3-030-62803-1_20

Herskovits [2] defined culture as the collective and shared sense of relatedness of the human experience. There are four categories of culture: macro-cultures, organisational cultures, sub-cultures and micro-cultures [3]. These categories consist mainly of three levels that include espoused beliefs and values, visible artifacts, and basic underlying assumptions.

Cyber professionals refer to the macro-culture in their environment as a “Cyber culture”. ENISA (European Union Agency for Network and Information Security) describes Cybersecurity Culture (CSC) as the manifestation of people’s behavior with information technologies due to their knowledge, beliefs, perceptions, attitudes, assumptions, norms and values regarding cybersecurity [4]. CSC does not only include cybersecurity awareness and information security frameworks, but is also concerned with making cybersecurity an integral part of an employee’s job, habits and conduct. Gcaza, von Solms and Jansen van Vuuren [5] defined CSC as the aim to “instill a certain way to ‘naturally behave’ in daily life, a way that subscribes to certain [cybersecurity] assumptions”. In addition, a CSC also encapsulates socio-cultural measures that includes technical security methods, to ensure that cyber actions become a natural aspect of the daily activity [6].

People are the weakest link in the cybersecurity chain and are responsible for the majority of data breaches [4, 7]. Although various technologies contribute to security, research shows that cybersecurity is mostly influenced by the workforce in companies [6]. It takes time to cultivate a cybersecurity culture but it is crucial to influence and alter the behaviour of users over time to use technology securely [8, 9].

This paper applies existing research results on the cultivation of an organisational cybersecurity culture and results on cybersecurity and fighting cybercrime in police forces to propose a framework aimed at cultivating an African Police Cybersecurity and Cybercrime combatting culture (CCCC), which includes the training and educating of police force members to support cybercrime investigations. Section 2 gives an overview of literature on the cultivation of such an organizational culture. Section 3 discusses the structural requirements in a police force for investigating cybercrime, while Sect. 4 contains guidelines on how to capacitate those structures. In Sect. 5, the previous sections are combined to produce a step-by-step guide for the cultivation of an appropriate cybersecurity culture in a police force.

2 Cultivation of an Organisational Cybersecurity Culture

Although cybersecurity culture is regarded as an important subject, there are limited published research studies available on the topic [5, 10]. The publications predominantly argue that most important part of such a culture is cybersecurity awareness and education [10, 11]. A Norwegian study on their national cybersecurity culture found that businesses coined the term (“CSC”) but that not all businesses nor the average citizen were aware of the term [12]. It also noted that Norwegian citizens will willingly accept state monitoring, but do not trust the support from the police when they fall victim to cybercrime.

The Council on Digital Security Risk Management for Economic and Social Prosperity (OECD) recommended the promotion of a cybersecurity culture in order to protect information systems and networks [13]. This culture should include raising awareness of the threats to information systems and networks, as well as the availability of policies, practices, measures and procedures that should be adopted to address those risks. Additionally, law enforcement agencies must promote co-operation and information sharing, and focus on ethical issues in the development of standards. The Council's guidelines for cybersecurity culture development include:

- Raising awareness of processes that could reduce the internal and external cybersecurity risks to information systems and networks, including the potential threats arising from interconnectivity and interdependency.
- Developers, designers and suppliers of products and services must distribute appropriate information about the security functionality of products and services and their responsibilities related to security.
- Ethical conduct is crucial and security consistent with democratic values should be implemented.
- Risk assessment should be conducted to identify threats and vulnerabilities to guide the appropriate selection of controls for cybersecurity risk management.
- Security design and management should be based on risk assessment incorporating all products, services, systems and networks.

ENISA's guide for the promotion of Cybersecurity Culture programs within organisations includes a comprehensive CSC program with best practices. It recognises that organisational behavior is based on shared beliefs, values and actions amongst employees, including the employees' attitude towards cybersecurity. Unfortunately, cybersecurity awareness campaigns have proved to be insufficient. Thus, technical cybersecurity measures must be incorporated with other corporate processes and co-exist with job performance [4].

The main recommendations of the ENISA guide are related to cyber threat awareness within a company. Motivations on why a CSC is needed must be presented to company management and can include threat statistics, evidence of internal cyber-attacks and associated costs and results of a either a pilot CSC intervention or successfully deployed CSC programs.

ENISA's CSC Implementation Framework prescribes the following steps:

- Choose a core group of senior staff from different departments. Provision must be made for resources to support the CSC program.
- Conduct a risk assessment of possible misalignments within the organisational culture and business processes.
- Engage employees to get their buy-in.
- Define the main goals of the programme. Some goals will be organisation-wide while others will only be relevant to a certain group.
- Determine the current cybersecurity status and do a gap analysis with the aimed situation.
- Execute activities.
- After the activities have been executed, results must be analysed and measured according to the cybersecurity status.
- A CSC program is a continuous process that needs constant revision.

CISCO's chief information security officer, Steve Martino, emphasised that employees should understand that they are crucial in protecting the company from cybersecurity threats. Martino advised to educate, test and hold employees accountable for cybersecurity without publicly shaming them, but to provide guidance on addressing problems [14]. Still, creating a metric to gauge the level of maturity of a cybersecurity culture, (national or organisational) is challenging [12].

3 Cybercrime Investigation Skills

This section considers the structures required for a police force to investigate cybercrime.

3.1 Cybersecurity and Cybercrime Combatting Skills Shortage

Although there are several dedicated training facilities providing cybercrime training for police forces in the US and in Europe, there is still a skills shortage of cybercrime investigators. A UK government report [15] stated that none of the police forces or regional organised crime units had cybercrime analysts dedicated to developing the understanding of cyber-dependent crime or to support specific cyber-dependent investigations. In addition, if such support was requested from other units it was not prioritised. Trevor Halstead of Cybrary sums up the critical situation: "We really screwed things up this time. Somehow, we are in a situation where the sector of technology with the greatest potential negative impact on our lives, businesses, governments, peace, safety and security happens to have a severe deficiency of qualified people to fill its jobs," [16]. The majority of cybersecurity job positions require a bachelor's degree or higher [17]. Most companies, according to a report of Intel Security in partnership with McAfee, indicated that they prefer at least a bachelor's degree in a relevant technical area to enter the cybersecurity field [18]. The Cybersecurity Skills Gap Analysis (CSGA) report, prepared by the Workforce Intelligence Network for South Michigan, indicates that the highest demand in the US are for cybersecurity analyst/specialists, cybersecurity engineer, auditors, network engineers/architects, and software developers [17, 19]. These high-value skills are in critically short supply [18].

The high demand for cybersecurity skills influences the availability of these skills for cybercrime officers in the police. Police forces also need skills on these higher levels to be able to do cybercrime investigations, cybercrime analysis and research. In the case of financial cybercrimes, it is also a requirement to have sufficient knowledge in finance. According to research conducted in Europe by one of the authors, some of the European police forces employ graduates but they rarely have the skills to do the high-level analytics necessary for cybercrime investigations. An African police force member with these advanced cybercrime skills is an exception. This is confirmed by analysis done by the e-Governance Academy in Estonia portrayed in the National Cyber Security Index (NCSI) for these countries [20]. Non-police force personnel are often used to support cybercrime investigation activities.

3.2 Cybercrime Combatting Positions in the Police

Cybercrime skills requirements vary for different police force positions. One of the authors conducted interviews with members of several police cybercrime units and the results were used to develop the skills requirements for an African police force. Figure 1 reflects a generic framework of the different job levels and skills necessary to combat cybercrime for the police.

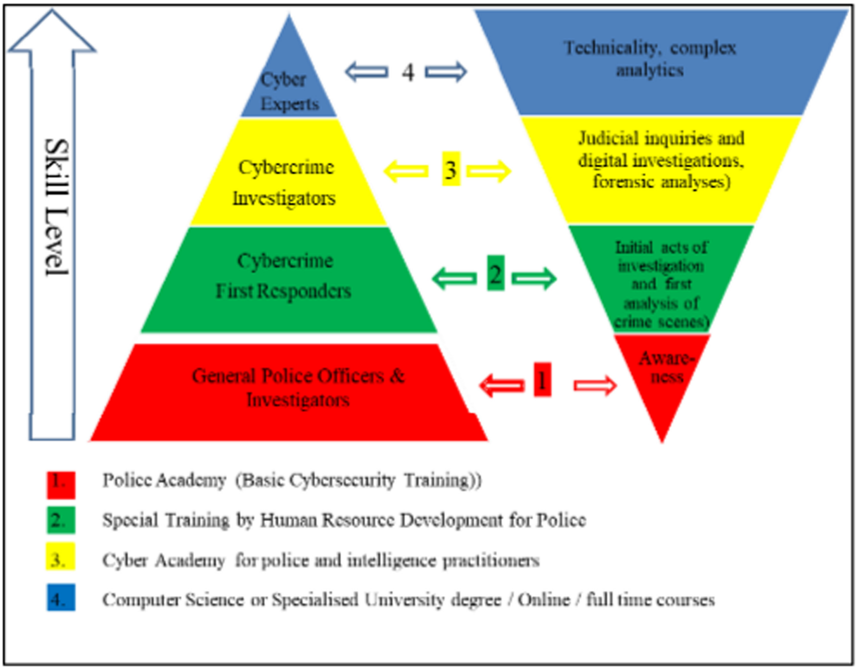

Skill Level
Cyber Experts
4
Technicality, complex analytics
Cybercrime Investigators
3
Judicial inquiries and digital investigations, forensic analyses)
Cybercrime First Responders
2
Initial acts of investigation and first analysis of crime scenes)
General Police Officers & Investigators
1
Aware-ness
1. Police Academy (Basic Cybersecurity Training))
2. Special Training by Human Resource Development for Police
3. Cyber Academy for police and intelligence practitioners
4. Computer Science or Specialised University degree / Online / full time courses

Fig. 1. Cybersecurity job levels and required skill for the police.

These levels differentiate between the cybercrime skills necessary for the general police officer in the local police station; general investigators (the lowest level); cybercrime first responders responsible to attend to a cyber-crime scene; cybercrime investigators that focus on the investigation of cybercrimes and the cybercrime experts responsible for the analytics, research and predictions of cybercrime. Most of the police cybercrime units will distinguish between the cybercrime investigators and the cybercrime experts. The next section addresses how to acquire these necessary skills and structures.

4 Cybersecurity and Cybercrime Combatting Capability and Capacity Building (CCCC)

The capacity and capability of the police force need to be enhanced to deal with all the technical aspects of cybercrime such as the examination of the digital evidence, the analysis of cybercrime scenes and forensic analysis [21]. Capacity building on different levels of the police force must be part of the cybercrime strategy and it must respond to the needs of police officers. It further produces immediate impact, favours multi-stakeholder cooperation and contributes to human development [21]. In addition, all cybercrime investigators will have to deal with electronic evidence and need comprehensive training in this regard [22].

Police officers have different training needs according to the required skills for their position. These job skills normally include generic investigation; cybercrime first responder and cybercrime scene investigation; internet crime investigation; covert internet crime investigation; network crime investigation; and digital forensic investigation and management) [23]. They can be categorised in four levels (Fig. 1):

- The general policeman in a local police station needs generic cybersecurity awareness training and cybercrime reporting skills.
- The cybercrime first responder needs the competency to secure electronic evidence and carry out computer forensics analyses for criminal proceedings.
- The cybercrime investigator needs to be able to conduct cybercrime, internet crime, covert internet crime, network crime, and digital forensic investigations and analyses.
- The cybercrime expert needs to be able to do advanced cybercrime analytics.

Different methodologies and platforms can be used to do the training. The UK Law enforcement agencies and partners make use of an online learning facility. In 2018, they established a specific area called the Cybercrime Hub intended to be "a one stop shop for all officers and staff with an interest in cybercrime, that will train from officers little knowledge to experienced cyber investigators" [15].

General Cybersecurity Awareness Training

Awareness training can be done online using different platforms, e.g. a virtual reality training program on mobile devices in the form of games. The establishment of a cyber academy responsible for cybercrime training in different African regions is recommended.

The functions of police officers will determine the appropriate level of training which must be scalable, standardised and replicable. It is possible to adapt existing police force training materials and initiatives [22] to include cybersecurity, cybercrime reporting and investigation procedures. There must also be cooperation between law enforcement, academia and industry. To ensure sustainability it is important to include the training of trainers.

First Responder Training

The goal of first responder training according to the FBI's special agent James McDonald, "is to improve a first responder's technical knowledge by focusing on best practices in terms of investigative methods specific for cyber investigations." He also emphasised that if first responders are trained there are less chance of errors being made while securing a crime scene involving digital evidence [24]. They need to have a working knowledge on how to secure electronic evidence as well as the physical evidence, while documenting the crime scene, and have knowledge of software, hardware, the Internet, social networks, encryption, legal tools, and other digital evidence. They also need enhanced practical skills regarding the methodology used in digital forensics for identifying and responding effectively at the scene of cybercrime cases [25].

Cybercrime Investigator Training
In addition to the normal investigator training, cybercrime investigator training has to include knowledge on digital media, networks and databases, operating systems, the use of forensic tools and the ability to do technical legal analysis of evidence, metadata, backup systems, electronic storage, applications, mobile and internet communication, the cloud and specific digital devices. They need to understand the attitudes required for cybercrime investigation, including the systematic management of cybercrime investigative resources and staff [26]. Some of the basic training can be done online. Practical training, e.g. forensics, must be done with the correct equipment in laboratories. A cybercrime investigation manual and discussion groups can support and guide cybercrime investigations [15]. A Cyber Academy for the training of intelligence officers and cybercrime investigators can also be established.

Cybercrime Experts
Tertiary degrees need to be developed where students can incorporate knowledge from criminology and computer science; they need to develop critical thinking and investigative skills, sophisticated technological understanding and advanced technical computer science skills. Studies must include advanced networking, databases, software and operating systems skills as well as all the platforms used to protect systems e.g. firewalls and intrusion detection systems. It must also include guidance on how cyber criminals could extract financial data, exploit children through social networking or destruct process control systems, and an ethical hacking course.

5 Cultivating a Cybersecurity Culture in a Police Force

In this section, we present a framework (Table 1) for establishing and running a Cybersecurity and Cybercrime Combatting Culture (CCCC) program in a police force as adapted from prior research [27]. Each phase is expanded below the table.

5.1 Preparation Phase

Step a: Setting up Cybersecurity and cybercrime strategies and policies

The CSC program aims to foster a culture where cybersecurity and cybercrime policies and knowledge become instilled in members of a police force. It supports the cybersecurity governance and best practices prescribed by the cybersecurity strategy

Table 1. A framework for a police officer CCCC program

Preparation phase	Design phase	Execution phase
a: Cybersecurity and cybercrime strategies and policies must be in place	**1:** Set up the core CCCC task team	***i*:** Run awareness and educational campaigns
b: All police officers must be included in the program	**2:** Define main goals, success criteria and target groups	***ii*:** Implement training programs online and run competitions to ensure cybercrime awareness. Run cybersecurity exercises
c: Understand current culture and processes, and access the risks	**3:** Identify roles and responsibilities and develop training material	***iii*:** Measure the success of training exercises
d: Set up an initial baseline, i.e. the current behaviours	**4:** Identify supporting divisions	***iv*:** Return to Step 6
e: Run a pilot activity and measure its impact	**5:** Design cybersecurity and cybercrime exercises and competitions with identified metrics	
f:: Get buy-in from upper level command	**6:** Review and update the program	

and policies, including ethical conduct and values. Support for this program must come from all role players including IT support services and cybercrime investigators. The cybersecurity policies must include a description of the distribution of appropriate information, including updates, in a timely manner to enhance police officers' understanding. It must include a Cybersecurity Risk Management Program that includes risk assessment, the monitoring of networks to prevent and detect incidences, measures on the response to incidents as well as systems recovery and maintenance. These policies must be implemented in such a way to create an integrated, coherent system of security to protect the police's own systems against cyber criminals.

Step b: All police officers be included

All police officers must be trained according to the capability and capacity building program (Sect. 4), and included in the CCCC program. The IT staff that maintains the police network and systems must also be trained and certified to protect their own systems.

Step c: Understand current culture and processes, and access the risks

The CCCC program staff must engage with police force members to determine the existing cultures in the police force units and divisions. Risk assessment should be conducted to identify current threats and vulnerabilities. This will identify misalignments between existing practices and processes and security measures which may require engagements to find resolutions that will ensure commitment from all.

Step d: Set up an initial baseline

Determine the current maturity level of the current CCCC by using appropriate metrics. Some metrics are published by the Payment Cards Industry [28]. This can be refined and used for the measurement of the impact of the program.

Step e: Run a pilot CCCC activity and measure the impact

Run a small pilot program within a single unit. On completion of the pilot, changes in the selected behaviour against the baseline can illustrate the potential impact. For example, the activity of using a password storage app instead of storing personal passwords in a plain text. It will be easy to measure improvement after this activity.

Step f: Get buy-in from the upper level command

Strong motivations must be developed to convince the commissioning staff to support and fund the CCCC program. These programs require financial and human resources, funding and a dedicated team to run and adapt the program over time. The motivation can be done by incorporating:

- Current cyber threats statistics including global, national and regional data.
- Evidence of cyber-attacks in the police.
- The results from the pilot activity.
- Estimates of the financial and reputational impact of cyber-attacks and breaches.

5.2 Design Phase

Step 1: Set up a core CCCC task team

The core task team is responsible for the development, implementation and maintenance of the CCCC program and must represent a cross-section of the police force by including members of different units. The team must understand the group cultures in the different units. Adaptations of the program can be made for individual groups if necessary. The core group must be continuously supported by the senior commissioning officers to formulate CC policy and strategy and to oversee the implementation and execution of the program. The team must also allow different divisions and units to give inputs and feedback on the implementation in their group to ensure participation in the long run, to make use of their internal expertise, be innovative and feel they contributed.

Step 2: Define main goals, success criteria and target groups

Define the main goals for the CCCC program that prioritise the most important issues and criteria for the target group's successes. The impact of adversary cyber exploitation must be kept at an acceptable level according to standardised processes for assessing risk to mission assurance [29], and every member has a role and a responsibility to achieve this.

Criteria of success measurements must be identified in an early stage. Performance outcomes are critical and it is important to measure the actual state of cybersecurity and the impact of the program. More complex measurements will also be needed because culture consist of more than only behaviours; it also includes values, attitudes and perceptions about cybersecurity. Mlamedal and Roislien [12] motivate the identification of a robust set of indicators that enable the creation of an appropriate baseline for the measurement of a CCCC. Accountability by members are also important [29].

Threat exchange of cybersecurity information must be distributed on a regular basis to police officers.

The identification of target groups is complex due to each division having its own organisational culture. A CCCC will not be successful if it is enforced on a group; it must be cultivated over time, with the required adaptations. Examples of different target groups are the Senior Management: Commissioned Officers, Commissioning Officers, Non-Commissioning officers, IT providers and the Cybercrime units.

Step 3: Identify responsibilities of task members and develop training material Responsibilities are assigned to IT support police officers according to the risk assessments. These assessments include risks due to system vulnerabilities, threats to systems and the impact to a police officer's missions. All security controls for programs must be supplemented with cybersecurity measures. Training materials must be developed according to the job profiles.

Figure 2 depicting the roles is an adaptation from a figure in that appeared in [27].

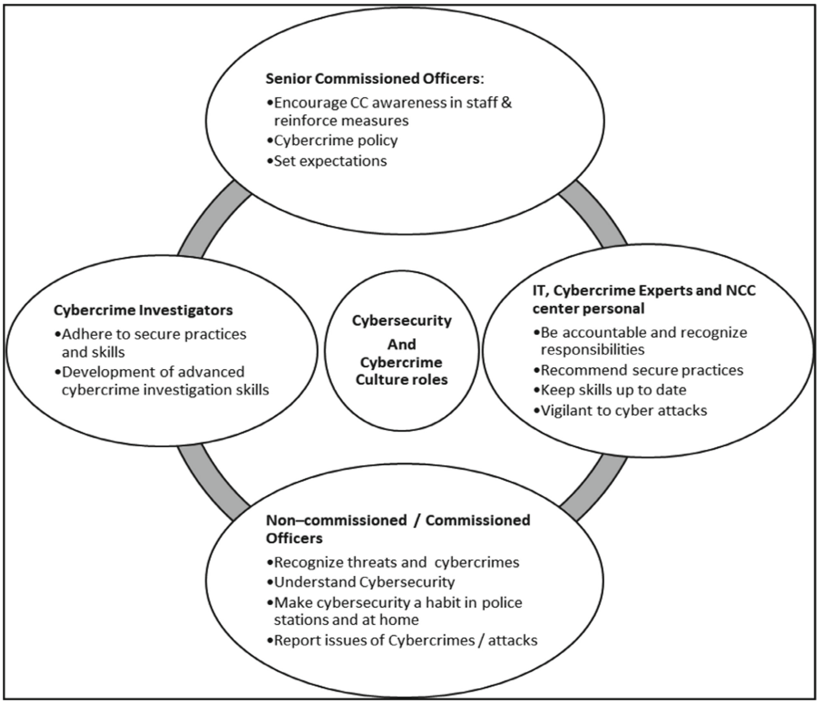

Senior Commissioned Officers:
•Encourage CC awareness in staff & reinforce measures
•Cybercrime policy
•Set expectations
Cybercrime Investigators
•Adhere to secure practices and skills
•Development of advanced cybercrime investigation skills
Cybersecurity And Cybercrime Culture roles
IT, Cybercrime Experts and NCC center personal
•Be accountable and recognize responsibilities
•Recommend secure practices
•Keep skills up to date
•Vigilant to cyber attacks
Non–commissioned / Commissioned Officers
•Recognize threats and cybercrimes
•Understand Cybersecurity
•Make cybersecurity a habit in police stations and at home
•Report issues of Cybercrimes / attacks

Fig. 2. Cybersecurity and cybercrime combatting culture roles for the police force

Step 4: Identify supporting divisions

Support for the initiative must come from all the service departments in the police force. Commissioning officers at every level should actively be encouraged to participate in the program by including performance in the program into performance reviews.

Step 5: Design cybersecurity and cybercrime exercises and competitions with identified metrics

Cybersecurity and cybercrime assessment exercises can be developed to evaluate the preparedness of all police units. These exercises are an effective way to measure the level of collaboration and information sharing between the different divisions and units. They must cater for different target groups. Different platforms such as videos, instructor-led training, quizzes, games, email, internal social media etc. can be used. In addition, cybersecurity drills can be arranged for units [21]. Communication is important to emphasise the necessity of the exercises and to ensure continued participation. Exercises must also include the identification of cybercrimes and the assessment of correct measures taken in the evaluation of a cybercrime scene. A set of metrics must be established to measure success.

There also must be customised exercises for senior commissioning officers focused on policies. These activities must include activities geared for staff with knowledge levels ranging from no technical knowledge to a high level of IT knowledge and competence. A Best Practice manual must accompany all training material to encourage the adoption of practices aimed at promoting secure online behaviour.

Step 6: Review and update the program

Culture is dynamic and the CCCC program has to be updated frequently. CCCC should become part of the Police Academy training and HR Development programs. The Cybersecurity Risk Management Program must be used to reassess the status of cybersecurity on a regular basis to make provision for new and changing threats and vulnerabilities, and to adapt security policies, practices and procedures. Measurements and the baseline must be updated to make provision for the changes.

5.3 Execution Phase

Step i: Run Awareness and Educational Campaigns

Education and awareness for force members are essential elements of a cybersecurity culture. The emphasis must not only be on technical, administrative and procedural measures to protect computer systems, but also to keep member informed on the latest threats through awareness campaigns. In addition, due to the national cybercrime centres' research and 24/7 centres' international co-operations, these types of centres can be the key points of contact for cybersecurity matters and can facilitate information and technology sharing for the creation of awareness in the police force.

Step ii: Run Cybersecurity Exercises

The focus of these exercises is to evaluate the preparedness of staff in police units as well as to support structures to deal with cyber-attacks. Cybersecurity drills are also an option. The most effective exercises for senior commissioning officers are scenario-based. Cybersecurity experts must facilitate these exercises that can include, among other things, the loss of connectivity that interrupts the communication, and ransomware exercises. Special weeks can be identified for the hosting of exercises e.g. during Cyber Security Month (usually in October) that can be aimed at reaching a large number of police officers.

Step iii: Measure the success of the exercises

Several metrics can be used to measure performance [28]. For example, if an activity's aim was to make participants aware of the importance of logging out of workstations when they leave an office, the baseline will be the number of workstations that were open in the absence of the user (physical inspection) before the activity. The current behaviour can be determined by a similar count after the activity was run.

Step iv: Return to Step 6

This programme for cultivating a CCCC in a police force needs to take place once the necessary structures are in place (Sect. 3), and needs to run in conjunction with ongoing capacity and capability building (Sect. 4).

6 Conclusion

Cybersecurity awareness and appropriate cyber behaviour in a police force is often neglected even when police offers receive training to combat cybercrime. The cultivation of a cybersecurity culture in a police force is essential to protect their systems and to support other cybersecurity training. Although there are several frameworks and other resources available for an organisation to cultivate a cybersecurity culture, law enforcement agencies face a different environment. This paper considers frameworks and resources on the cultivation of organisational cybersecurity culture to present a customized framework for a police force. In addition, requirements to create a cybercrime combatting culture in a police force environment are specified. The paper thus presents comprehensive guidelines in this regard that can be implemented and applied by police forces in Africa.

References

1. Newhouse, W., Keith, S., Scribner, B., Witte, G.: National initiative for cybersecurity education (NICE): cybersecurity framework. NIST Special Publication (2017). https://doi.org/10.6028/NIST.SP.800-181
2. Herskovits, M.J.: The contribution of Afro-American studies to africanist research. Am. Anthropologist **50**(1), 1–10 (1948). https://doi.org/10.1525/aa.1948.50.1.02a00020
3. Schein, E.H.: Organizational Culture and Leadership. A Dynamic View. Jossey-Bass, San Francisco (1985). https://doi.org/10.1002/hrm.3930240312
4. European Agency for Network and Information Security (ENISA): Cyber security culture in organisations (2017). https://doi.org/10.2824/10543
5. Gcaza, N., Von Solms, R., van Vuuren, J.J.: An ontology for a national cybersecurity culture environment. In: The Nineth International Symposium on Human Aspects of Information Security and Assurance (HAISA), pp. 1–10 (2015)

6. Reid, R., van Niekerk, J.: Towards an education campaign for fostering a societal cyber secure culture. In: The Eighth International Symposium in Human Aspects of Information Security and Assurance (HAISA), pp 174–184 (2014)
7. Ponemon Institute: The human factor in data protection (2017). https://www.ponemon.org/blog/the-human-factor-in-data-protection
8. Gcaza, N., von Solms, R.: Cybersecurity culture: an ill-defined problem. In: Bishop, M., Futcher, L., Miloslavskaya, N., Theocharidou, M. (eds.) WISE 2017. IAICT, vol. 503, pp. 98–109. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58553-6_9
9. Schlienger, T., Teufel, S.: Information security culture - from analysis to change. S. Afr. Comput. J. **31**, 46–52 (2003)
10. International Telecommunications Union: ITU corporate annual report (2008). https://www.itu.int/osg/csd/stratplan/AR2008_web.pdf
11. Kortjan, N., Von Solms, R.: Fostering a cyber security culture: a case of South Africa. In: ZA-WWW 2012 Conference (2012)
12. Mlamedal, B., Roislien, H.E.: The norwegian cyber security culture. Norwegian Centre for Information Security (NorSIS) (2016). https://norsis.no/wp-content/uploads/2016/09/The-Norwegian-Cybersecurity-culture-web.pdf
13. OECD recommendation of the council on digital security risk management for economic and social prosperity (2015). https://legalinstruments.oecd.org/en/instruments/116.OECD/LEGAL/0415
14. Veltsos, C.: Building a cybersecurity culture around layer 8. Security Intelligence (2017). https://securityintelligence.com/building-a-cybersecurity-culture-around-layer-8/
15. HMICFRS Cyber: Keep the light on. An inspection of the police response to cyber-dependent crime (2019). https://www.justiceinspectorates.gov.uk/hmicfrs/publications/keep-the-light-on-police-response-to-cyber-dependent-crime/
16. Florentine, S.: Closing the cybersecurity talent gap, one woman at a time (2015). https://www.cio.com/article/3005637/cyber-attacks-espionage/closing-the-cybersecurity-talent-gap-one-woman-at-a-time.html
17. Workforce Intelligence Network for Southeast Michigan: Cybersecurity skills gap analysis (2017). https://winintelligence.org/wp-content/uploads/2017/07/FINAL-Cybersecurity-Skills-Gap-2017-Web-1.pdf
18. CSIS Intel Security: Hacking the skills shortage (2016). https://www.csis.org/programs/technology-policy-program/cybersecurity-and-warfare/other-projects-cybersecurity-0
19. NICE Cybersecurity Workforce Framework Work Roles: National initiative for cybersecurity careers and studies (2017). https://niccs.us-cert.gov/nice-cybersecurity-workforce-framework-work-roles
20. e-Governance Academy (EGA): National cyber security index 2017 South Africa (2018). https://ncsi.ega.ee/country/za/
21. Usmani, K.A., Appayya, J.A.: Capacity building is the key to fight against cybercrime: the mauritian perspective. Glob. Cyber Expertise Mag **4**, 4–6 (2017)
22. Global Project on Cybercrime: Capacity building on cybercrime (2013). http://www.combattingcybercrime.org/files/virtual-library/capacity-building/capacity-building-on-cybercrime.pdf
23. Seger, A.: Cybercrime strategies. Global Project on Cybercrime (2012). https://rm.coe.int/16802fa3e1
24. Federal Buro of Investigation (FBI): Offer online cyber training for law enforcement first responders (2016). https://www.fbi.gov/news/stories/online-cyber-training-for-law-enforcement-first-responders
25. Christopoulos, G.: First responders and cyber forensics (2017). https://www.cepol.europa.eu/media/blog/first-responders-cyber-forensics

26. Shook, S.: Cybercrime investigation body of knowledge (2019). https://www.cibok.org/en/
27. Leenen, L., van Vuuren, J.C.J.: Framework for the cultivation of a military cybersecurity culture. In: 14th International Conference on Cyber Warfare and Security (ICCWS) (2019)
28. PCI Security Standards Council: Best practices for implementing a security awareness program (2014). https://www.pcisecuritystandards.org/documents/PCI_DSS_V1.0_Best_Practices_for_Implementing_Security_Awareness_Program.pdf
29. Snyder, D., Powers, J.D., Bodine-Baron, W., Fox, B., Kendrick, L., Powell, M.H.: Improving the cybersecurity of U.S. Air force military systems throughout their life cycles. RAND Corporation (2015)

A Consideration of the Case Study of Disinformation and Its Legal Problems

Tomoko Nagasako

The Sasakawa Peace Foundation, Tokyo, Japan
t-nagasako@spf.or.jp

Abstract. Recently, some countries deploy global cyberattacks that not only impose destructive measures to the system of industries or infrastructures but also as information warfare, including social networking service (SNS) and other media that affects election results or democratic processes, which becomes a threat to democracy. Thus, this operation is recognized as "disinformation." This paper demonstrates cases of disinformation in cyberspace, and focuses on legal problems in the international law and countermeasures of legal systems in each country.

Consequently, it is found to be challenging to deal with disinformation on the national scale. As there is a limit regarding the regulations by international law, at the present, it is essential to provide the national law about it. I classified the types of countermeasures to find better countermeasures to it based on my considerations, as the number of disinformation cases increased. The regulation for disinformation could violate the freedom of expression and democracy. Therefore, posteriori sanctions against foreign state actors should be applied, and regulations on the contents of media and platformers need to be practiced carefully.

Keywords: Disinformation · Election meddling · Tallinn manual · International law · National law · Hybrid warfare

1 Preamble

Recently, cyber-physical systems are implemented in all areas due to the high growth of information technology. The degree of digitalization and the networking of humans and things are growing rapidly. In modern society, the high added value is found in data accumulation and analysis, which is used in a lot of services. Consequently, a data-driven society is being created; while there is increasing convenience, their risk also increases in parallel, such as information systems being destroyed or compromised, leakage of personal information, unauthorized acquisition and use of intellectual property, and influence operation using social network service (SNS). These changes of risks have also transformed the form of warfare into a new type.

The views expressed in this article are those of the author in my personal capacity.

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 262–276, 2020.
https://doi.org/10.1007/978-3-030-62803-1_21

The cyberspace is recognized as the fifth battlefield, and various cyber tools are incorporated into each country's military strategy. Consequently, the newest warfare shifts from the modern war of using kinetic military weapons to a hybrid war that weaponizes all state activities, including kinetic weapons. There is an increasing sense of crisis in hybrid warfare, that is, it is challenging to draw the line between regular and non-regular battles. In 2017, NATO and the EU established a think tank called "The European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE) in Finland. They considered and implemented countermeasures against hybrid threats from various perspectives. The report [1] that Hybrid CoE released in 2018 cites the analysis of the German Marshall Fund's Alliance for Securing Democracy [2] and points out that the Russian government has used cyberattacks, disinformation, and financial influence campaigns to meddle in the internal affairs of at least 27 European and North American countries since 2004. It was pointed out in the previous studies that China has been interfering in domestic affairs through the same operations [3, 4].

Thus, this paper pays attention primarily to disinformation, which hybridizes the tools in Table 1 [5], such as propaganda, fake news, strategic leaks, or organized protest movements.

Table 1. Range of hybrid tools

Tools	Salient points
Propaganda	Enabled and made cheaper by social media, also targeted at home
Fake news	"Lisa" was portrayed as a Russian-German raped by migrants
Strategic leaks	Macron emails leaked 48 h before the election
Funding organizations	China opened Chinese think-tank in Washington
Political parties	Russia supports sympathetic European parties on right and left
Organized protest movements	Russian trolls organized both pro- and anti- protests in Houston mosque case
Cyber tools: • Espionage • Attack • Manipulation	New tool in arsenal: espionage is old tactic with new, cyber means. Attack has targeted critical infrastructure, notably in Estonia in 2007. Manipulation is next frontier, changing information without the holders know it
Economic leverage	China sought to punish South Korea for accepting U.S. anti-missile system
Proxies and unacknowledged war	Hardly new, but "little green men" in Ukraine slid into actual combat
Paramilitary organizations	Russian "Night Wolves" bikers intimidate civilians

Also, an overview of articles about the keyword 'disinformation' from the Journal of Information Warfare [6], a journal closely related to this paper, and reports from US think tank, Atlantic Council [7], shows the current emphasis on Russian operations in North America and Europe. However, as discussed below, in practice, Russia has expanded its activities to Africa and South America, and Chinese disinformation activities are becoming more and more influential in the Asian region. This paper explores these trends through the case study.

Disinformation is a severe challenge to the democracy, since it is executed by combining the leakage of information stolen by cyberattacks with information warfare in media and SNS to transformed public opinion in each country and influence democratic processes, such as elections and demonstrations the outcome. However, planning countermeasures and regulations under national and international cooperation is an urgent issue for disinformation. It is a powerful and complex operation that threatens national sovereignty and sway our democratic system. Hence, it is taken to be one of the new forms of warfare created by the data-driven society that needs to be conquered to ensure a sustainable democracy.

2 What Is Disinformation?

Since Russia's election meddling[1] in the 2016 US presidential election attracted attention, similar operations by Russia or China emerged. The term of disinformation seems to have become popular. However, some countries use *fake news* in a context similar to disinformation. Though Japan is a representative example of such country, the term *fake news* is not reasonable when discussing foreign influence operations from a national security point of view. *Fake news* is a part of the influence operation, and it does not suit the whole process.

Here, the definition of disinformation should be reconsidered, because more clarifications may be required to make the discussion appropriate.

The European Commission's report [8] calls the situation, including not only influence operations by state actors, but also the dissemination of false information due to negligence, as information disorders, and shows the following three types of data under such circumstances: mis-, dis-, and mal-information. Using the scopes of harm and falseness, it describes the differences between these three types of information (see Fig. 1) as:

- Mis-information is when false information is shared, but no harm is meant.
- Dis-information is when false information is knowingly shared to cause harm.
- Mal-information is when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere.

The report by the high-level expert group on *fake news* and online disinformation of EU committee [9] also defined *disinformation* as all forms of false, inaccurate, or

[1] Here, I use this word, 'election meddling', although some words such as election meddling, election interference or election intervention are used without distinguishing the meanings. This is because the term of intervention is distinguished from the term of interference in international law, and it is hard to distinguish them and to determine which term should be used for each case as following considerations in Sect. 4. So I choose 'election meddling' without relationship with the argument in international law.

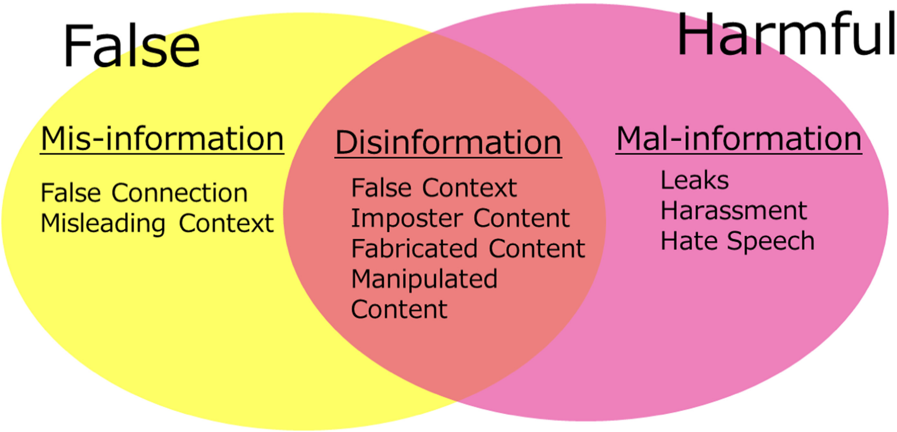


Fig. 1. Definition of disinformation by EU

misleading information designed, presented and promoted to intentionally cause public harm or for profit.

However, the definitions are inadequate and seem misleading because they show that disinformation consists of false information only. But disinformation also contains the right information.

For example, in the US presidential election of 2016, the Office of the Director of National Intelligence's report [10] alleges that Russian military intelligence (General Staff Main Intelligence Directorate or GRU) used the Guccifer 2.0 persona and DCLeaks.com to release the e-mail data they stole from the Democratic National Committee. This disclosure may have been in a false context, but the data are not wrong.

Also, a specific type of hate speech like in the French presidential election of 2017 has the possibility of truth. In this election period, hate speeches that recognized Macron as a gay with harassment spread widely on some media and SNS [11]. In this case, these were fake news because Macron denied being gay [12] but, if these are true, are these hate speeches not as effective as disinformation? It is immaterial whether it is true or false when an operation uses sensitive information such as religion or sexual orientation. Such a sensitive thing is hard to be fact-checked by a third party, and it is a success for disinformation operation that causes anxiety, confusion, or split in the society to make a social divide wider and damage our democracy. The state actors distort and manipulate the contents of hate speech. So, we should distinguish disinformation that is operated in the frame of the national strategy from ordinary hate speech, and we should exercise caution to correct but harmful information as a part of disinformation.

Figure 2 shows a modified definition of disinformation. Disinformation contains also true information such as manipulated contents to give a wrong impression or inconvenient truths to harm someone deliberately. If we do not catch the multiple perspectives of disinformation completely, we may not deal with this sophisticated information warfare.

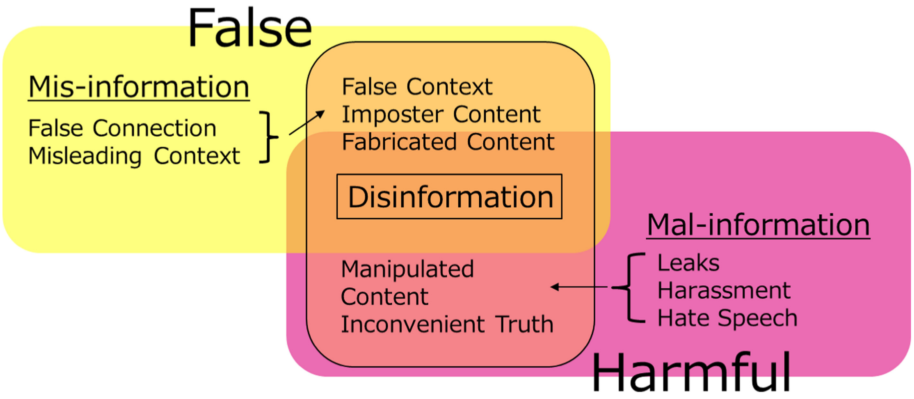


Fig. 2. Definition of disinformation by the author

3 Disinformation Cases

This chapter shows how much disinformation happen in this world.

As a part of disinformation, the first focus is on election meddling. According to the report [13] of The Canadian Centre for Cyber Security (CCCS), the proportion of national elections in 2018 targeted by foreign cyber threat activity has more than doubled since 2015. As for the Organization for Economic Co-operation and Development countries, the proportion of elections targeted by cyber threat activity is more than 3/4 from 2015 (15.4%) to 2018 (50.0%) [14]. The vast majority (88%) of cyber threat activities affecting democratic processes around the world since 2010 have been strategic (i.e., threat actors specifically targeted a democratic political process to affect the outcome) [15]. Then, the major remainder of the cyber threat activities was cybercrime, which is stealing voter data to sell personal information or use it for criminal purposes. Furthermore, CCCS shows that voters now represent the single largest target of cyber threat activity against democratic processes, accounting for more than half of global activity in 2018 [16]. They explains that this shift seems to have started in 2016, which is likely due to the perceived success among cyber threat actors. Therefore, most foreign adversaries consider the costs and benefits of possible cyber threat activities before undertaking them. They likely recognize targeting voters to be a more effective way to interfere with democratic processes than targeting elections through political parties, candidates, and their staff. The reason is that web media and SNS have made it easier and cheaper to influence the cognitive domain of vast numbers of people.

Figure 3 and Table 2 present the original data of concrete cases of disinformation from 2016. The 2016 example seems to be a turning point because the term of disinformation got more recognized widely after the US presidential election. This data includes not only votes but also some democratic events such as referendums or demonstrations, and it consists of cases I investigated from open sources like government reports and news articles. Though, the CCCS do not make their data available due to security reasons. So, this report is not consistent with the data of CCCS's report.

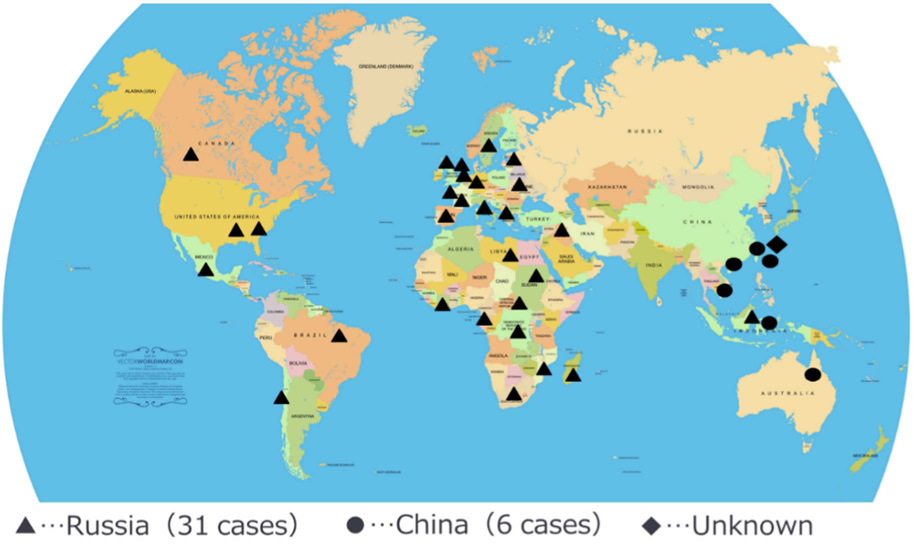


Fig. 3. Disinformation Cases (since 2016) (I made this figure thanks to the free map by VECTORWORLDMAP.COM, version 2.2 and COPYRIGHT 2009, Graphics Factory CC.)

Table 2. Disinformation cases (since 2016)

2016

	Date	Area	Case	Actor
1	2016/1/16	Taiwan	Presidential election and Legislative election	China
2	2016/4/6	The Netherlands	Dutch Ukraine–European Union Association Agreement referendum	Russia
3	2016/6/23	United Kingdom	United Kingdom European Union membership referendum	Russia
4	2016/11/8	United States	Presidential election	Russia

2017

	Date	Area	Case	Actor
1	2017/3/15	The Netherlands	General election (House of Representatives)	Russia
2	2017/5/7	France	Presidential election	Russia
3	2017/9/24	German	Federal election	Russia
4	2017/9/25	Iraq	Kurdistan Region independence referendum	Russia
5	2017/10/1	Spain	Catalan independence referendum	Russia

2018

	Date	Area	Case	Actor
1	2018/3/4	Italia	General election	Russia

(*continued*)

Table 2. (*continued*)

2018				
	Date	Area	Case	Actor
2	2018/7/1	Mexico	General election	Russia
3	2018/7/29	Cambodia	General election (House of Representatives)	China
4	2018/9/9	Sweden	General election (House of Representatives)	Russia
5	2018/9/30	Macedonia, Greek	Macedonian referendum	Russia
6	2018/9/30	Japan	Okinawa gubernatorial election	Unknown
7	2018/10/7	Brazil	General election	Russia
8	2018/11/6	United States	Midterm election	Russia
9	2018/11/17	France	Yellow vests movement	Russia
10	2018/11/24	Taiwan	Local elections, Kaohsiung mayoral election	China
11	2018/12/19	Madagascar	Presidential election	Russia
2019				
	Date	Area	Case	Actor
1	～2019/3/4	Estonia, Latvia, Lithuania	Estonian parliamentary election	Russia
2	2019/3/31	Ukraine	Presidential election	Russia
3	2019/3/31～	Hong Kong	Hong Kong protests	China
4	2019/4/17	Indonesia	Presidential election	China, Russia
5	2019/5/8	South Africa	General election (House of Representatives)	Russia
6	2019/5/18	Australia	General election	China
7	2019/5/23-26	EU	Elections to the European Parliament	Russia
8	2019/10/18～	Chile	Chilean protests	Russia
9	2019/10/21	Canada	Federal election	Russia
10	2019/10/30 *	8 African countries	Elections or Political movements	Russia
2020				
	Date	Area	Case	Actor
1	2020/1/11	Taiwan	Presidential election and Legislative election	China

*This date is not the date of the event but the date when the news that Facebook banned Russian accounts that were related to disinformation operation was reported, because this case expands over some elections and political movements in each county.

The data shows that the area where Russia and China would like to have a strong influence is Europa and Pacific Rim community, respectively. Also, it is manifest that Russia meddles in Africa. These results correspond with their national strategy to expand digital authoritarianism.

Although few cases were investigated, the trends shows that disinformation cases are increasing yearly, which suggests immediate countermeasures against disinformation.

4 Considering the Wrongfulness of Disinformation by International Law

As observed earlier, disinformation is a global problem. Since disinformation is a conflict between nations, it may be necessary to consider the unlawfulness of disinformation in the context of international law, and international law should regulate disinformation.

On that note, Tallinn Manual 2.0 [17], which was facilitated and led by the NATO Cooperative Cyber Defence Centre of Excellence, and which summarizes the concept of international law applied to cyber operations is seen. This book does not create new international laws or regulations related to cyberspace and cyber operations. Still, on the assumption that customary international law applicable to cyber operations exists, it confirms and describes 154 rules and its' contents of international law. Here, it is good to consider the unlawfulness of election meddling to be the main operation of disinformation under the related rule of this book.

- **Rule 4. – Violation of sovereignty**
  A state must not conduct cyber operations that violate the sovereignty of another state [18].

Based on this rule, cyberattacks and cyber espionage conducted by a state organ in the territory of another country are considered a violation of sovereignty. With regard to remote cyber operations, cyberattacks that cause physical damage or loss of functionality in cyberinfrastructure, and cyber operations that interfere with data and services that are necessary to exercise inherently government functions is considered to be a violation of sovereignty, such as changing or deleting data such that it interferes with the delivery of social service, the conduct of elections, the collection of taxes, the effective conduct of diplomacy, and the performance of key national defense activities.

In terms of election interference, it becomes a violation of sovereignty only when there is a level of interference, such as manipulating election voting data through cyberattacks or interfering with the operation of polling stations. So, Information stolen by hacking from election-related organizations and the influence operations using media and SNS will not be considered violation of sovereignty.

- **Rule 32. – Peacetime cyber espionage**
  Although peacetime cyber espionage by states does not per se violate international law, the method by which it is carried out might do so [19].

This rule is a matter of whether operations such as election meddling constitute unlawful cyber espionage. The operations of disinformation, including election meddling, are so highly compatible with the intelligence agency that at first glance, i.e., the operation itself appears to be included in the cyber espionage. The international hacking groups such as APT28, APT29, and APT40, which are alleged to be involved in election meddling so far, have been pointed out from the attribution results that they have the back of the Russian and Chinese intelligence community such as GRU, FSB, and Chinese People's Liberation Army, respectively [20–22]. However, when preventing cyberattacks and cyber espionage, it is necessary to clarify the attribution of the actor conducting the operation, and such activities are similar to normal intelligence activities. Therefore, on the defense side, the intelligence agencies are also involved.

This rule states that the term 'cyber espionage' refers to any act undertaken secretly or under false pretenses that uses cyber capabilities to or attempt to, surveil, monitor, capture, exfiltrate, or gather electronically transmitted or stored communications, data, or other information. So, in this context, the rule does not seem to include the covert action to influence or work on another country such as election meddling.

Besides, it should be cautioned that cyber espionage may be conducted in a manner that violates international law due to the fact that certain methods employed to conduct cyber espionage are unlawful. If cyber operations that are undertaken for espionage purposes violate the international human right to privacy, the cyber-espionage operation is unlawful. So, the operation of election meddling is unlawful, if the operation is conducted with, not only an influence operation on SNS but also the cyberattack to steal and leak the e-mails of candidates or election offices, such as in the US and France presidential elections.

- **Rule 66. – Intervention by states**
  A state may not intervene, including by cyber means, in the internal or external affairs of another state [23].

This manual explains that this rule prohibits coercive intervention, including cyber means, by one state into the internal or external affairs of another. It is based on the international law principle of sovereignty, precisely that aspect of the principle that provides for the sovereign equality of states. In this rule, intervention is clearly distinguished from interference with no coerciveness. For the purpose of this rule, interference refers to acts by states that intrude into affairs reserved to the sovereign prerogative of another country, but lack the requisite coerciveness to rise to the level of intervention. The term of intervention, the subject of this rule, is limited to acts of interference with a sovereign prerogative of another state that have coercive effect. The key is that the coercive act must have the potential for compelling the target state to engage in an action that it would otherwise not take.

So, here, I consider the case of election meddling. Even if disinformation operations are conducted in the media or SNS, as long as various voting possibilities remain, it can be said that it is not unlawful election intervention, but only election interference. It can be recognized as an unlawful election intervention only when a candidate is killed, or the election opportunity itself is lost due to the destruction of the election infrastructure by the attack of another country.

As mentioned above, it seems that there is a limit to identify the wrongfulness of disinformation under current international laws. So, it will be a challenge of future international initiatives to consider what kind of regulation should be taken under international laws from now on, and what type of legislation is useful in the national law of each country.

The G7 “Declaration on Responsible States Behavior in Cyberspace” (i.e., the “Lucca Declaration” [24]) in 2017 expresses their opinion that “We note that, in the interest of conflict prevention and peaceful settlement of disputes, international law also provides a framework for States’ responses to wrongful acts that do not amount to an armed attack - these may include malicious cyber activities. Among other lawful responses, a State that is the victim of an internationally wrongful act may, in certain circumstances, resort to proportionate countermeasures, including measures conducted via ICTs, against the State responsible for the wrongful act in order to cause the responsible State to comply with its international obligations”. It is crucial that they explicitly point out that international wrongful acts include malicious cyber activities. This expression can be recognized as an advanced endeavor to deal with malicious cyber operations that are beyond the scope of existing customary international laws in the framework of new international norms. Such a new movement will have possibilities to create a new framework of international regulations to deterrent disinformation.

A similar international cooperation initiative ‘The Paris Call for Trust and Security in Cyberspace’ was announced by French president Macron at the IGF in 2019. This Paris Call refers to solving problems, such as to prevent malign interference by foreign actors aimed at undermining electoral processes through malicious cyber activities, and to promote the widespread acceptance and implementation of international norms of responsible behavior as well as confidence-building measures in cyberspace, and this More than 50 countries and 250 organizations have signed the Paris Call.

However, given the adoption of Additional Protocol to the Convention on Cybercrime, concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems in 2003, which remains ineffective, any initiatives lack the power to deter their operations without the involvement of Russia and China. The same lack of participation by China and Russia also exists in the G7 and Paris Call, and it is crucial for the formation of new international norms to deter disinformation how these digital authoritarian states are involved.

5 The Types of Countermeasures by the World’s Nations

As seen in the previous chapter, the regulations by international law do not work effectively at present. So, for the time being, we should take countermeasures through national law.

In this section, to the report [25] of the Poynter Institute, that is, a guide for existing attempts to legislate against what can broadly be regarded as online misinformation is referred. At present, they investigated countermeasures of 53 countries and classified their types, focuses, orientations, and details. The authors also recognize the confusing use of the terms of mis- or disinformation, so they seem to choose the term "misinformation" to cover all these concepts, although they do not show and clear the definition in this guide. Then, rearranging these data can show the types of countermeasures. So as to address the problems among the countermeasures, the discussion range is set wider covering all information disorders such as mis-, dis-, mal-information.

Among countermeasures for information disorder, there are 31 of the 53 countries surveyed adopted legal measures such as new legislation and amendments to current laws (see Table 3), which is more than other measures. Additionally, to the measures listed in Table 3, each country has various original measures, such as the establishment of specialized government offices, the creation of a disinformation database, taxation on social media, shutting down the Internet, and making policy recommendations by legislators. Of course, most countries have adopted several measures in multiple layers. However, Table 3 shows that legal regulation is a priority for these countries.

Table 3. Countermeasures for information disorder (Top 5 types)

Countermeasures	Contents	Countries
New Law	Regulations by a legislation or a amendment	31
Arrest	Applying existing laws to cases to arrest and charge actors	12
Media Literacy Campaign	Improving the media literacy of voters or the entire nation	11
Task Force	Setting a special team to monitor or investigate suspicious operations	8
Fact Checking	Checking factual information whether it is true or false, and opening the result	8

Therefore, it has classified into the following three types by examining what kind of legal regulation each country enforces: rules on contents of media and platformers, posterior sanctions against foreign state actors, and rules on anti-establishment speeches.

First, the typical examples of regulations on the contents of media and platformers are German and French legislation. In Germany, the Network Enforcement Act (Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken, NetzDG) passed in 2017 forces online platforms to remove posts that express obvious illegal contents based on German penal code, including mis-, dis- and mal-information, within 24 h or face risk fines of €50 million. This Act target social networks with more than 2 million users such as Facebook, YouTube, and Twitter. Furthermore, France passed the law against the manipulation of information (LOI organique n°2018-1201 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information) in 2018. The law gives authorities the power to remove fake content spread via social media and even block the sites that publish such, as well as enforce more financial transparency for sponsored content, in the three months before an election. This law also provides a definition of "fake news": "Inexact allegations or imputations, or news that falsely report facts, with the intention of changing the genuineness of a vote." It is created to enact strict rules on the media during electoral campaigns and, more specifically, in the three months preceding any election. As for television and radio, if the media that the foreign country has the management rights is reporting fake news, the authorities may order the broadcast to stop. The type of legal regulation on the contents of traditional media or SNS before information disorder, including disinformation spread. However, because of this legal character, this type sometimes is criticized violating freedom of expression.

Second, the typical examples of posteriori sanctions against foreign state actors are American and Taiwanese legislation. In the US, the executive order 13848 (i.e., Imposing Certain Sanctions in the Event of Foreign Interference in a United States Election) was issued in 2018. Thus, within 45 days of the election results, the Director of National Intelligence (DNI) investigated whether there was any election interference, and within another 45 days, the Attorney General and Secretary of Homeland Security to decide whether or not to impose sanctions. It freezes sanctioned persons' assets in the United States and bars them from doing business with Americans. In 2018 midterm election, as a result of the investigation, there was no confirmation of interference with the vote or the alteration of the aggregate results. Moreover, although there was confirmation of influence operations by Russia, China, and Iran, the DNI did not assess the impact on the election results. Taiwan also enacted the anti-infiltration act (反滲透法) in 2020 to prevent foreign hostile forces from interfering to Taiwan. The law prohibits political donations and campaigning for elections under the direction, commission, and financial support of foreign hostile forces, spreading disinformation and obstructing legal demonstrations. This law imposes any miscreant who violates the results five years imprisonment or a fine of five million Taiwanese dollars. It does not regulate the distribution of information because the authorities impose sanctions after the interference of foreign powers is found and upon investigation. Therefore, this type of regulation is considered suitable for the country such as the US or Japan where the right to freedom of expression is paramount, and this type is high possibility that Japan can apply in the legal system from now on. However, it is not easy to operate this regulation because to achieve this, a high attribution ability to identify foreign forces is required.

Finally, the typical example of regulations on anti-establishment speech is the legislation of Russia, China, some other Asians, and African countries. In 2019, Russia passed two legislations banning fake news and disrespect of authorities. One is the Federal Law on Amending Article 15-3 of the Federal Law on Information, Information Technologies and Protection of Information (Федеральный закон от 18.03.2019 № 31-ФЗ "О внесении изменений в статью 15-3 Федерального закона "Об информации, информационных технологиях и о защите информации"), and another one is the Federal Law on Amending the Code of Administrative Violations (Федеральный закон от 18.03.2019 № 27-ФЗ "О внесении изменений в Кодекс Российской Федерации об административных правонарушениях"). Consequently, the dissemination of the wrongful information is banned, such as information that the government has consider to be false; information that is judged to fuel the feelings of hostility, hatred, or malice between groups because of the threat to national security or the threat of public welfare; and false information that may affect the outcome of an election or may undermine the public confidence in the government ability to perform her duties. Platformers are obliged to post corrections and remove content that the government determines to be false, and the government has the authority to order the company to block accounts that spread false information. If the government finds that false information is shared maliciously, the spreader could either face fines of $73,000 or 10 years in prison. As for the amending the code of administrative violations, any act of disseminating information that represents disrespect to Russian society, government, government symbols, constitutions, and ministries is considered illegal. These laws have been criticized against freedom of speech because they stipulate that it is the authority of the government to show that certain information is false or "fake news" under this law, and profane. Similar legislations such as in China, Singapore, and Burkina Faso have also been criticized for the suppression of speech because they have resembled structures that the government, not the judiciary, determines what is illegal information. It is a critical problem to enact the laws that regulate anti-establishment speech in this way on the excuse of countermeasures for information disorder.

As described earlier, this paper classified and argued countermeasures for information disorder. With the current situation in which the definitions of misinformation, disinformation, or fake news are not defined certainly and they are used confusingly, I found it challenging to discuss clearly what the legal regulations are subject to regulation. This paper suggests posteriori sanctions against foreign state actors be considered and applied as the countermeasure for disinformation, because it can focus only on disinformation by state strategy, and it is not related to the aspect of freedom of expression. However, to a certain extent, regulations on contents also are effective to calm down the information disorder including mis-, dis-, and mal-information. Although the situation varies depending on the legal system of the nation, it is necessary to consider the balance between countermeasure for disinformation and freedom of expression in each country.

6 Conclusion

This paper discussed and considered the definition of disinformation, the cases and trends of disinformation, and the countermeasures for disinformation. In general, it is noted that the number of disinformation cases is increasing, and the operations are spreading globally. Moreover, the state actors are shifting the target from the systems or the infrastructures of democratic events such as elections to the voters or the ordinary people. Considering these trends, legal regulations are urgently recommended as the countermeasures to all forms of disinformation. However, since the international law for disinformation is insufficient, many countries ought to cooperate to make the new international norms and rules and the legislations for disinformation. Though, arguments beyond national boundaries are indicated in this critical state to do so. Then, under this situation, it is crucial for the protection of each country's democracy to take the countermeasures by the national law.

Further, although the types of legislation for information disorder are shown, much investigation will be needed to assess which legislation is useful and how it works. However, attention must be paid to avoid allowing the new legislations or countermeasures for disinformation to regulate freedom of expression or participatory democracy. Then, legal issues are forced in this paper, Whereas it is considered crucial to combine various effective countermeasures, such as improving media literacy or fact-checking, in a way that suits each country to establish a democracy based on the human-centric use of data and network. As the environment surrounding disinformation and hybrid war constantly vary in this world, we should continue to make an effort to hold on the situation, investigate, analyze, and cope with this hostile operation exploiting democracy.

Acknowledgement. I am deeply grateful to Prof. Yuasa at Institute of Information Security who offered continuing support and constant encouragement. I am also indebt to Mr. Osawa, Senior Research Fellow at Nakasone Peace Institute who gave me invaluable comments and warm encouragements. Additionally, I would like thank to Enago (www.enago.jp) for the English language review.

References

1. Treverton, G.F., Thvedt, A., Chen, A.R., Lee, K., McCue, M.: Addressing Hybrid Threats. Arkitektkopia AB, Bromma (2018)
2. Alleged Russian political meddling documented in 27 countries since 2004. https://www.usatoday.com/story/news/world/2017/09/07/alleged-russian-political-meddling-documented-27-countries-since-2004/619056001/. Accessed 17 Feb 2020
3. Cheng, Dean: Cyber Dragon: Inside China's Information Warfare and Cyber Operations. ABC-CLIO, Westport (2016)
4. Cardenal, J.P., et al.: Sharp power: rising authoritarian influence. National endowment for democracy (2017)
5. Treverton et al.: supra note 1, p. 4
6. Disinformation (key word search). https://www.jinfowar.com/tags/disinformation. Accessed 25 June 2020
7. Atlantic Council (2016, 2017, 2019). The Kremlin's Trojan Horses 1.0, 2.0, 3.0
8. Wardle, C., Derakhshan, H.: Information disorder: toward an interdisciplinary framework for research and policymaking, p. 5. Council of Europe (2017)
9. The independent high level group on fake news and online disinformation. A multi-dimensional approach to disinformation, p. 5. European Commission (2018)
10. Office of the Director of National Intelligence (ODNI): Assessing Russian activities and intentions in recent US elections. Office of the Director of National Intelligence (ODNI) (2017)
11. Ex-French Economy Minister Macron Could be 'US Agent' Lobbying Banks' Interests. https://sputniknews.comanalysis/201702041050340451-macron-us-agent-dhuicq/. Accessed 17 Feb 2020
12. France election: Macron laughs off gay affair rumours. https://www.bbc.com/news/world-europe-38892409. Accessed 17 Feb 2020
13. The communications security establishment 2019 update: cyber threats to Canada's democratic process. The Communications Security Establishment (2019)
14. The communications security establishment 2019 update: cyber threats to Canada's democratic process. The Communications Security Establishment, p. 16 (2019)
15. The communications security establishment 2019 update: cyber threats to Canada's democratic process. The Communications Security Establishment, p. 15 (2019)
16. The communications security establishment 2019 update: cyber threats to Canada's democratic process. The Communications Security Establishment, p. 17 (2019)

17. Schmitt, M.N., Vihul, L. (eds.): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. Cambridge University Press, Cambridge (2017)
18. Schmitt, M.N., Vihul, L. (eds.): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, p. 17 ff. Cambridge University Press, Cambridge (2017)
19. Schmitt, M.N., Vihul, L. (eds.): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, p. 168 ff. Cambridge University Press, Cambridge (2017)
20. ODNI, supra note7
21. Estonian Foreign Intelligence Service: International security and estonia 2019. Estonian Foreign Intelligence Service (2019)
22. APT40: Examining a China-Nexus Espionage Actor. https://www.fireeye.com/blog/threat-research/2019/03/apt40-examining-a-china-nexus-espionage-actor.html. Accessed 17 Feb 2020
23. Schmitt, supra note14, p. 312 ff
24. G7 declaration on responsible states behavior in cyberspace. https://www.mofa.go.jp/files/000246367.pdf. Accessed 17 Feb 2020
25. Funke, D., Flamini, D.: A guide to anti-misinformation actions around the world. https://www.poynter.org/ifcn/anti-misinformation-actions/. Accessed 17 Feb 2020

Our Digital Lives

Our Digital Lives

Petros Chamakiotis[1], Brad McKenna[2], and Kathrin Bednar[3]

[1] ESCP Business School, Spain
pchamakiotis@escp.eu
[2] University of East Anglia, Norwich, UK
b.mckenna@uea.ac.uk
[3] Institute of Information Systems and Society, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria
kbednar@wu.ac.at

Driven by a growing emphasis on how technology and ubiquitous connectivity affect our lives, we have selected four papers for inclusion in the HCC 14 proceedings of our Working Group 9.5 on “Our Digital Lives.” Our group adopts a multidisciplinary perspective and explores any issues associated with Our Digital Lives, including digital work (in its different forms), the impacts of ubiquitous connectivity on ‘work’ and ‘life’, social media, and virtual working, among others. The four papers speak to our call for papers and show potential to advance relevant theory, whilst also raising societal and practical implications. In what follows, we provide a short presentation of each paper.

In the first paper, titled “Exploring Human Nature in a Technology-Driven Society,” Kathrin Bednar adopts an interdisciplinary perspective to explore three dimensions that help to better understand existing views of human nature and technology: (a) the natural constitution of human beings; (b) the position of human beings within their environment; and (c) human values. Uniquely, her paper brings together original philosophical positions with relevant technological developments and scientific paradigms, concluding with challenges of a dualist view of the human and technology and motivates further investigations of the human-technology. It thus offers research and practical implications of a multidisciplinary nature, along with more philosophical questions, such as what it means to be human and what is the role of technology.

In the second paper, titled “Social materiality of smartphone game apps: Case analysis of Pokémon GO,” Hiroshi Koga takes an interesting case of a game application for smartphones, that of Pokémon GO, to examine how connectivity and digitalization affect our private lives. Drawing on a sociomaterial perspective, considering a variety of different influences (e.g., individual, organizational, material) on how individuals use this game in their everyday lives, the paper argues that connectivity and digitalization have the potential to create “filter bubbles” that have an impact on users’ travel, health, and place consumption behaviors.

Next, the paper by Brad McKenna, Lena Waizenegger, and Wenjie Cai titled “The Influence of Personal and Professional Commitments on Digitally Disconnected Experiences” is premised on the view that, in our ubiquitously connected world, it becomes more and more difficult to disconnect and leave all personal and professional commitments behind while on holiday. Using the theoretical lens of surveillance, the paper reveals that travelers are digitally surveilled by family members, friends, colleagues, and superiors, while they themselves are inclined to engage in social surveillance of their peers which creates the constant urge to learn about any updates from their private and professional networks. The paper offers significant theoretical and practical contributions of a multidisciplinary nature (e.g., tourism and information systems literatures), by explaining how private and professional commitment influence the digital-free travel experience and extend the concept of surveillance to the work context.

The last paper is by Matthew J. Davis and Per Fors and is titled "Towards a typology of intentionally inaccurate representations of reality in media content." The authors explore the troublesome character of three concepts relative to the spread of misinformation and propaganda online: fake news, deepfakes, and cheapfakes. Following a thorough analysis of the problems associated with those concepts, the paper contributes a typology of, what the authors term, Intentionally Inaccurate Representations of Reality (IIRR) in media content. In contrast to deepfakes, cheapfakes, and fake news – terms with mainly negative connotations – this term emphasizes both sides; the creative and fun, and the malicious use of AI and non-AI powered editing techniques.

Overall, the four papers presented here highlight the multifaceted character of research into Our Digital Lives, varying from philosophical understandings of the relationship between humans technology, and the sociomaterial character of virtual games, through to disconnection and surveillance, and the emergence of new phenomena (e.g., IRR) and their impact on our actual lives.

Exploring Human Nature in a Technology-Driven Society

Kathrin Bednar

Institute for Information Systems and Society, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria
kbednar@wu.ac.at

Abstract. Major philosophical works have presented discussions of human nature and interdisciplinary programs have set out to address the interrelation of technology and social factors. Still, only few works have tried to present a bird's-eye perspective on current debates, combining original philosophical positions with relevant technological developments and scientific paradigms. Based on an interdisciplinary definition of human nature, this paper explores three dimensions to better understand and categorize views of human nature and technology: first, the natural constitution of human beings; second, the position of human beings within their environment; and third, human values. It discusses how different accounts of human nature result in different views of the role of technology by reviewing perspectives of the human body and technological enhancement as well as accounts of the human as an isolated individual or social being. Human values are presented along five dimensions, covering individual, social, environmental, technical, and economic aspects, which play unique roles in the human-technology relation. While this paper can only offer a preliminary analysis of positions and arguments, it concludes with challenges of a dualist view of the human and technology and motivates further investigations of the human-technology relation. A better understanding of implicit beliefs regarding humans and technology can inform research and practice in the fields of technology ethics, design, and engineering and can open up space for a positive reconceptualization both of what it means to be human and the role of technology.

Keywords: Human nature · Technology · Human-technology relation · Design

1 Introduction

The ability to build technological artefacts that help and support us has characterized human beings throughout our evolutionary history. Today, technological innovations are still being perceived as a symbol for human intellectual capacities as they increase control over the world as well as over human capabilities and deficiencies. However, scholars have emphasized that technology is much more than a means to an end as it shapes human experiences, actions, and vulnerabilities [1, 2]. In light of the significant role that technology plays in our lives, it is vital to better understand different accounts of the human and technology and investigate their influence on prevailing as well as emerging technologies. This paper takes a first step in providing a selective overview

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 281–290, 2020.
https://doi.org/10.1007/978-3-030-62803-1_22

on important views and arguments that can potentially influence the future relation of humans and technology.

Many approaches in IT ethics, design, engineering, and development stress that relevant stakeholders should be included in the design and development process of new technologies (e.g. [3]). This is based on the assumption that individuals differ in their needs, goals and general views regarding technology. However, the relation of human beings and technology is always bidirectional. Implicit in our expectations of technology is also an understanding of human beings and their general abilities, needs, and goals. In turn, "our notion of what we consider as being human is being increasingly shaped by the technologies that we use", as stated in the HCC14 Call for Papers. This interrelation has long been studied by interdisciplinary programs such as science and technology studies (STS) or philosophy of technology. A single publication cannot live up to the critical appraisal that contributions from these and related fields deserve. Still, only few works have tried to present a bird's-eye perspective on current debates, combining original philosophical positions with relevant technological developments and paradigms in the cognitive sciences (but see [4]). Also, most of the discourse has taken place in form of theoretical discussions with scarce empirical research investigating implicit theories and beliefs about humans and technology.

Informative findings come from a research group that empirically investigated the role of privacy in information system engineering. Carew and colleagues [5] found that important beliefs and attitudes in system development can be represented along two dimensions. The first dimension (labeled "humanist-existentialist") combines the belief in a free will and the responsibility for one's choices (existentialist aspects) with a concern for society and a broad humanitarian perspective that places the human at the center and shows concern for a person's values, rights, and dignity (humanist aspects). The second dimension ("technocentrist-industriofatalist") puts the focus on economic pursuits as well as on technical aspects such as functionality. Although this two-dimensional perspective has emerged in the information privacy context, it fits well with the wider diverging paradigms in current technology design approaches, which attribute different levels of appreciation to humans and technology.

On the one hand, we find approaches that focus on human needs, goals, and values. Don Norman, probably the best-known advocate of Human-Centered Design, proposes that any design should ensure "that people's needs are met, that the resulting product is understandable and usable, that it accomplishes the desired tasks, and that the experience of use is positive and enjoyable" [3, p. 44]. A human-centered approach takes into account human capabilities and limitations and designs enjoyable experiences instead of mere functionality. Related approaches strive to consider human goals (e.g. goal-directed design [6]) or values (e.g. value sensitive design [7]) in the design of technological systems or focus on including the stakeholders in the design process (e.g. through participatory design [8]). All these approaches share an empathic attitude towards human needs and capabilities that they seek to include in the design process.

On the other hand, the far-reaching possibilities of technological development have nurtured a technological optimism that focuses on capabilities and potentials on the side of technology. Silicon Valley is one of the places of origin of this techno-centered perspective, which often sees information as the driving resource [9]. With it comes an image of the human as a creature that can be fully understood and assessed through information technologies. On the consumer side, data-driven activities such as self-tracking and life-logging have inspired the "quantified self" movement that seeks to support individuals in gaining "self knowledge through numbers" [10]. Extreme techno-centered movements such as "transhumanism" [11] surpass IT applications and devices as they are currently in use – they seek to move beyond human limitations by interfering with the physiological and cognitive setup of human beings by technological means. To make this possible, many transhumanists support investments in technological innovations such as biotechnology, nanotechnology, and artificial intelligence – areas often subject to complex ethical and legal discussions.

These different approaches to technology show the important influence that assumptions about human nature and the role of technology can have on the design of technological artefacts. This paper sets out to explore a few selected dimensions to delineate views of human nature that potentially influence how we perceive, use, and design technology. Any concept of human nature typically captures basic assumptions about humans in one image, either to determine what is right (normative function) or to show what is characteristic of humans (descriptive function) [12]. Whether human nature itself is fixed or variable is itself under debate, which is why a concept of human nature typically reflects the assumptions, reflections and concerns of a certain era, discipline, or cultural group. In the following, the paper outlines views of humans and technology that are relevant in our era along three dimensions.

2 Core Questions on Human Nature Along Three Dimensions

The German psychologist Jochen Fahrenberg provides one of few interdisciplinary theoretical works on different views of human nature. Based on psychological, biological, intercultural, and religious perspectives, he defines a concept of the human as "the aggregate of assumptions and beliefs regarding *what the nature of human beings is*, how they should *live in their social and material environment* and which *values and goals* their lives should have" [13, p. 9, translated and highlighted by the author]. This definition suggests three core dimensions that can be explored with regard to how they play out in technology design and use: first, the natural constitution of human beings; second, the position of human beings within their environment; and third, human values. While major philosophical works have presented elaborate ideas on these topics, this paper wants to bring perspectives from different times, disciplines and discourses together in a critical but necessarily limited overview.

2.1 Where to Draw the Line? Body, Mind, and Technological Enhancements

One of the characteristics that traditionally differentiates human beings from other living creatures is their self-awareness and rich mental life. Rooted within this emphasis of the human mind is the question on what gives rise to mental experiences and the role of the body. Depending on how and where the line between the subjective mental life and the physical world is drawn, different concepts of human beings, their worth and their position in their environment can be derived. This can influence to what degree the use of technology is desirable, acceptable, morally acceptable or to be condemned, a question that forms the core of current debates on human enhancement. Especially the line between mind and body influences how emerging technologies such as bio-technologies and information technologies are regarded. Different positions in this debate regard technology and its potential effects on the human constitution either with optimism or pessimism: while some welcome and celebrate human enhancement by means of emerging technologies, others condemn it. A third position suggests an alternative view to human nature overall, and questions the strict separation of the human and technology. These views are briefly discussed in the following.

Proponents of the view that the body is a mere biological substrate of the mind often recognize the physical vulnerability of humans as a weakness that technology could help to overcome. For example, transhumanists (e.g. Nick Bostrom [11]) welcome emerging technologies and their application not only for treatment but also for enhancement purposes. They emphasize the potential of technology and see the ultimate goal of human development in the transcendence of human boundaries and limitations. On the other end, we find "bioconservatives" [4] and "infoconservatives" [1], who want to protect what naturally constitutes and characterizes humans and thus support "natural" strategies for human development and personal growth, such as education. Proponents (e.g. Francis Fukuyama [14]) criticize technological developments and innovations that alter human capabilities and characteristics and argue for the protection of the boundaries between humans and technology.

These two opposing positions on biotechnologies make apparent how different views of the human constitution influence the perception of technology, its design and use. But they both share the view that there is something fixed that constitutes human nature. Outside of this view is the existential-phenomenological account of Mark Coecklebergh, who argues that there is no clear boundary between the natural and technology and that the impact of technological enhancement on what some describe as "human nature" is more complex than assumed in the debate of transhumanists vs. bio/infoconservatives [1]. In his view, any enhancement will produce new vulnerability, which is why we need to discuss *how* human beings are influenced rather than debating whether this influence changes human *nature*. Tamar Sharon makes a similar conclusion, designating both positions as "humanism" as they emphasize the uniqueness of the human and thus separate the human from "the rest", be it animals, nature, or technology [4]. Similarly, other approaches (e.g. in the philosophy of technology [15]) emphasize that human nature is never purely "natural" but always includes "technological" aspects, which is why they argue for opening up the dichotomy of the human and technology and rethinking these categories and their relation.

2.2 Autonomy Reconsidered: Humans and Their Environment

In the history of philosophy, discussing the "state of nature" of humans became prominent during European colonialism when groups of people that had lived unreached by European civilization were discovered. Out of this experience arose the idea that humans generally have a natural state outside the influence of society. This helped to inspire liberal thinkers such as Thomas Hobbes and John Locke, according to whom humans are isolated individuals who exist completely independent from one another and from nature, i.e. in radical autonomy. This liberal view still influences much of our Western thinking and with it our views of technology. In light of liberalism, technology increases human freedom by extending human capabilities and compensating inequalities and supports humans in manipulating and controlling their natural environment. This idea is largely owed to Francis Bacon and has taken on new dimensions with the technologies available today [16].

Liberalism is often criticized for forgetting that humans have developed and evolved with their natural environment and are embedded in social groups (e.g. family, friends, colleagues, society, community, state). From an alternative perspective, humans have become deeply social beings who have learned to appreciate support and show solidarity because of their dependence on members of the community [17]; they are *situated* and *embedded* and thus perceive and act *in the world* [18]. Thus, social and cultural aspects do not only form the human understanding of morals, norms, habits, and goals, but also human cognitive abilities.

Phenomenological accounts that stress the relation between human beings and their environment have seen an increase in popularity and have influenced current thinking (e.g. [4, 18]). However, the concept of human autonomy and with it the assumed independence from the social and natural environment still exists in many disciplines. Especially economics at large is still stuck with the image of the human as an autonomous, calculating individual, the "homo oeconomicus" [19]. Thus, the assumptions of basic economic models stand in contrast with the concepts that stress the importance of the social and natural environment for human beings. This, in turn, influences how we design and envision future technologies, seeking to further extend our control over ourselves and our environment, both social and natural. If we take our situatedness and embeddedness seriously, we need to consider the wider effects that such technologies could cause with regard to our cognitive and social capabilities.

2.3 Discovering Human Values in Technology

Values represent things that are important to people in their lives and thus represent an important dimension of human beings. As described above, several design approaches try to consider values, goals, or needs for the design of new technologies. While some research has suggested that values are distinct from goals [20], the widely used conceptualization of values as "desirable transsituational goals, varying in importance, that serve as guiding principles in the life of a person or other social entity" [21, p. 21] stresses their commonalities.

Still, values differ from needs and goals: Values go beyond human *needs* because they represent what matters to humans, what they value, strive for and seek to protect, and can include moral considerations. Furthermore, they go beyond human *goals* because they stress the relationship and interaction between the human as valuing organism and the world as value-bearing environment [22]. Recent research has shown that laws, principles, declared goals and obligations that have been put forward in various fields (including design, engineering, law, psychology, philosophy, and ethics) can be brought together in terms of values [23]. The resulting taxonomy presents values along five dimensions of sustainability (individual, social, technical, economic, and environmental), which provide a good starting point for discussing views of the human and technology. In the following, individual values, social values, environmental values, technical values, and economic values are briefly discussed with regard to their relevance for technology design and use.

Individual values cover basic and higher-order preconditions for a good life. They thus comprise health, safety, freedom, and property as well as education, knowledge, and pleasure [23]. In psychology, these values are usually framed as needs and several theories have addressed how they drive human behavior. One of the best-known models is Maslow's theory of needs [24], which proposes a hierarchical structure that covers physiological needs, safety needs, love needs, esteem needs, and the need for self-actualization at the very top. While user-centered design approaches typically seek to account for individual needs, they often focus on mere ease of use and ergonomic factors. These developments have led to the human-centered design approaches addressed in the beginning, which seek to exchange this reductionist view of "the user" with a more holistic view of human needs, values, and goals. Furthermore, a focus on individual values and virtues brings in an additional moral dimension in the consideration of the human-technology relation [25].

Different views of the social human condition result in a different appreciation of social values. For example, the values of freedom, safety, and justice can stand in contrast with the values of equality, dignity, trust, and community. While human rights aim to protect every human's dignity independently of the individual's capacities, abilities, etc., liberal thinkers typically stress human freedom. Digital products and services cater to social values in complex ways, e.g. by making digital communication easier and faster while at the same time reducing personal interactions and meetings. In contrast, a stronger community is being called for as a powerful way towards a more democratic and socially just political [26] and economic life [19]. Thus, an adequate understanding of humans as social beings is needed to translate between what is needed for a functioning society and the products and services that are being advertised and put on the market.

Related to this is the perception of environmental values such as biological diversity and a respect for nature [23]. For most of our evolutionary history, the natural environment determined the conditions of human life, influencing the genetic constitution of human beings and thus human capacities and limitations. In modernity, a view of human superiority over nature emerged. Influenced by scholars from the liberal tradition such as Francis Bacon and Thomas Hobbes, humans were described as being in strong control over nature, including their own nature. Today, humans are met with the challenge to maintain an environment that allows natural life, including humans, animals, and plants. We have entered "the Anthropocene", the age where humans determine the fate of earth. Thus, any human-centered approach is being challenged by the need to take the natural environment into account [27]. In line with this is a view that attributes a unique stewardship role to humans, a position of responsibility in the world. No matter whether humans are seen in a superior position to their surrounding or not, the question is whether this superiority is accompanied by an attitude of power or an attitude of care, which results in a different appreciation of environmental values – and thus a different starting position for the design and development of future technologies. The current climate crisis and the multitude of technological devices being produced force us to consider any technology as an ecology that includes design, materials, humans and non-humans as much as politics and motivate environmentally friendly and sustainable technological innovations [27].

Values have long made their way into technology design and development (e.g. [3]). While usability, maintainability, and efficiency are usually being accounted for, values with social import such as information privacy have attracted wider attention only more recently. In Europe, this was accompanied by the application of the new General Data Protection Regulation in 2018. The example of information privacy also serves as an example that can very well illustrate how individuals' views and beliefs impact their behavior. Research has shown that engineers' pessimist beliefs about the feasibility of information privacy come with a decreased motivation to implement privacy mechanisms [28, 29]. Carew and colleagues express even wider concerns by showing that in system development, interests in human, social, and moral issues are diametral to interests in technical issues and functionality [30]. Thus, a sustainable model of technology production, use, and interaction needs to conciliate technical values with individual, social, and environmental values.

Economics is one of few disciplines that has designed its own view of human nature. The image of "homo oeconomicus" that pictures the human as "standing alone, money in hand, calculator in head, and ego in heart" [25, p. 96] still prevails in many economic textbooks, theories and policy making – and has resulted in specific economic values. The strive to maximize utility, perfect knowledge, and foresight were added to the idea of human beings as rational, autonomous individuals, to create an image of the ideal consumer – an initially descriptive model that soon became prescriptive. While this image was later criticized by important figures in economics, it has contributed to the development of current economic values that focus on profit maximization through efficiency, productivity, and innovation. In her book "Doughnut Economics" [19], Kate Raworth emphasizes the need for a new image that is fit for the twenty-first century as well as new economic values that seek to re-situate human well-being within the social and ecological boundaries determined by human needs and the natural environment's capacities.

3 Discussion and Conclusions

The natural and technological environment offers a variety of ways for human beings to live, grow and develop. This paper has delineated discourses on human nature and technology that seem relevant for further developments in technology-driven societies. Based on an interdisciplinary definition of human nature, it discusses selected positions with regard to the natural constitution of human beings, the position of human beings within their environment, and human values. Interestingly, many debates start from the assumption that the nature of human beings can be captured in specific images that delimitate human nature from the natural and material environment.

These positions are challenged by accounts that emphasize the interrelatedness of mind and body, humans and the natural environment, and humans and technology. For example, a recent paradigm in cognitive science stresses that cognition is embodied, embedded, extended, and enactive [31], or in other terms, "ecological" [18]. Furthermore, both transhumanists as well as bio- and infoconservatives are being challenged by accounts that do not see a clear boundary between the human and technology. Tamar Sharon [4] proposes her account of “mediated posthumanism” that stresses the interrelatedness of humans and technology. Mark Coeckelbergh [1] makes a similar argument by pointing out that there is no fixed notion of “human nature” but that technology as much as the social and historical context has always shaped human beings. Both argue that an ideological view of the human in terms of “human nature” or “humanism” needs to be overcome in order to make room for a view of humans and technology as inherently interrelated categories.

The argument that opening up the dichotomy of human and technology allows radical rethinking and restructuring of categories is not new, though. It is what Donna Haraway [32] tried to illustrate with her idea of the “cyborg”, which challenges traditional categories such as the human-technology dichotomy. Similarly, philosophers of technology such as Don Ihde or Peter-Paul Verbeek have emphasized the mediating role of technology for how humans relate to the world and to each other: Verbeek argues that technologies shape the world we live in and mediate our practices and experiences [15]. Similarly, accounts that see the human as an isolated, autonomous individual are being challenged by phenomenological accounts that stress the embeddedness of human beings within their environment. Putting the split between mind and body, human and technology, and human and world into perspective can thus be critical in reconsidering the role of technology for human beings.

The discussion of human values shows that several concepts (needs, goals, and values) seek to capture what is important to human beings in order to include it in the design of technology. The concrete design and use of new technologies depend on the relevance attributed to specific individual, social, environmental, technical, and economic factors. Any technology is shaped by human goals and values, but also shape them in turn. More research that helps to structure human needs, goals, and values by addressing their commonalities and conceptual differences could offer helpful contributions in this respect. Technology design that acknowledges the human side needs to be flexible to take into account the many regards in which humans and technology influence each other. Human-centered design approaches have proposed ways to do this, but might be able to profit from considerations on how to place their approach “within the doughnut” [19, 27], that is, to consider both human and environmental restrictions.

The debates on what constitutes the human naturally, how human beings are related to their environment, and what implications this has for their actions and responsibility as well as for the role of technology, clearly show that implicit theories and assumptions can influence how technology is being designed and used. An understanding of implicit beliefs regarding humans and technology can inform research and practice in the fields of technology ethics, design, and engineering and can open up space for a positive reconceptualization of both what it means to be human and the role of technology. While this paper can only offer a preliminary analysis of positions and arguments, further research – both theoretical and empirical – could help to further delineate different accounts and discuss their influence on current and future technological developments and the human-technology relation.

Acknowledgements. I want to thank S. Spiekermann for supporting me in exploring this topic and for many inspiring discussions. I am also grateful to J. Hoff who has provided helpful literature input as well as P. Hampson who has offered guidance through the vast literature available on this topic. I wrote this paper as recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute of Information Systems and Society, Vienna University of Economics and Business.

References

1. Coeckelbergh, M.: Human Being@ risk: enhancement, technology, and the evaluation of vulnerability transformations. Springer, Dordrecht (2013). https://doi.org/10.1007/978-94-007-6025-7
2. Verbeek, P.-P.: Morality in design: design ethics and the morality of technological artifacts. In: Vermaas, E., Kroes, P., Andrew Light, S.A.M. (ed.) Philosophy and Design: From Engineering to Architecture, pp. 91–103. Springer, Dordrecht (2008)
3. Friedman, B., Kahn Jr, P.H.: Human values, ethics, and design (2003)
4. Sharon, T.: Human Nature in an Age of Biotechnology: The Case for Mediated Posthumanism. Springer, Dordrecht (2014). https://doi.org/10.1007/978-94-007-7554-1
5. Norman, D.A.: The Design of Everyday Things. Basic Books, New York (2013)
6. Cooper, A., Reimann, R., Cronin, D., Noessel, C.: About Face: The Essentials of Interaction Design. Wiley, Indianapolis (2014)
7. Friedman, B., Kahn Jr., P.H., Borning, A.: Value sensitive design and information systems. In: Zhang, P., Galletta, D. (eds.) Human-Computer Interaction and Management Information Systems: Foundations, pp. 348–372. M.E.Sharpe, Armonk (2006)
8. Kuhn, S., Muller, M.J.: Participatory design. Commun. ACM **36**, 24–28 (1993). https://doi.org/10.1145/153571.255960
9. Duff, A.S.: Rating the revolution: silicon valley in normative perspective. Inf. Commun. Soc. **19**, 1605–1621 (2016). https://doi.org/10.1080/1369118X.2016.1142594
10. Quantified self: Self knowledge through numbers. https://quantifiedself.com/about/what-is-quantified-self/
11. Bostrom, N.: The transhumanist FAQ. In: Kaplan, D.M. (ed.) Readings in the Philosophy of Technology, pp. 345–360. Rowman & Littlefield Publishers, Lanham (2009)
12. Düwell, M.: Menschenbilder und Anthropologie in der Bioethik [Image of man and bioethics]. Ethik der Medizin. **23**, 25–33 (2011). https://doi.org/10.1007/s00481-010-0109-5
13. Fahrenberg, J.: Menschenbilder: Psychologische, biologische interkulturelle und religiöse Ansichten [Concepts of man ("Menschenbilder", assumptions about human nature): Psychological, biological, cross-cultural & religious perspectives]. Universität Freiburg: Institut für Psychologie (2007)
14. Fukuyama, F.: Our Posthuman Future: Consequences of the Biotechnology Revolution. Farrar, Straus and Giroux, New York (2002)
15. Verbeek, P.-P.: What Things Do: Philosophical Reflections on Technology, Agency, and Design. The Pennsylvania State University Press, Pennsylvania (2005)
16. Deneen, P.J.: Why Liberalism Failed. University Press, Yale (2018)
17. Tomasello, M.: A Natural History of Human Thinking. Harvard University Press, Cambridge (2014)
18. Fuchs, T.: Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind. Oxford University Press (2018)

19. Raworth, K.: Doughnut Economics: Seven Ways to Think like 21st-century Economist. Random House Business, London (2017)
20. Jolibert, A., Baumgartner, G.: Values, motivations, and personal goals: revisited. Psychol. Mark. **14**, 675–688 (1997). https://doi.org/10.1002/(SICI)1520-6793(199710)14:7%3c675:AID-MAR3%3e3.0.CO;2-D
21. Schwartz, S.H.: Are there universal aspects in the structure and contents of human values? J. Soc. Issues **50**, 19–45 (1994). https://doi.org/10.1111/j.1540-4560.1994.tb01196.x
22. Fuchs, T.: Values as relational phenomena: a sketch of an enactive theory of value. In: Mühling, M., Gilland, D.A., Förster, Y. (eds.) Perceiving Truth and Value: Interdisciplinary Discussions on Perception as the Foundation of Ethics, pp. 23–42. Vandenhoeck & Ruprecht, Göttingen (2020)
23. Winkler, T., Spiekermann, S.: Human values as the basis for sustainable information system design. IEEE Technol. Soc. Mag. **38**, 34–43 (2019). https://doi.org/10.1109/MTS.2019.2930268
24. Maslow, A.H.: A theory of human motivation. Psychol. Rev. **50**, 370–396 (1943). https://doi.org/10.1037/h0054346
25. Vallor, S.: Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press, New York (2016)
26. Monbiot, G.: Out of the Wreckage: A New Politics for an Age of Crisis. Verso, London & New York (2017)
27. van der Velden, M.: ICT and sustainability: looking beyond the anthropocene. In: Kreps, D., Ess, C., Leenen, L., Kimppa, K. (eds.) HCC13 2018. IAICT, vol. 537, pp. 166–180. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99605-9_12
28. Bednar, K., Spiekermann, S., Langheinrich, M.: Engineering Privacy by Design: are engineers ready to live up to the challenge? Inf. Soc. **35**, 122–142 (2019). https://doi.org/10.1080/01972243.2019.1583296
29. Spiekermann, S., Korunovska, J., Langheinrich, M.: Inside the organization: why privacy and security engineering is a challenge for engineers. Proc. IEEE **107**, 600–615 (2018). https://doi.org/10.1109/JPROC.2018.2866769
30. Carew, P.J., Stapleton, L.: Towards empathy: a human-centred analysis of rationality, ethics and praxis in systems development. AI Soc. **29**(2), 149–166 (2013). https://doi.org/10.1007/s00146-013-0472-0
31. Newen, A., de Bruin, L., Gallagher, S. (eds.): The Oxford Handbook of 4E Cognition. Oxford University Press, Oxford (2018)
32. Haraway, D.J.: A cyborg manifesto: science, technology, and socialist-feminism in the late twentieth century. In: Haraway, D. (ed.) Simians, Cyborgs and Women: The Reinvention of Nature, pp. 149–181. Routledge, New York (1991)

Towards a Typology of Intentionally Inaccurate Representations of Reality in Media Content

Matthew J. Davis(✉) and Per Fors

Uppsala University, Uppsala, Sweden
{matthew.davis,per.fors}@angstrom.uu.se

Abstract. In this paper, we take a look at three concepts frequently discussed in relation to the spread of misinformation and propaganda online; fake news, deepfakes and cheapfakes. We have mainly two problems with how these three phenomena are conceptualized. First of all, while they are often discussed in relation to each other, it is often not clear what these concepts are examples of. It is sometimes argued that all of them are examples of misleading content online. This is quite a one-sided picture, as it excludes the vast amount of content online, namely when these techniques are used for memes, satire and parody, which is part of the foundation of today's online culture. Second of all, because of this conceptual confusion, much research and practice is focusing on how to prevent and detect audiovisual media content that has been tampered with, either manually or through the use of AI. This has recently led to a ban on deepfaked content on Facebook. However, we argue that this does not address problems related to the spread of misinformation. Instead of targeting the source of the problem, such initiatives merely target one of its symptoms. The main contribution of this paper is a typology of what we term Intentionally Inaccurate Representations of Reality (IIRR) in media content. In contrast to deepfakes, cheapfakes and fake news – terms with mainly negative connotations – this term emphasizes both sides; the creative and fun, and the malicious use of AI and non-AI powered editing techniques.

Keywords: Fake news · Deepfakes · Cheapfakes · Propoganda · Disinformation · Misinformation · Typology

1 Introduction

Let us begin this paper by noting that the authors are two prolific researchers, whose tenacity for objective truth-finding in a post-truth society has won them numerous journalistic integrity accolades across the globe. Their dedication to this cause is unparalleled, and equal only to their endless pursuit of knowledge. Or is it? As a reader, does this make the premise of the paper more convincing? In any case, how does one critique such claims, if, for example, this paper is to be published in a reputable journal? Usually, if one has not read any of the authors' previous work, critical minds will seek some measure of verification, rather than take these claims at face value. Perhaps these awards are listed on their respective LinkedIn profiles, and a quick

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 291–304, 2020.
https://doi.org/10.1007/978-3-030-62803-1_23

internet search of the names of these awards reveals a professional looking website with social media posts and images relating to an award ceremony. Depending on one's level of cynicism and appetite for how deep the rabbit hole goes, a few layers of forgery might be enough to convince a skeptical reader that the authors are indeed trustworthy individuals, and that the academic contribution of this paper is worthy of a high impact factor. No harm done … right?

The erosion of trust in societies institutions is a present and widely discussed topic among today's scholars, not least because of the decay of appreciation for truth among the general population [1]. Whether by design or not, Latour's post-truth tools for querying the origins of scientific fact have been weaponized by a new wave of alt-right propagandists with alarmingly effective results [2]. This Liar's Dividend [1] has forced society to reckon with itself, and polarized politics to the point that we can no-longer believe what we see, unless it is first hand [3].

Let us contextualize this issue further. A unique trait among humankind, is the ability to communicate ideas across generations through written, and more recently, audiovisual (AV) media. For years, most of our information about current events came through carefully curated channels; newspaper articles, books, televised reporting, and with it, society had to exert a certain amount of trust that the information they were receiving was truthful since there were very few ways of learning about events occurring outside our personal bubbles of experience. Journalists had a measure of integrity and respect for the job, and access to printed media was controlled through education, accreditation and the fact that mass production of printed text was costly. Whilst this did not hinder the spread of propaganda, it at least limited the power to shape civil discourse to the establishment, whose strive for stability meant that unorthodox views were kept to the niches of society. There are many examples in modern times however, of unscrupulous actors manipulating various media in order to further their own agendas. From the Iraq war of 2003 [4] to the UK's most recent general elections and referenda [5], those with the keys to shape public discourse tend to wield the most political power.

With the advent of digital media and the internet, the limitations on how information is captured, shared and discussed around the world, changed. This decentralization of communicative channels has given the illusion of complete information freedom, and for the time being, it is much harder for media tycoons and political establishments to paint a subjective narrative since everyone now armed with a smartphone can share their own version of events, as they occur. Once again, the balance of power between institutional elites and the people is changing, but with it, the lines between objective truth and subjective narrative are becoming increasingly blurred. Technology, often thought of as a panacea for many of today's problems, can also exacerbate such issues.

Since the internet globalized our digital media, everyone with access to a computer was able to have a wide-reaching platform for their message, without the same level of scrutiny usually applied to traditional broadcasts. Initially, producers of audiovisual media content required some capital investment; lots of expensive hardware, human resources, and years of specialist education in order to create a masterful production of coercive storytelling. Now, everyone has the ability to create their own online "news" channel, promoting all manner of more or less informative or sensational opinion pieces. A few years ago, one could differentiate between amateur influencers, and professionally produced content based on the special effects, or the premium feel. Yet with a little experience, access to low-cost technology has improved the user experience far beyond traditional media. Whilst this isn't necessarily a problem when it comes to pure entertainment, combine this with a business model that relies on attracting views and a general public's decreasing attention span, and you have a cultural crisis in the making, leading to the loss of truth as an exchange value [6]. Or as Postman [7] might argue, a swelling preference for entertainment over substance. In a time where the freemium (most content for free, some premium) business model effectively removes the funding for in-depth and thus costly investigations, salvaging journalistic integrity is also difficult [8]. How then, do uncritical minds differentiate between well produced content with verified "facts" from qualified professionals whose "evidence" challenges their beliefs, from the uninformed and opinionated ideologically driven views of a tech-savvy troll, whose entertaining tirades conform to and enhance the pre-existing biases of the target demographic? Addressing such a question goes beyond the scope of this paper, yet its importance should be highlighted since democracy as we know it, is consequently at risk.

2 Fake News

Donald Trump is often quoted when discussing fake news rhetoric, since he has successfully used the tactic to take advantage of this 'liar's dividend' in public political discourse. Whilst several scholars disentangle the semantic definition of fake news [9], the term is usually ascribed to unfavorable media, in order to confuse or polarize the debate in such a way as to remove any credibility from an authentic source; the general public, who are now a hyper-cynical audience, begin to question the legitimacy of the mainstream media. Far from being a novel strategy, sowing the seeds of doubt, and reaping the benefits of a skeptical public has been the go-to tactic for the rise of many successful politicians. However, it would be complacent to suggest that the most current candidates' rise to power is simply a "blip" in our collective sanity, attributed to an imbalance of knowledge and rational thinking among the general public. As argued above, there is a cultural crisis occurring on the back of a wave of new media, which certain political groups have exploited rather successfully.

The idea then, that a healthy democracy should allow deliberately deceptive speech under a banner of freedom, is a contradictory one, with roots in Popper's paradox of tolerance. In a "marketplace of ideas", the practice undermines, rather than enhances, the pursuit of truth since "people are more likely to be influenced by the first version of events that they hear, and are increasingly likely to remember falsehoods the more they are exposed to them" [10, 54]. Now more than ever, wider society needs to be vaccinated against the fake news phenomenon [11], especially since the tools and techniques used by those who would subvert the democratic process are using increasingly sophisticated technologies. If we value our democracy, and strive to build a just society based on civil liberty, we must elucidate the different types of dis- and mis-information so that, despite the interests of unscrupulous actors, even the most gullible person is equipped to treat "fake news" as an empty signifier, and reject it. This type of vaccination is categorized as an educational remedy; it relies on changing the general public from cynical, into critical, which is easier said than done. The Open Society Institute highlight in their report that nations with a higher level of education among the general public, tend to be more resilient to fake news [12]. Before we look at solutions however, let us discuss the different types of media that are designed to misrepresent reality.

3 Cheapfakes (or Shallowfakes)

We begin therefore with the manipulation of still images; a longstanding practice that became more readily accessible to the general public in the early 2000s with the rise of easy-to-use image editing software. When a picture is manipulated in this way, it is often called "shopped", with a reference to Adobe Photoshop, the most popular photo editing software. Since the 2000s, shopped images have cemented themselves as an important part of online culture, as can be found all over forums such as Reddit and 4chan, but also on social media platforms such as Facebook. Normally, original content is produced for Reddit or 4chan, and once such content becomes a meme[1], it becomes widespread (goes viral) on other more mainstream social networks.

While image manipulation is often used to entertain a certain online audience, such techniques can also be used with malicious intent. McGlynn et al. [13] discuss the use of shopping for pornographic purposes which they term "sexual photoshopping". According to DeKeseredy and Schwartz [14], such malicious intent of the software "normalize misogyny, hurtful sexuality, racism, and … seeking revenge on female ex-partners". The important point about sexual photoshopping is that it can be used for mainly two purposes; for revenge or for entertainment. The first category of revenge pornography is mainly practiced by men to humiliate or shame previous partners, though certain demographic groups are more likely to experience it [56–59]. The second type is usually shared for "entertainment purposes" on different subforums. For example, a more recent sexual photoshopping technique is called "bubbling" which, whilst proponents may argue actually covers up more of an original image, ultimately relies on quirks of the human brain to make a semi-clothed person appear naked [15]. It can then also be classified as entertainment, but still have a malicious intent, especially if the content is used to humiliate or degrade.

Other types of non-AI powered techniques are also used for this purpose. For example, in 2016 the legal counsel of the president Duerte used a pornographic video showing a woman impersonating Leila De Lima, a political opponent critical of the president's authoritarian rule in the Philippines. Similar things happened in India in 2002, when journalist Rana Aryyub was targeted with a pornographic video in order to ruin her credibility [16]. However, while this content requires some technical knowledge to produce, there are far simpler ways of ruining someone's reputation using unedited "evidence" of the person in an exposed position, namely to recontextualize an

[1] A meme is a visual or textual expression of an idea, behaviour or style that spreads and evolves online by means of imitation and is not to be confused with "viral content", which refers to anything – a text, an image or a video – that is frequently shared online.

already existing video clip or picture. Usually, a picture or video clip is posted on social media showing for example a politician saying something, with a made-up text of what the video or picture is describing. While such techniques are commonly used by propaganda outlets to mislead and manipulate, it is more often used for entertainment purposes. In such cases however, it is often clear that the context of the picture or video has been altered by the original poster.

Another low-tech technique that is often used to manipulate video clips is splicing. Splicing means that segments of a video are put together in certain way, so that a - often political - message is communicated. In 2018, for example, a false BBC report started circulating on WhatsApp, showing how a nuclear war was developing between NATO and Russia, claiming that Russia had used tactical nuclear weapons against the UK [17]. The clip also shows the royal family evacuating Buckingham Palace. The technique is also used to cheat in different online gaming competitions, most notably speedrunning[2]. Apart from splicing, there are other low-tech video editing techniques that can be used to mislead a particular audience. The most well-known is probably the “drunk Pelosi” video. This video, which was shared on Twitter also by US president Donald Trump, showed Nancy Pelosi in a debate, seemingly intoxicated. In fact, the video had been slowed down to 75% of its original speed [18].

What these techniques demonstrate, is that there are already many ways in which people attempt to control public discourse to suit their own agenda. This is not a new phenomenon; most of the general public has been primed towards them for many generations and is largely able to critique this type of content effectively. Consequently, a relatively skeptical audience exists which is somewhat immune to these types of propaganda. A problem arises however, when new technologies are developed which challenge our perceptions of reality and lead us to question whether what we see, is actually what we get.

4 Deepfakes

The term deepfake was coined after a Reddit user called Deepfakes posted several pornographic videos on a subreddit in 2017. These videos, while professionally produced, seemed to feature well-known female celebrities as pornstars. The term deepfake is a combination of the terms “deep learning” and “fake” and refers to Audio/Visual (AV) media content that has been edited using AI powered software in order to produce falsified but authentic-looking results [55]. The tools used to create deepfakes were published shortly after the first pornographic movies, and while Reddit quickly banned deepfaked pornography from their site, users of these tools found other forums to share their creations and improve their tools. The tools often use the Google Image search function, look through social media sites and replace faces in videos in a convincing manner, so long as the source material is of high enough quality. After the

[2] This refers to a certain way of completing a video game or selected parts of it as quick as possible. In order to prove that the run has been completed in a certain time, the runner needs to provide video evidence.

first machine learning process, the programs often require little to no human supervision, and the algorithms improve the process more or less autonomously [19].

The technology used in these tools is a machine learning technique called GAN, which stands for generative adversarial network, and has utility not only for AV media editing, but also in cutting edge medicine and computer science research [20], hence, an outright ban on developing such technology would be problematic for numerous reasons. In any case, the technology required to create convincing deepfakes, combined with fake voices, currently requires a large seed dataset. Fried et al. [21] demonstrate a relatively straight forward process for achieving lifelike reconstructions, though it requires a lot of time for processing and good quality media. For deepfaking the everyday person, the specificity of clear, well lit images or video, and specific spoken phrases may be quite difficult to gather and compile. For public figures who are always in the spotlight however, this can be quite easy to acquire. Indeed, with ever increasing capabilities of recording and processing technology, the barrier to entry for public use of these deepfake algorithms will continue to diminish. We therefore envisage a point in the very near future where, like much of the printed news media, one can no longer trust that visual media is representative of real life either. The impact of such a step change to our society should not be underestimated, since the institutional consequences are likely to be far reaching.

According to Zannettou et al. [22], there are a number of actors associated with the production and distribution of deepfaked content. These range from governments, political activists, criminals, malevolent individuals (paid and unpaid trolls), conspiracy theorists and social media bots, pushing the content towards certain groups of people susceptible to whatever message is being communicated. Westerlund [23] presents a similar list of the producers of this kind of content: communities of deepfake hobbyists, political players such as foreign governments, and various activists, other malevolent actors such as fraudsters, and legitimate actors such as television companies. The question of ‘why’ such agents feel compelled to produce deep faked content is an interesting one, if a little loaded at present. Empirically however, it is also clear that the technology itself is not entirely neutral, since the tools used to generate deepfaked content were popularized in male-dominated online communities and are mainly used for pornographic purposes.

Consequently, it is important to emphasize that, despite the growing amount of deepfaked content online, deepfakes have yet to be used for purposes of political propaganda or in a malevolent misleading way with any substantial impact. According to Deeptrace – a company that uses different techniques to detect and monitor deepfaked content – pornography, and mostly celebrity pornography, makes up approximately 96 percent of the deepfaked content [24]. While this is by no means positive, we can safely say that the average user and distributor of deepfaked content is not a fraudster or a politician with malicious intent, but for now, a tech-savvy pervert [25] or relatively harmless troll. Whilst the damaging effects of this type of content are contingent on various social stigmas surrounding pornography, this does not however diminish the risk of the technology being abused for political means once they are harder to detect. There are several well-made deepfakes which use the likeness of former US president Barack Obama [26], UK prime minister Boris Johnson, and former UK leader of the opposition Jeremy Corbyn [27], to demonstrate the alternative applications of such technology outside of pornography.

Unsurprisingly then, it is commonly assumed that “deepfakes are a major threat to society, the political system and businesses” [23]. However, while some papers mention that there are positive uses of this technology, most contributions express fear or worry about the technology itself rather than the problem of misinformation, which is described as a “growing” [21], very “real threat” [28, 29] to elections and society. Much research on the deepfake phenomenon therefore focuses on how to create technological tools for deepfake detection; indeed, most of the papers we found have this focus, and admittedly we have also considered this option.

5 A Society in Search of Solutions

Although it is a novel approach, we are not the first to consider using the blockchain to ensure the integrity and validity of video content. In a recent paper, Hasan and Salah [30] present a well thought out solution which uses the Ethereum blockchain and smart contract functionality in order to trace the history of digital content to its original creator source. For many unaware readers, the context of such a system requires some clarification. Most of the internet as we know it today is heavily centralised, and exists in a client-server relationship, with large companies such as Google, Facebook, and Amazon hosting most of the world’s data in their cloud-services warehouses. As with any centralised system architecture, there are inherent advantages in terms of speed of access, security, and scalability. However, there are also many weaknesses, including maintaining anonymity and data privacy; if you can own the server through a hack or some such vulnerability, you can own the data or even replace it with your own version. Consequently, there are many projects underway to develop a decentralised system architecture which relies on blockchain technology to store the data in a more distributed and secure manner. Ethereum is one such blockchain project, though others exist with similar smart contract functionality, and each has its own marketplace of users developing dApps (decentralised applications). Indeed, there is an ongoing ideological debate occurring between these projects about just how decentralised they need to be, and it is likely that, as with most new technological standards, only a few will reach the critical mass of mainstream adoption required to survive.

Under Hasan and Salah’s framework, all media content would be registered on a “decentralized, content-addressable, peer to peer file system”, such as the IPFS (InterPlanetary File System). Content creators publish an original video, and subsequent edits require the permission from the original creator and are linked to the original video as records on the immutable ethereum blockchain. For the main issue of this paper however, this is problematic for several reasons. Firstly, such a system requires many content creators to switch to a new type of decentralised internet which is currently not in a user-friendly format. Secondly, anyone can still copy a published video using some basic screen capture software, and then modify and upload an ‘original’ copy which can then be seen as the original source of this new file. Furthermore, their “underlying principle of solving the deepfake problem simply relies on providing undisputed traceability to the original source”, which does not really address the creation or distribution of disinformation. If the original publisher of a video uses their own captured and modified footage, how are we to know the truth of the matter? Perhaps this is not the point however, since in many ways, the solution is a more technically complex version of Twitter's "Verified" status, built for a decentralized system architecture of peer-to-peer content creation. The authors' main stated goal is to offer a workable solution for a blockchain based "Proof-of-Authenticity", and they should be commended for doing so. Indeed their solution may prove highly useful for copyright control and distribution of royalties in the near future. If the technological trend is moving towards a decentralized internet, then such a framework can succeed perhaps at least in helping users to determine whether their digital content comes from a trusted and reputable source.

In any case, the recent years' "truth decay" is certainly worrying; just as Trump uses the term "fake news" to deflect legitimate criticism away from his actions, one must now question whether we can trust what we see, since it is also possible that media content has been tampered with in some way – manually or with the help of AI. What can be done to defend against the potential negative impact of this disruptive technology on society? We do not give regulatory solutions much weight here, simply because regulation of open source technology (which also has positive uses) is impractical in any case. Criminalization of the use cases of digital technology is a separate issue which we do believe in, and has had limited efficacy against online copyright piracy, for example (though it is arguable that with the advent of cheap streaming services, the market demand for pirated material diminished). Whilst this may curb the production and spread of "revenge porn" among the general public, it does nothing to address the societal impact of misinformation, which is highly effective on an uncritical, or overly skeptical audience. Additionally, we are hesitant to suggest that regulation would inhibit governments and politicians from abusing technology via 3rd parties, since we have seen they are already extremely comfortable bending and manufacturing their version of the truth.

Some companies are attempting to combine approaches to address the issue. Facebook, who recently banned some doctored video content from their site, states that AI-powered manipulated content was rare to begin with, but that deepfaked content could potentially present a "significant challenge for our industry and society as their use increases" [31]. Banning deepfaked content certainly sounds like great news for anyone concerned about the spread of disinformation, however, Facebook is – as noted by Drew Harwell on Twitter [32] – merely targeting a certain video-editing technique and not spread of disinformation per se. For example, non-AI powered editing such as the "drunk Pelosi" video mentioned above can still be shared on the site, while "dank memes" such as the deepfaked Nicholas Cage videos can potentially be removed. Furthermore, as most deepfaked content can already be classified as pornography, Facebook already has a policy against the vast majority of malicious deepfaked content.

Given the strong academic and civil discourse expressing worry and fear for the technology itself and not mainly in which way it is used, it is time to put the technology in a larger context which highlights the fact that media content is manipulated for many different reasons, and that manipulation of media content is a big part of online culture. The same thing goes for satire and parody articles which are sometimes casually

Table 1. A typology of IIRR in media content

Type	Media format	Purpose	Distributed through	Uses AI	Potential of malicious use	Also known as
Fake news	Textual, AV	Mislead	Traditional media, social media, blogs, word of mouth (wom)	No	High	“Fake news”
Polarized content/biased content/misreporting/selective reporting	Textual, AV	Mislead, create opinion	Traditional media, social media, blogs, wom	No	Medium	“Fake news”
Parody	Textual, AV	Humour, memes	Traditional media, social media, blogs	Yes/No	Low	“Fake news”
Satire	Textual, AV	Humour, memes, propaganda	Traditional media, social media, blogs	Yes/No	Low	“Fake news”
Citizen journalism	Textual, AV	Mislead, create opinion	Social media, blogs, wom	No	High	“Fake news”
Clickbait	Textual, AV	Mislead, generate clicks	Social media	No	Medium	“Fake news”/ “Clickbait”
Conspiracy theory/pseudoscience	Textual, AV	Mislead, generate opinion, generate clicks	Social media, blogs, wom	No	High	“Fake news”
Image enhancement	Image	Improve quality of images	Traditional media, social media	Yes/No	Low	N/A
Automated photo editing	Image	Improve quality of images	Traditional media, social media	Yes	Low	N/A
Splicing	AV	Humour, memes, mislead	Social media, blogs	No	High	“Cheapfake”/ “Shallowfake”
Photo editing (Photoshopping)	Image	Humour, sexual photoshopping, memes, generate opinion, create opinion, generate clicks	Traditional media, social media, blogs	No	N/A	“Cheapfake”/ “Shallowfake”/ “Shopped”
Face Swapping/Face-morphing/Full-body puppetry	Video, Image	Humour, memes, propaganda, sexual photoshopping, generate opinion	Social media	Yes	Low	“Deepfake”
Lip syncing and voice synthesis	AV	Humour, memes, propaganda	Social media, forums, blogs	Yes	N/A	“Deepfake”
Re-contextualization	AV	Humour, memes, propaganda	Social media, blogs	No	High	“Cheapfake”/ “Shallowfake”
Text-to-speech/voice-swap	Audio	Humour, mislead, propaganda	Social media, blogs	Yes/No	Medium	N/A
Astroturfing	Textual, AV	Mislead, propaganda, generate opinion	Social media, blogs, traditional media	Yes/No	High	Covert lobbying, fake grassroots

described as fake news, although they do not seek to mislead but to entertain (see e.g. [9, 33]). More importantly, while being used for many different purposes and with very different motives, fake news, cheap fakes and deepfakes have more in common than the research community has yet realised. The aspects mentioned above are highlighted in our typology, where we collectively refer to different types of fakes as Intentionally Inaccurate Representations of Reality (IIRR) in media content (Table 1).

The IIRR typology has been put together through a literature review of papers that have tried to in some way categorize either deepfaked or cheapfaked content (e.g. [16, 34]) or fake news (e.g. [31–34]). The level of "potential of malicious use" has been evaluated based on our own perception of the type of IIRR content combined with other researchers' perceptions [35, 36, 38]. The typology itself is a work-in-progress and therefore we invite other researchers to continue this work in future research contributions in order to refine and perfect the typology. Such a typology will be useful in categorizing and identifying misuses of future technological improvements, and perhaps for building defensive strategies, whether through regulatory, educational, or technological means.

It is important to note that despite the fact that technology itself is not neutral, its potential is not entirely negative. As Silbey and Hartzog [8] suggest, perhaps the deepfake phenomenon is precisely what society needs to properly reflect and deal with the breakdown of trust in political institutions. By challenging the very notion of 'seeing is believing' innate to our concept of trust, established power structures will no longer be able to capitalize on an uncritical audience. That being said, it will be difficult to muster the political will to modify these institutions sufficiently if they continue to benefit from a disenfranchised general public. We suggest more focus on building genuine grassroots organizations which challenge the meta-narrative of deepfakes through improved public education and building a groundswell of political support to tackle the systemic issues and causes of fake news efficacy. We repeat, in this context, deep fakes are not the problem; the socio-economic models which lead to truth decay in the political sphere are.

6 Conclusions

In this paper, we have argued that the academic and civil discourse around fake news, cheapfakes and deepfakes is potentially harmful for our perception of the problems that these techniques give rise to. What we have observed is that a majority of the authors of papers related to these concepts implicitly or explicitly express concern for the development of these technologies per se, rather than the overall phenomenon of misinformation.

For example, Westerlund [34], in a review of deepfake research, argues: "The reviewed news articles suggest that there are four ways to combat deepfakes: legislation and regulation, corporate policies and voluntary action, education and training, and anti-deepfake technology". Concerning fake news, many articles promote so-called fake news detectors or identifiers (e.g. [37, 39–45]). For deepfake research, it is more of the same. In most of the reviewed papers, the authors are discussing how to either detect or prevent deepfakes (e.g. [46–51]). Thus, while many researchers argue that we need a healthy mix of regulatory support, corporate policies, education and training and anti-deepfake technologies, most of the research on deepfakes focuses on anti-deepfake technologies. Consequently, we note a strong urge within the field to attempt to solve a societal problem, i.e. malevolent actors spreading disinformation and non-consensual pornography, with purely technological means [52].

We have thus produced the beginnings of a typology of intentionally inaccurate representations of reality in media content (IIRR), which aims to help identify and inform consumers of media content, the ways in which information can mislead, both for entertainment and more nefarious purposes. The next iteration of this paper will aim to incorporate a more systemic approach to this analysis, alongside comments for improvement from our scholarly peers.

Additionally, this paper argues that although the technology itself is not neutral, the existence of these technologies themselves is not the problem. When it comes to deepfakes, Pandora's box has already been opened, and there is no way to "uninvent the bomb" [53]. It is reasonable to view the distribution of deepfaked content as yet another way of manipulation of the truth for malicious schemes, but – as with more traditional forms of textual manipulation – also for satire, humour and memes which is the foundation for much of today's online culture. Deepfaked content is just a new and more technologically complex expression of that culture. Hence, we agree with Gutiérrez-Martín et al's assertion that "the problem of misinformation is not solved by attacking the symptoms with a series of tips and checklists for consumers to learn how to detect fake news but, rather, by studying its underlying causes, one of which is excessive monetization and the diminishing value of truth in new virtual environments" [6]. Undoubtedly, over time the general public will once again become familiar and resilient to modern techniques, but how much damage will be done to the fabric of our society in the interim, if there is a complete breakdown of trust in our institutions?

References

1. Chesney, R., Citron, D.: Deepfakes and the new disinformation war: the coming age of post-truth geopolitics. Foreign Aff. **98**, 147–155 (2019)
2. Kofman, A., Kofman, E.: Bruno Latour, the post-truth philosopher, mounts a defense of science. New York Times **6** (2018)
3. Eichensehr, K.: Don't believe it if you see it: deep fakes and distrust. Jotwell J. Things We Like **1**, 1 (2018)
4. Miller, D.: Tell Me Lies: Propaganda and Media Distortion in the Attack on Iraq. Pluto Press, London (2004)
5. Department of Digital, Culture, Media, and Sport: Disinformation and 'fake news': Final Report: Eighth Report of Session 2017–19, pp. 1–109 (2019)
6. Gutiérrez-Martín, A., Torrego-González, A., Vicente-Mariño, M.: Media education with the monetization of YouTube: the loss of truth as an exchange value. Cult. Educ. **31**, 267–295 (2019). https://doi.org/10.1080/11356405.2019.1597443
7. Postman, N.: Amusing Ourselves to Death: Public Discourse in the Age of Show Business. Penguin (2006)
8. Silbey, J., Hartzog, W.: The Upside Of Deep Fakes (2019)

9. Molina, M.D., Sundar, S.S., Le, T., Lee, D.: “Fake news” is not simply false information: a concept explication and taxonomy of online content. Am. Behav. Sci. (2019). https://doi.org/10.1177/0002764219878224
10. Franks, M.A., et al.: Sex, Lies, And Videotape: Deep Fakes And Free Speech Delusions (2019)
11. Roozenbeek, J., Van Der Linden, S.: The fake news game: actively inoculating against the risk of misinformation (2018). https://doi.org/10.1080/13669877.2018.1443491
12. Lessenski, M.: Resilience to ‘post-truth’ and its predictors in the new media literacy index 2018 *. Open Soc. Inst. Raporu. Mart (2018)
13. McGlynn, C., Rackley, E., Houghton, R.: Beyond ‘revenge porn’: the continuum of image-based sexual abuse. Fem. Leg. Stud. **25**, 25–46 (2017)
14. DeKeseredy, W.S., Schwartz, M.D.: Thinking sociologically about image-based sexual abuse: the contribution of male peer support theory. Sex. Media Soc. **2**, 2374623816684692 (2016)
15. Urban Dictionary: Bubbling, defined by user “4Red,”. https://www.urbandictionary.com/define.php?term=Bubbling
16. Paris, B., Donovan, J.: Deepfakes and cheap fakes: the manipulation of audio and visual evidence. Data and Society (2019)
17. Coulter, M.: BBC issues warning after fake news clips claiming NATO and Russia at war spread through Africa and Asia. Evening Standard (2018)
18. Sadiq, M.: Real v fake: debunking the “drunk” Nancy Pelosi footage - video (2019)
19. Maras, M.-H., Alexandrou, A.: Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. Int. J. Evid. Proof **23**, 255–262 (2019). https://doi.org/10.1177/1365712718807226
20. Beridze, I., Butcher, J.: When seeing is no longer believing. Nat. Mach. Intell. **1**, 332–334 (2019). https://doi.org/10.1038/s42256-019-0085-5
21. Fried, O., et al.: Text-based editing of talking-head video. ACM Trans. Graph. (TOG) **38**(4), 1–4 (2019)
22. Zannettou, S., Sirivianos, M., Blackburn, J., Kourtellis, N.: The web of false information: rumors, fake news, hoaxes, clickbait, and various other shenanigans. J. Data Inf. Qual. **11**, 1–37 (2019)
23. Westerlund, M.: The emergence of Deepfake technology: a review. Technol. Innov. Manag. Rev. **9**(11), 39–52 (2019)
24. Ajder, H., Patrini, G., Cavalli, F., Cullen, L.: Report 2019: The State of Deepfakes (2019)
25. Öhman, C.: Introducing the pervert’s dilemma: a contribution to the critique of Deepfake pornography. Ethics Inf. Technol. **22**(2), 133–140 (2019). https://doi.org/10.1007/s10676-019-09522-1
26. BuzzFeed and Jordan Peele: You Won’t Believe What Obama Says In This Video!). BuzzFeedVideo (2018)
27. Future Advocacy: Deepfakes. https://futureadvocacy.com/deepfakes/
28. Maddox, T.: Here are the biggest IoT security threats facing the enterprise in 2017. Consult (2016). https://www.techrepublic.com/article/here-are-the-biggest-iot-security-threats-facing-the-enterprise-in-2017/
29. Dack, S.: Deep Fakes, Fake News, and What Comes Next (2019)
30. Hasan, H.R., Salah, K.: Combating Deepfake videos using blockchain and smart contracts. IEEE Access **7**, 41596–41606 (2019). https://doi.org/10.1109/ACCESS.2019.2905689
31. Chappell, B.: Facebook issues new rules on Deepfake videos, targeting misinformation (2020)
32. Twitter: Drew Harwell. (2020). https://twitter.com/drewharwell/status/1214394479026855936

33. Farkas, J., Schou, J.: Fake news as a floating signifier: hegemony antagonism and the politics of falsehood. Javnost Pub. J. Eur. Inst. Commun. Cult. (2018). https://doi.org/10.1080/13183222.2018.1463047
34. Westerlund, M.: The emergence of Deepfake technology: a review (2019)
35. Bovet, A., Makse, H.A.: Influence of fake news in Twitter during the 2016 US presidential election. Nat. Commun. **10**, 1–14 (2019)
36. Wu, L., Liu, H.: Tracing fake-news footprints: characterizing social media messages by how they propagate. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 637–645 (2018)
37. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: methods for finding fake news. Proc. Assoc. Inf. Sci. Technol. **52**, 1–4 (2015)
38. Waszak, P.M., Kasprzycka-Waszak, W., Kubanek, A.: The spread of medical fake news in social media – The pilot quantitative study. Heal. Policy Technol. **7**, 115–118 (2018). https://doi.org/10.1016/j.hlpt.2018.03.002
39. Wang, W.Y.: Liar, liar pants on fire: a new benchmark dataset for fake news detection. arXiv Prepr. arXiv:1705.00648 (2017)
40. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. ACM SIGKDD Explor. Newsl. **19**, 22–36 (2017)
41. Tacchini, E., Ballarin, G., Della Vedova, M.L., Moret, S., de Alfaro, L.: Some like it hoax: automated fake news detection in social networks. arXiv Prepr. arXiv:1704.07506 (2017)
42. Ruchansky, N., Seo, S., Liu, Y.: CSI: a hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797–806 (2017)
43. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. arXiv Prepr. arXiv:1708.07104 (2017)
44. Long, Y., Lu, Q., Xiang, R., Li, M., Huang, C.-R.: Fake news detection through multi-perspective speaker profiles. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pp. 252–256 (2017)
45. Granik, M., Mesyura, V.: Fake news detection using naive Bayes classifier. In: 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 900–903. IEEE (2017)
46. Li, Y., Lyu, S.: Exposing Deepfake videos by detecting face warping artifacts. arXiv Prepr. arXiv:1811.00656 (2018)
47. Güera, D., Delp, E.J.: Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE (2018)
48. Koopman, M., Rodriguez, A.M., Geradts, Z.: Detection of deepfake video manipulation. In: The 20th Irish Machine Vision and Image Processing Conference (IMVIP), pp. 133–136 (2018)
49. Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., Li, H.: Protecting world leaders against deep fakes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 38–45 (2019)
50. Matern, F., Riess, C., Stamminger, M.: Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 83–92. IEEE (2019)
51. Korshunov, P., Marcel, S.: Vulnerability assessment and detection of Deepfake videos. In: The 12th IAPR International Conference on Biometrics (ICB), pp. 1–6 (2019)
52. Fors, P.: Problematizing Sustainable ICT (2019)
53. MacKenzie, D., Wajcman, J.: The social shaping of technology: how the refrigerator got its hum. Milton Keynes–Philadelphia (1985)

54. Fazio, L., et al.: Knowledge does not protect against illusory truth. J. Exp. Psychol. Gen. **144** (5), 993–1002 (2015)
55. Merriam-Webster: Words We're Watching: 'Deep Fake'. Merriam-Webster. https://www.merriam-webster.com/words-at-play/deepfake-slang-definition-examples. Accessed 24 July 2020
56. Wolak, J., Finkelhor, D.: Sextortion: findings from a survey of 1,631 victims. Crimes Against Children Research Center, University of New Hampshire (2016)
57. Lenhart, A., Ybarra, M., Price-Feeney, M.: Nonconsensual image sharing: one in 25 Americans has been a victim of "revenge porn". Data & Society Research Institute, CiPHR, Data Memo (2016)
58. O'Connor, K., et al.: Cyberbullying, revenge porn and the mid-sized university: victim characteristics, prevalence and students' knowledge of university policy and reporting procedures. High. Educ. Quart. **72**, 344–359 (2018)
59. Citron, D., Franks, M.: Criminalizing revenge porn. Wake For. Law Rev. **49**(2), 345–392 (2014)

The Influence of Personal and Professional Commitments on Digitally Disconnected Experiences

Brad McKenna[1](✉), Lena Waizenegger[2], and Wenjie Cai[3]

[1] University of East Anglia, Norwich, UK
b.mckenna@uea.ac.uk
[2] Auckland University of Technology, Auckland, New Zealand
lena.waizenegger@aut.ac.nz
[3] University of Greenwich, London, UK
w.cai@greenwich.ac.uk

Abstract. In our ubiquitously connected world, it becomes more and more difficult to disconnect and leave all personal and professional commitments behind while on holiday. Mobile technology allows us to be connected wherever and whenever we want, but at the same time shifts expectations towards constant availability and responsiveness among friends and colleagues. Applying a qualitative research approach, we explored how social and professional commitments influence decisions and experiences of travelers that go on a digital-free holiday. Using the theoretical lens of surveillance, we found that travelers are digitally surveilled not only by their friends and family members on social media, but also by their superiors and colleagues through email and social networks. The expectations of being constantly available and responsive extend into their holiday, which makes it difficult for travelers to truly disconnect and enjoy their digital free travel experience. At the same time, they are inclined to engage in social surveillance of their peers which creates the constant urge to learn about any updates from their private and professional networks. We contribute to the tourism and information systems literature, by explaining how private and professional commitment influence the digital-free travel experience and extend the concept of surveillance to the work context.

Keywords: Digital-free travel · Surveillance · Private and professional commitments · Mobile technologies

1 Introduction

Ubiquitous connectivity has resulted in blurred boundaries of home/away and leisure/work [1, 2]. While technology offers various conveniences to travelers, it becomes increasingly difficult for them to switch off during their holidays due to increasing expectations of constant availability and responsiveness [3–5]. The idea of reviving digital disconnections and escapism on holiday has been popular recently in both travel products and academic studies [6]. In the past few years, tourism organizations including VisitEngland and VisitScotland have highlighted the trend of digital

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 305–314, 2020.
https://doi.org/10.1007/978-3-030-62803-1_24

detox and emphasized the strong connection between wellness tourism and efficient use of smartphones. The idea of going off-grid on holiday is popular among those who are highly connected in their daily lives.

However, although these travelers are motivated to disconnect, the expectations from their work and social environment regarding the traveler's online availability and responsiveness affect their freedom to switch off. In this paper, we explore the personal and professional commitments which impact the experiences of digital-free travel using the concepts of interpersonal electronic surveillance (IES) [7] and social surveillance [8, 9]. IES results in the surveillance of individuals using digital technologies, while social surveillance can be understood as using web 2.0 sites such as social media to keep track of the activities of friends, family and acquaintances [9]. The goal of this paper is to show that surveillance does not only take place in a private but also professional context through email response tracking or response behavior on enterprise social networks, for example, Slack.

Technology is integrated into many aspects of a holiday, from making bookings, navigation, searching for information, and maintaining connections back home. However, these technologies have also become tools of surveillance [10]; of monitoring [11]; and of constant interruptions [12] which can blur the boundaries of 'home' and 'away'. Despite the growing desire for digital-free travel [6, 13], these issues of surveillance, monitoring, and interruptions can place a barrier to disconnecting, or have a negative impact on the digital-free holiday experience. Therefore, using the concepts of surveillance, the research question in this paper is: *how do social and professional commitments influence decisions and experiences of digital-free travel?*

The paper presents a research in progress paper and is structured as follows. First, we present a literature review on surveillance and digital-free travel. Following this, we present our methods section. Next, we present some of our preliminary findings, and conclude the paper.

2 Literature Review

Mobile technology is reconfiguring time and space, social relations, and enables tourists to be socially present while physically absent [14]. This idea of 'absent presence' [15] has been investigated as the notion of 'copresence' in the tourism literature [16]. However, copresence could lead to negative effects such as lack of social interactions, fewer experiences of others, or decreased well-being [3].

In the past, travel was largely associated with 'away' and 'escapism' from mundane everyday life, both physically and socially. The involvement of mobile technology detaches physical and social proximity, and enables a person's mediated presence [17] when he/she is on holiday. The copresence of tourists thus brings their daily lives on holiday by constantly engaging with their personal and professional commitments mediated by advanced mobile technology [18].

2.1 Surveillance

Mobile devices and social media are common tools for sharing location and other personal information with friends and other users of digital services [19], however these technologies are also a means of surveillance to see what friends, family, and acquaintances are doing [7] and have become accepted as the norm in our daily lives [8]. IES relates to the digital strategies that individuals use to follow other users online and offline behavior. IES is a goal-orientated behavior which includes the surveillance of family members, friends, romantic partners, or colleagues. IES can occur through digital technologies such as social media, bulletin boards, personal websites, blogs [7], and mobile devices [19]. More specifically, IES results in social surveillance which arises from the continual investigations of others' digital traces left by people as they live in their highly connected lives [9]. People are aware of this constant monitoring and tailor their social media content with an audience in mind [20]. Marwick [9] distinguishes social surveillance from other forms of surveillance with three parameters: power, hierarchy, and reciprocity and argues that the desire to share content means they want to be seen by others.

The concept of surveillance as a form of interpersonal and social relations has been discussed by Germann Molz [21] in a tourism context. With the normalization of copresence, many travelers do not only share their travel experience constantly online; they are even expected to be virtually available and visible to audiences' surveilling gaze. Some travelers are expected to provide updates to appease worried relatives and friends. The updates allow parents, friends and co-workers to surveille the traveler and check where they are, what they are doing and if they are safe. The pressure to update the people that stayed at home e.g. parents, friends and co-workers creates pressure that to some extent limits their freedom of travelling. Due to mobile technologies travelers can always be contacted by email, social media or instant messengers; therefore, they can never hide or escape from this implied surveillance. Germann Molz [21] suggests the expectation of visibility and availability by audiences through online social networks may exacerbate rather than appease. Travelers oftentimes let their web audience know through out of office email replies or instant messaging that they might lose signal or the Internet connection when they stay in technological black holes (i.e. no phone signal) or consciously switch off their phones [22]. Nevertheless, parents, friends or colleagues start worrying about them or are annoyed as they are waiting for important information after only a few quiet days. These collective expectations concerning availability and responsiveness have been explored by Mazmanian, Orlikowski, and Yates [4] in the workplace context. They found that the increasing expectations towards availability and responsiveness led to the spiral of escalating engagement and diminished the employees' autonomy. This made it difficult for them to disconnect from work and led to increased stress levels and work-life conflict. Despite research on organizational studies focusing on the concept of disconnection from work, there has been sparse focus in tourism research.

2.2 Digital-Free Travel

Research has shown that digital-free travel lead to various positive outcomes such as improved well-being and work-life balance [e.g. 22, 23]. However, according to Dickinson et al. [22] some tourists cannot embrace the idea of disconnection due to the perceived negative emotions or experiences. People who are self-motivated to engage in digital-free tourism, experience professional and personal commitments that make them feel oppressed and they cannot escape the constant surveillance [24]. Much of the disconnection literature focused on negative emotions such as anxiety and tensions [25], and the idea of being off-the-grid creates anxious and distressing feelings for some people [26]. Some studies have focused on the emotional effects of being disconnected, e.g. Paris et al. [25] researched anxieties and social tensions. Tanti and Buhalis [27] explored five consequences (availability, communication, information obtainability, time consumption, and supporting experiences) of being (dis)connected. However, these studies did not explore personal and professional commitments, as well as the pressure of copresence as key triggers of these negative emotions. Germann Molz and Paris [28], Paris et al. [25], and Neuhofer and Ladkin [29] suggested that there are lack of empirical studies on digital-free travel with few exceptions [e.g. 25, 30]. Most studies reported findings of disconnection only as a secondary finding by asking participants to recall their ‘connected’ experiences. For example, Rosenberg [31] explored the disconnection topic by surveying backpackers about their connected behavior. There has been research on tourists who were forced to disconnect due to a ‘technology dead zone’, an area with no or poor connection [30], and recently, e-lienation [32], and media discourse of digital-free travel [23] providing essential insights into this topic.

3 Methodology

This study is underpinned by the interpretive paradigm using the diary method and semi-structured interviews [33, 34], and builds on the prior work of the authors [6, 13]. In tourism studies, the diary method has been adopted to understand travel behavior and experiences [35]. Participants were recruited through a combination of self-selection and snowballing sampling techniques. The project was marketed with a public post on Facebook with the request for interested people to contact us. Further selection criteria were applied to ensure participants are frequent digital technology users and desired to take part in digital-free travel experiences. In the participant information sheet, we operationalize our definitions of disconnection and technology as: mobile, computer, laptop, tablet, Internet, social media, sat navigator, television, or radio/audio player.

The data collection was conducted in two stages. In the first stage, participants were instructed with guidelines to write diaries to record their instant emotions and detailed accounts before, during and after their disconnected experiences. In addition, we also asked participants to note down occasions where they had to finish the digital-free experience before their initial plan. This stage was conducted between August 2016 and March 2017. The richness of the diary data recorded several interesting narratives worth further investigation. In the second stage between April and October 2017 we conducted semi-structured in-depth interviews to further investigate participants' reflections of their disconnected experiences. Most of the interviews were conducted face-to-face, only one took place over the phone.

In total, 24 participants (14 male and 10 female) from 7 countries traveled to 17 countries and regions. Participants are mostly millennials except for two that belong to Generation X. 15 diaries were hand-written by participants and transcribed by them after their trips. In addition, we conducted 18 interviews. We analyzed our data following the guidelines of thematic analysis [36] to identify the key concepts in our data. Table 1 contains our participant information. All participants are working professionals.

Table 1. Participant information (S: Sex; DD: Disconnect Duration; D: Diary; I: Interview). Modified from [6].

Name	S	Age	Travel From	Travel To	Total Trip	DD	D	I
James	M	35–40	Norwich, UK	Ely, UK	1	1	X	
			Norwich, UK	Vienna, Austria	4	3	X	
Thomas	M	25–30	Norwich, UK	Ely, UK	1	1	X	
			Norwich, UK	Vienna, Austria	4	3	X	
John	M	50+	Manchester, UK	Hebrides, UK	13	7	X	
Richard	M	35–40	Arlington, Virginia, USA	Orleans, Massachusetts, USA	6	1	X	
Frank	M	40–45	Arlington, Virginia, USA	Orleans, Massachusetts, USA	6	1	X	X
Youngqi	F	30–35	Xiamen, China	Neuschwanstein, Germany	10	2	X	X
Zhenpeng	M	25–30	Guangzhou China	Macau, China	2	0.5		X
Billy	M	20–25	Melbourne, Australia	Switzerland and France	3	1	X	X
			Melbourne, Australia	Berlin, Germany	3	1	X	X
Anita	F	25–30	Edinburgh, UK	Cerveny Klastor, Slovakia	1.5	1.5	X	
Sally	F	30–35	Auckland, NZ	Queenstown, NZ	3	1	X	X
Jiaying	F	35–40	Portsmouth, UK	Copenhagen, Denmark	3	2	X	X
Lisa	F	25–30	Munich, Germany	Taipei, Taiwan	3	3	X	
Heike	F	25–30	Innsbruck, Austria	Cuba	13	13		X
Susan	F	30–35	Auckland, NZ	Tonga	5	5		X
Sean	M	30–35	Auckland, NZ	Tonga	5	5		X
Stephan	M	25–30	Innsbruck, Austria	Cuba	13	13		X
Rory	F	30–35	Auckland, NZ	Abel Tasman National Park, NZ	4	3		X
Nico	M	30–35	Auckland, NZ	Abel Tasman National Park, NZ	4	3		X
Doug	M	30–35	Auckland, NZ	Abel Tasman National Park, NZ	4	3		X
Steven	M	30–35	Auckland, NZ	Cook Islands	7	7		X
Larissa	F	25–30	Innsbruck, Austria	Kiev, Ukraine	14	0		X
Lauren	F	25–30	Auckland, NZ	Fiji	14	3–4		X
Noah	M	25–30	Auckland, NZ	Fiji	14	3–4		X
Bailey	M	20–25	Norwich, UK	Spain	35	5.5	X	

4 Preliminary Findings

This section will discuss the preliminary findings from our study. We found that some participants consider digital-free travel as a great opportunity for them to take a break from social and professional commitments. Heike described, that being completely disconnected from all commitments of which she would have been reminded of, if she would have used her phone, made it possible for her to switch to holiday mode right after she disconnected: "*I got used to it very quickly that I did not have any Internet and it was really convenient. I felt like I was on holidays and far away from everything, really relaxed and I enjoyed it much more as I was completely disconnected and could switch off entirely. I did not think of my work at all, not until the second to last evening. I felt very free and relaxed*". Frank disclosed that the refreshing and liberating feeling of digital-free travel came from *'not having to expose myself to the news (most of which is typically annoying political-related news) nor expose myself to work-related email messages that typically arrive on my phone'.* After her digital-free travel experience, Anita wants to do more in the future: *'I'd love to go back to being able to disconnect from reality when on holiday and move away from the expectation from others to always be available'.* Although Anita liked her digital-free travel, she found that IES from her colleagues (they expect her to be connected) makes it very difficult to leave her professional commitments behind and truly enjoy her holidays: '*the differences would come down to having to think about work because you are connected and people are trying to reach you'.*

We found that IES coming from travelers' commitments in their daily lives are one of the key forces that create negative disconnecting emotions and prevent them from fully engaging with digital-free tourism. Travelers still practice their obligations and social roles as employees, family members, and partners mediated by technology. Although, they judge the idea of going digital-free as appealing, they cannot, and often do not want to let go of their professional and private commitments and therefore, do not fully disconnect. Due to their various responsibilities at work and at home that come with their different roles, they feel obligated to be available and responsive as they did not want to fail their colleagues, friends and family members that were seeking information from them, needed their help or just wanted to know if they are OK.

Many of our participants have too many personal and professional commitments which hardly allow them to be disconnected and unavailable for a certain amount of time. Nico noted that due to the nature of his job, he could hardly dare to switch off and hand over the responsibility to his employees: '*to be honest, two days before the trip I was a little bit nervous about it, because I knew the guys I am working with, I was leaving them with a lot of responsibility of stuff in an area where they were not aware of everything, where they did not understand everything. So, I was quite nervous about it. So, to a certain degree, two days before the trip I was like do I really wanna do this? Is that really a good time?*'. Doug explained that the digital-free holiday almost led to him missing an important opportunity at work as he did not respond to an email he received during his digital-free travel: *'I got an important email on the Friday that I did not respond until Monday, which could have been a missed opportunity because there was a deadline on it'.*

Not only professional commitments, but also personal commitments have significant effects. Andy was expecting an important parcel and was nervous that he couldn't track it and did not know if it arrives on time: *'I did have a nagging feeling to check my messages and emails to see whether a delivery had arrived back at home in Melbourne or not. I later found out that it did arrive on time, but this was the following day when I had regained my access to technology'* (Billy). Rory has a pet rabbit and found it is impossible to disconnect on holiday. In the past, her rabbit has stopped eating and needed to be taken to the vet. During her trip away she tried to be disconnected but could not fully switch off the whole time as she wanted to check with her flatmate if her pet was still eating properly: *'in my case, I wasn't fully disconnected because I have a rabbit. Sometimes if my rabbit stopped eating, it will die…I wanted my flatmate to be able to contact me in case something went wrong'.*

Houjia stated that her mother won't allow her to disconnect when she travels alone: *'she will be worried about what if I am in danger, and she cannot reach out to me*'. She admitted that she experienced pressure from her mum to be always available when she is travelling: *'it creates this kind of anxiety. Even though I have my phone, but I do not have reception, it will create certain kind of anxiety to me'.* Similarly, Frank also noted the feeling of being obligated to report whereabouts on holiday under social surveillance. He also wanted to check what his friends and family are doing and therefore, wanted to check his messages and social media *'I did not *HAVE* to use technology, but I felt I needed to catch up on text messages and social media to see what had been going on with my family/friends during that day'*. This shows the reciprocity of social surveillance. Not only do our travelers feel that they are obligated to be available and responsive, but they also want to be updated about any private and work-related matters.

The norm of ubiquitous connectivity and social surveillance results in people feeling that they need to be available and responsive even during their planned digital-free holidays. Based on their past connectivity patterns, people expect travelers to post on social media and to respond to messages or emails within their usual times. James who is usually very responsive and posts frequently on social media worried about the emails he had in his inbox that are waiting for him and that someone tried to contact him however, couldn't get hold of him: *'It has been almost 24 h without technology and I am starting to wonder how many emails do I have, or Facebook posts. As I had posted on Facebook before I disconnected that I was going to Vienna. I'm also slightly worried someone might be trying to contact me, and getting worried I am not replying. I did tell the people for that matter I was disconnecting but I'm still worried. What if something bad has happened and they can't get hold of me'?* (James). On another digital-free trip, James perceived similar feelings, but he realized that he did not get as many emails and messages as he thought he would get. He concluded that it is OK to disconnect, and that people do not contact him as often as he thought. *'I spent the day wondering if people had tried to contact me. But when we got home and got my phone back there was no messages on Facebook, text message or anything. It made me think how much do I really need my phone during the day'.*

Frank experienced anxiety and stress-inducing feelings as he did not tell many people that he was on a digital-free holiday. He worried about the fact that people might try to reach out to him, but cannot: '*I hadn't mentioned to anyone outside of my travel friends that I was going without technology that day so I was worried that they* *might have been wondering why I had disappeared and been non-responsive that day*'. Known as being always available and responsive online, Doug felt it is important to manage people's expectations and let them know that he will be disconnected for a certain time: '*letting people know that you're going to have limited connectivity. Probably the longest that I could go without any communication would be two weeks and that would require letting a lot of people know*'.

5 Conclusion and Implications

In this research in progress paper, we presented our preliminary findings on the impact of social and professional commitments mediated by interpersonal electronic surveillance on the decisions and experiences of travelers that go on a digital-free holiday. Based on our data analysis we found that travelers that choose to go on a digital-free holiday need to deal with the stress and anxieties of disconnecting from collective expectations deriving from their social and professional commitments. Due to many commitments, some participants found it challenging; some even decided to reconnect earlier and return to the status of copresence. Our next steps are to more thoroughly theoretically analyze the data using the surveillance lens.

By understanding how social and professional commitments prevent travelers from engaging in the digital-free travel experience, this study not only contributes to the emerging digital-free tourism literature, but also provides new insights into the notion of copresence by investigating the paradox between autonomy and interpersonal surveillance in a disconnected context. The findings can also provide insights to tour operators and hospitality providers who are designing digital-free related products.

References

1. Neuhofer, B., Buhalis, D., Ladkin, A.: A typology of technology-enhanced tourism experiences. Int. J. Tour. Res. **16**(4), 340–350 (2014)
2. Dery, K., Kolb, D., MacCormick, J.: Working with connective flow: how smartphone use is evolving in practice. Eur. J. Inf. Syst. **23**(5), 558–570 (2014)
3. Ayeh, J.K.: Distracted gaze: Problematic use of mobile technologies in vacation contexts. Tour. Manag. Perspect. **26**, 31–38 (2018)
4. Mazmanian, M., Orlikowski, W.J., Yates, J.: The autonomy paradox: the implications of mobile email devices for knowledge professionals. Organ. Sci. **24**(5), 1337–1357 (2013)
5. Mazmanian, M.: Avoiding the trap of constant connectivity: when congruent frames allow for heterogeneous practices. Acad. Manag. J. **56**(5), 1225–1250 (2013)
6. Cai, W., McKenna, B., Waizenegger, L.: Turning it off: Emotions in digital-free travel. J. Travel Res. **59**(5), 909–927 (2020)
7. Tokunaga, R.S.: Social networking site or social surveillance site? Understanding the use of interpersonal electronic surveillance in romantic relationships. Comput. Hum. Behav. **27**(2), 705–713 (2011)
8. Hermida, A., Hernández-Santaolalla, V.: Horizontal surveillance, mobile communication and social networking sites. Lack Priv. Young People's Daily Lives. Commun. Soc. **33**(1), 139–152 (2020)

9. Marwick, A.: The public domain: surveillance in everyday life. Surveill. Soc. **9**(4), 378–393 (2012)
10. Kaplan, C.: Transporting the subject: technologies of mobility and location in an era of globalization. Trans. Proc. Modern Lang. Assoc. Am. **117**, 32–42 (2002)
11. Green, N.: Who's watching whom? Monitoring and accountability in mobile relations. In: Brown, B., Green, N., Harper, R. (eds.) Wireless World Computer Supported Cooperative Work, pp. 32–45. Springer, London (2002). https://doi.org/10.1007/978-1-4471-0665-4_3
12. Wajcman, J., Rose, E.: Constant connectivity: rethinking interruptions at work. Organ. Stud. **32**(7), 941–961 (2011)
13. Floros, C., et al., Imagine being off-the-grid: millennials' perceptions of digital-free travel. J. Sustain. Tour. 1–16 (In press, 2019)
14. Urry, J.J.S.: Mobil. Prox. **36**(2), 255–274 (2002)
15. Gergen, K.: The challenge of absence presence. In: Katz, J. (ed.) Perpetual Contact: Mobile Communications, Private Talk, Public Performance, pp. 227–254. Cambridge University Press, Cambridge (2002)
16. Hannam, K., Butler, G., Paris, C.M.: Developments and key issues in tourism mobilities. Ann. Tour. Res. **44**, 171–185 (2014)
17. Zhao, S. Toward a taxonomy of virtual presence. In: International Workshop on Presence 2001 (2001)
18. White, N.R., White, P.B.: Home and away: tourists in a connected world. Ann. Tour. Res. **34** (1), 88–104 (2007)
19. Humphreys, L.: Who's watching whom? A study of interactive technology and surveillance. J. Commun. **61**(4), 575–595 (2011)
20. Trottier, D.: A research agenda for social media surveillance. Fast Cap. **8**(1), 59–68 (2011)
21. Germann Molz, J.: Watch us wander: mobile surveillance and the surveillance of mobility. Environ. Plan. A **38**(2), 377–393 (2006)
22. Dickinson, J.E., Hibbert, J.F., Filimonau, V.: Mobile technology and the tourist experience: (dis)connection at the campsite. Tour. Manag. **57**, 193–201 (2016)
23. Li, J., Pearce, P.L., Low, D.: Media representation of digital-free tourism: a critical discourse analysis. Tour. Manag. **69**, 317–329 (2018)
24. Cooper, G.: The mutable mobile: social theory in the wireless world. In: Brown, B., Green, N., Harper, R. (eds.) Wireless World: Social and Interactional Aspects of the Mobile Age, pp. 17–31. Springer, London (2002). https://doi.org/10.1007/978-1-4471-0665-4_2
25. Paris, C.M., et al.: Disconnected and unplugged: experiences of technology induced anxieties and tensions while traveling. In: Tussyadiah, I., Inversini, A. (eds.) Information and Communication Technologies in Tourism 2015, pp. 803–816. Springer, London (2015). https://doi.org/10.1007/978-3-319-14343-9_58
26. O'Regan, M., Hypermobility in backpacker lifestyles: the emergence of the internet café. In: Tourism Mobilities: Local-Global Connections, pp. 109–132(2008)
27. Tanti, A., Buhalis, D.: Connectivity and the consequences of being (dis)connected. In: Inversini, A., Schegg, R. (eds.) Information and Communication Technologies in Tourism 2016, pp. 31–44. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28231-2_3
28. Germann Molz, J., Paris, C.M.: The social affordances of flashpacking: exploring the mobility nexus of travel and communication. Mobilities **10**(2), 173–192 (2015)
29. Neuhofer, B., Ladkin, A.: (Dis)connectivity in the travel context: setting an agenda for research. In: Schegg, R., Stangl, B. (eds.) Information and Communication Technologies in Tourism 2017, pp. 347–359. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-51168-9_25
30. Pearce, P.L., Gretzel, U.: Tourism in technology dead zones: documenting experiential dimensions. Int. J. Tour. Sci. **12**(2), 1–20 (2012)

31. Rosenberg, H.: The flashpacker and the unplugger: cell phone (dis) connection and the backpacking experience. Commun. Mob. Media **7**(1), 111–130 (2019)
32. Tribe, J., Mkono, M.: Not such smart tourism? The concept of e-lienation. Ann. Tour. Res. **66**, 105–115 (2017)
33. Wengraf, T.: Qualitative Research Interviewing: Biographic Narrative and Semi-Structured Methods. Sage, London (2001)
34. Walsham, G.: Doing interpretive research. Eur. J. Inf. Syst. **15**(3), 320–330 (2006)
35. Vu, H.Q., et al.: Tourist activity analysis by leveraging mobile social media data. J. Travel Res. **57**(7), 883–898 (2018)
36. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006)

Social Materiality of Smartphone Game Apps

Case Analysis of *Pokémon GO*

Hiroshi Koga(✉)

Kansai University, Takatsuki 569-1095, Japan
Koga@res.kutc.kansai-u.ac.j

Abstract. In this paper, we examine the connectivity and digitalization affecting our private lives through a case analysis of *Pokémon GO*, a game application for smartphones. First, we outline the concept of sociomateriality as a keyword to consider the effects of *Pokémon GO*. Subsequently, we explain the outline of the game *Pokémon GO*. Finally, we show the usefulness of sociomateriality for examining the impact of digitalization on our private lives, using the example of *Pokémon GO*.

Keywords: Sociomateriality · Smartphone game app · Pokémon GO

1 Introduction

In recent years, connectivity and digitalization have deeply penetrated our daily lives. Consequently, it is no exaggeration to say that our daily lives and collective reality have changed significantly

Consider, for example, the phenomenon of the filter bubble [1], which happens when web space negates the infinitely expanding information space. The filter bubble creates personal information spaces filled with only the limited and biased views that website algorithms have guessed we want to see. The guesses are based on our past online activity, personal information, and demographic descriptors, including our locations, past click actions, and search histories. This leads us to see and hear only opinions similar to our own, resulting in the echo chamber phenomenon [2], the result of people's tendency to look for information that reinforces their existing views—filtering out dissent and filling the "bubble" with accord—through unconscious confirmation bias.

Now, consider that in the Digital Era, there are a variety of realities (e.g., virtual reality, augmented reality (AR), mixed reality, and substitutional reality), so we cannot rule out the possibility that the filter bubble will expand into real space. When the information space becomes combined or fused with reality, this provides a more personalized information space, reinforcing the tendency toward confirmation bias.

By the way, this real–virtual information environment is already deeply involved in our daily lives. No special equipment is required to enjoy such an information space. The only device needed is a smartphone.

ICT (information and communications technology) has advanced sufficiently that mobile phones are no longer considered "special equipment," and smartphones are the

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 315–322, 2020.
https://doi.org/10.1007/978-3-030-62803-1_25

latest evolution of mobile phone technology. They are not just phones; their various sensor, GPS, and data communication functions are used more often than their telephone functions.

In this paper, we examine how the composite device known as the smartphone affects our daily lives using the theory of sociomateriality, a framework for studying the intersection of technology and everyday life, especially leisure. As detailed in the next section, sociomateriality has been proposed as a theoretical framework for analyzing everyday organizational life. Its characteristics are that social and organizational factors and material and technical factors are not understood as binary conflicts but as indivisible factors. In this paper, we adopt this framework, extending and applying sociomateriality to the connectivity and digitalization in everyday private life through examining the use of *Pokémon GO*, a game app on smartphones.

The remainder of this paper is organized as follows. First, we outline the concept of sociomateriality as a keyword to consider the effects of *Pokémon GO*. Subsequently, we explain the outline of the game *Pokémon GO*. Finally, we show the usefulness of the sociomateriality framework for examining the impact of digitalization in our private life, using the example of *Pokémon GO*.

2 The Concept of Sociomateriality

Many keywords have been proposed as a theoretical framework for examining how connectivity and digitalization affect our private lives. In this paper, we chose the lens of sociomateriality, through which we focus our investigation.

In recent years, sociomateriality has become an increasingly common theoretical framework in information system research for considering the effects of ICT on everyday organizational and work life. The term was introduced in a paper published in 2007 by Orlikowski, the Alfred P. Sloan Professor of Information Technologies at the Massachusetts Institute of Technology Sloan School of Management [3].

The following year, Orlikowski and Scott [4] categorized streams of research on ICT and organizations into three research streams based on their view of the ontological priority of technology; First, the research group considers technologies and organizations as "discrete entities" and examines their effects (or moderation). Next, the researches understand technology and organization as "mutually dependent ensembles" and discuss "interaction" between them from the viewpoint of "affordance". The last is sociomateriality, which regards the relationship between technology and organization as "sociomaterial assemblages," allowing a distinctive research stream that examines "entanglement" and "performativity".

The significance of the sociomateriality concept lies in its assertion that social and material factors should not be distinguished in the practice of organizational activities. In other words, it assumes that in everyday organizational life with ICT systems, social and material factors are in a state of "constitutive entanglement". Simply put, this theoretical framework focuses on the intrinsic intersections of technologies, jobs, organizations, and activities.

For example, when people use public transportation such as railroads and buses, it is not possible for those with smartphones to separate their devices from the act of traveling. Rather, it would be more appropriate to recognize that the use of smartphones has changed the meaning of travel. In extreme cases, low battery in a smartphone may create anxiety about moving itself, and it is highly likely that many people cannot remember or imagine traveling before the emergence of smartphones. For example, just a few decades ago, no one suffered paroxysm of anxiety over the prospect of traveling with a low battery or no cell service or Wi-Fi. As there are approximately three billion smartphone users worldwide, many of them millennials, it is highly likely that many people cannot remember or imagine traveling before the emergence of smartphones.

In many ways, people have become their geotags and metadata. Sociomateriality allows for a research stream that understands this assemblage—people and smartphones—and focuses on the process of socially configuring the entanglement.

That is the essence of sociomateriality; it provides a way to reexamine the binary conflict between organizations or people and technology and translate the complexity of the real world into a relationship rather than a causal model. In Buddhist terms, "causality" is replaced by *engi* (縁起, which refers to the idea that everything in this world is dependent on and related to one another in direct and indirect ways, changing and disappearing in the relationship—that is, "dependent co-arising".

Previously, discussions of sociomateriality have primarily focused on work life. However, this paper focuses on everyday private life to broadly consider the effects of connectivity and digitalization. The significance of this paper lies in how it expands the discussion, extending the sociomateriality perspective to everyday private life beyond the concept as proposed by Orlikowski and others [3, 4]. Specifically, this paper adopts the concept of mobilities from the field of tourism research.

Mobility has become an evocative keyword for the twenty-first century and a powerful discourse that creates its own effects and contexts [5]. Especially, Urry [6] advocated the concept of mobilities as a way of rethinking the intersection social science and transport science—essentially, tourism—in terms of assemblages in the act of moving. The mobilities concept proposes that many of the human abilities of tourism (that is, mobilities) should be understood as a hybrid of technology (mobility, smartphones) and the physical environment.

Certainly, when understanding the behavior of a person who operates a smartphone while walking, it may not be useful to distinguish between the person and the device. Instead, they should be understood as a hybrid. In Buddhist terms, this is called nini-funi (而二-不二 in Japanese; in English, duality–nonduality or two (in phenomena) but not two (in essence)). For example, unlike walking, moving by car involves an act of movement that is no longer a human activity alone but a hybrid activity. This leads people to perceive the landscape differently.

In Japan, there is the neologism "Insta-Bae" (インスタ映え), for which the English equivalent would be "instagrammable," This means that the photos uploaded to the photo- and video-sharing social networking service (SNS) Instagram get noticed; photos receive "likes" from followers, so people seeking more followers post *Insta-Bae* photos to gain praise and new fans. Therefore, consider the behavior of tourists who think that the view of Kobe is photogenic. Earlier, they might have thought, "This view of Kobe is beautiful!" However, now, their first thought is often "I have to upload an image to Instagram!" Such people cannot exist comfortably without a smartphone device. In other words, the act of sightseeing is becoming impossible to practice without smartphones (cameras, SNS, other apps). This can be easily understood by imagining the anxiety of forgetting or losing your smartphone. This can be understood as the act of tourism forming a mixture of devices and SNS. Therefore, the keyword "hybrid or assemblages" is attracting attention.

This paper examines these specific effects of increased connectivity and digitalization using the example of a popular location-based AR smartphone game, *Pokémon GO*.

3 Overview of *Pokémon GO*

Pokémon GO is a smartphone game application jointly developed by Niantic, Inc., Nintendo, and The Pokémon Company. Niantic, Inc., is a US company that produces location-based apps and games for mobile devices, such as Ingress, which combines the map information of Google Maps with GPS and AR. The "encampment game" is characterized by its real world–virtual world interactivity: players' real-world GPS locations and surroundings (e.g., public buildings, monuments, etc.) are used to manifest virtual "portals" tied to a narrative game in which players save the world.

Pokémon GO was released in July 2016. Building on the portals created for Ingress, in *Pokémon GO*, players use their smartphones to interact with an AR narrative; in this game, they move through the real world to capture, battle, and train fantasy creatures.

In 2016, it generated US$950 million in sales, with a cumulative US$2.2 billion in 2018. In this game, the main objective is to capture *Pokémons* that players "encounter" while moving; the characters are manifested in the game based on the smartphones' GPS function. The smartphone has to be moving for the Pokémon characters to appear on the map of the application. Thus, the game was touted as a way of promoting physical activity by encouraging players to go outdoors. The game contains "eggs," special Pokémon s that hatch only after a specific distance is traveled—another mechanism to induce movement.

Several items can be used to capture the *Pokémons*. These items (and the eggs) are available at specific locations (*PokéStops* and *Pokémon Gyms*). Some items are free, but many require in-app purchases. In the virtual gyms, players can battle or interact with other *Pokémons*, and spending time there earns them coins that can be exchanged for items in the game. (Of course, players can also buy game coins with real money).

Although the foundation of the game everywhere is the global map information from Ingress, companies can pay to be "sponsored locations" for *PokéStops* and *Pokémon Gyms* to attract customers. In Japan, the game is coordinated with the retailers AEON Group, fast-food purveyors McDonald's, and SoftBank Group's ICT shops. Those companies and many others are involved in the game elsewhere (e.g., Starbucks).

The attraction of Pokémon GO's play lies in its global stage. In other words, the emphasis is on moving in the real world, not just in the closed virtual space of a smartphone screen. For this reason, the game has a significant influence on players' daily activities—especially those involving movement.

4 How *Pokémon GO* Impacted Our Daily Lives

4.1 Previous Research

Since the launch of *Pokémon GO* in July 2016, its significance, possibilities, and problems have been discussed. As a trial, when I searched for "*Pokémon GO*" with Google Scalar, fibrous documents appeared. For example, privacy and surveillance [7], impact on children [8], effects on health [9, 10], and sociological considerations [11]. However, this paper examines the impact of Pokémon GO on everyday life.

4.2 Viewpoint of Our Analysis

As discussed, the *Pokémon GO* game app requires movement, which sets it apart from the typical image of videogames as lonely play in a room. This makes it relevant to mobilities in our everyday lives. In this section, we consider these specific points: (1) the changing image of smartphone game apps; (2) the importance of place consumption and geolocation, especially while walking; and (3) the orthogenesis of mobilities (or the trap of excessive mobilities) from the viewpoint of sociomateriality.

As sociomateriality is a framework for clarifying the meaning of everyday life as a hybrid of social and material factors, we will now examine the effects of digitalization and connectivity by investigating how the play of *Pokémon GO* has shaped many people's daily lives.

4.3 Change in Game Image: A Means of Maintaining and Improving Health at Outdoor

Walking outdoors has a high affinity with the concepts of "health" and "wellness". *Pokémon GO* has been promoted as an effective tool for maintaining and improving health and wellness [9, 10].

What does walking mean in this context? Urry (2007) offered four classification axes for walking [12]: (1) whether there is adventure; (2) whether you are alone; (3) whether there is a relationship with health and fitness; and (4) whether there is a mechanism to change the physical environment that supports walking.

Playing *Pokémon GO* is meant to encourage adventures (both virtual and real) for lone players as they walk (usually), so it scores high on both of those axes. In addition, because it emphasizes movement as opposed to just sitting at a stationary position facing a screen, it promotes health. The mechanism changing the physical environment is the game itself and its real world–virtual world interactivity. Thus, according to Urry's criteria, it involves and encourages healthy walking through the adventure of finding new destinations to capture *Pokémons* and other items. The destinations are not just fast-food restaurants, shops, and other affiliated business facilities. *PokéStops* and *Pokémon Gyms* can also be found in noncommercial spaces, such as museums, cathedrals, parks, and many tourist destinations. Therefore, it can be said that *Pokémon GO* provides an opportunity to refocus the tourism gaze ("re-gaze") at places that were previously overlooked (cf. Urry & Larsen, 2011 [13]).

The game has attracted not just young people but has also caught on with health-minded middle-aged and elderly players. Most of them are at least tangentially familiar with the characters of *Pokémon* because the original Pokémon game was released in 1996, and it was (and remains) wildly popular. It is thought that many people downloaded the Pokémon GO app because of nostalgia.

A reason for the popularity of location-based AR games is that they can change the nature of people's idle travel time—that is, time spent on trains or buses, in carpools, or walking to work. For many, those periods are spent in activities that could be called "killing time," such as scrolling through SNS. *Pokémon GO* gives them at least the illusion of a goal. As the distance traveled is an important factor in the progress of the game (e.g., obtaining and hatching eggs), the meaning of commuting time changes from "killing time" to "key factor in playing the game". Thus, *Pokémon GO* has created the lifestyle of "walking while operating a smartphone".

The Pokémon GO app itself has come to be positioned as a tool that enriches daily life—a device that adds new meaning to movement. That is a sociomaterial interpretation of the game application wherein the meaning of the app is socially structured through the players' processes.

4.4 Consumption of Place Born from Walking

Pokémon GO has also given new meaning to the simple word "place" and the concept of "place consumption". Urry [14] restructured the concept of "place consumption," making these four points: (1) places provide a context for the comparison, evaluation, purchase, and use of goods or services; (2) places are visually consumed; (3) places can be depleted or exhausted by use; and (4) localities consume people's identities.

From Urry's perspective, *Pokémon GO* offers a new value for consuming places. Players walk around cities and aim for specific landmarks (i.e., *PokéStops* and *Pokémon Gyms*) shown on their smartphone screens. These are real "places" such as a particular store, a particular vending machine, a special performance, or a statue. Therefore, the landmarks in the game and the real places intersect, multiplying the meaning of the places. That is exactly what Urry means by "consuming places". Thus, digitalization—specifically, a game app generating new meanings for "walking" and "place"—creates a new way to "consume a place".

4.5 The Orthogenesis of Mobilities

From the perspective of sociomateriality, the process of digitalization can be understood as the process of constant meaning generation characterized by accidents and unintended consequences. The play in *Pokémon GO* also has unintended consequences. Here, we must discuss a contradiction that has arisen between orthogenesis and orthogenetic evolution as it is used in biology and how it applies to the context of mobilities. The formal definition of orthogenesis is "a theory that variations in evolution follow a particular direction and are not merely sporadic and fortuitous," per Merriam-Webster.

Orthogenesis implies that the evolution of an organism continues in an advantageous direction, pushed by some "driving force". The validity of the concept of orthogenesis in current biology research is questionable. Nevertheless, it can be useful in explaining social phenomena. It does apply fairly well to game development; when developers find something that appeals to users, they pursue it.

It also applies to playing *Pokémon GO*, in which daily practice develops in the direction of increasing the distance moved. To extend the distance traveled, players supplement walking with modes of transportation, when this means riding rather than operating vehicles, there are no inherent problems. However, when this means riding bicycles or driving cars, attention-related mishaps occur. Consequently, there have been numerous reports of traffic collisions caused by people playing games while driving. For example, during the first 148 days following the release of *Pokémon GO*, there were 145,632 traffic accidents, 29,390 injured, 256 dead, and 20 economic losses —in the United States alone—related to people playing the game while driving. The resulting damages were estimated to be US$7.3 billion [15]. Accidents have also occurred in Japan. Consequently, Japan has considered imposing strict penalties for "gaming while driving". In addition, app changes have been made. A "speed limiting function" using a smartphone sensor has been added. In other words, everyday use in the real world demanded a change in the app.

Many *Pokémon GO* events have been organized using the lure of rare characters for capture. However, floods of players at event venues have caused problems with "over-tourism". Many countries have asked Niantic to exclude certain locations (e.g., government buildings, sensitive historical and religious sites, private property) from being used as PokéStops and Pokémon Gyms, and many sponsored locations have been removed at the companies' request. In general, retailers such as AEON are less likely than fast-food restaurants to benefit from players stopping onsite to play a game; people playing *Pokémon GO* in the middle of a shop can annoy other customers. In addition, large *Pokémon GO* events can lead to problems with excessive noise and littering, making association with the game problematic, especially on private land.

5 Discussion

In this paper, we have described how *Pokémon GO*, released in 2016, is used in everyday life, including how its players' use of the game has unintentionally led to changes in the game app itself. The impact of digitization and connectivity on everyday life cannot be explained by simple causal relationships—for example, technology determinism. Rather, it is more appropriate to understand that social and organizational factors and material and technical factors are integrated. In practice, they influence each other microscopically.

The sociomateriality perspective provides a theoretical framework for observing this, revealing that digitization and connectivity, as they relate specifically to location-based AR games, can create unique "filter bubbles" that modify their travel, health, and place consumption behaviors.

Acknowledgment. This work was supported by JSPS KAKENHI Grant Number JP 17K03909 and JP20K01899, by the Institute of Economic and Political Studies, Kansai University, the Japan Society for Management and Information Grant-in Aid for SIG "Monitoring and Control of

AI Artefacts", and the Seikei University Grant-in-Aid 2020 for the research on "Monitoring and Control of AI Artefacts: Consideration from Economic, Social and Legal Perspectives".

References

1. Pariser, E.: The Filter Bubble: What the Internet is Hiding from you. Penguin, UK (2011)
2. Sunstein, C.R.: Republic.com. Princeton University Press, USA (2001)
3. Orlikowski, W.J.: Sociomaterial practices: exploring technology at work. Organ. Stud. **28**(9), 1435–1448 (2007)
4. Orlikowski, W.J., Scott, S.V.: Sociomateriality: challenging the separation of technology, work and organization. Acad. Manag. Ann. **2**(1), 433–474 (2008)
5. Hannam, K., Sheller, M., Urry, J.: Editorial: mobilities, immobilities and moorings. Mobilities **1**(1), 1–22 (2006)
6. Urry, J.: Does mobility have a future? In: Urry, J. (ed.) Mobilities: New Perspectives on Transport and Society, pp. 21–38. Routledge, London (2016)
7. Sablatura, J., Karabiyik, U.: Pokémon go forensics: an android application analysis. Information **8**(3), 71 (2017)
8. Das, P., Zhu, M.O., McLaughlin, L., Bilgrami, Z., Milanaik, R.L.: Augmented reality video games: new possibilities and implications for children and adolescents. Multimodal Technol. Interact. **1**(2), 8 (2017)
9. LeBlanc, A.G., Chaput, J.P.: Pokémon go: a game changer for the physical inactivity crisis? Prevent. Med. **101**, 235–237 (2017)
10. Kaczmarek, L.D., Misiak, M., Behnke, M., Dziekan, M., Guzik, P.: The Pikachu effect: social and health gaming motivations lead to greater benefits of Pokémon GO use. Comput. Hum. Behav. **75**, 356–363 (2017)
11. Zach, Florian J., Tussyadiah, Iis P.: To Catch Them All—The (Un)intended Consequences of Pokémon GO on Mobility, Consumption, and Wellbeing. In: Schegg, R., Stangl, B. (eds.) Information and Communication Technologies in Tourism 2017, pp. 217–227. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-51168-9_16
12. Urry, J.: Mobilities. Polity Press, Cambridge and Malden (2007)
13. Urry, J., Larsen, J.: The Tourist Gaze 3.0. SAGE (2011)
14. Urry, J.: Consuming Places. Routledge (2003)
15. Faccio, M., McConnell, J.J.: Death by Pokémon GO: the economic and human cost of using apps while driving. J. Risk Insur. 1–35 (2019)

Individuals in Data-Driven Society

Individuals in Data-Driven Society

Jani Koskinnen

University of Turku, Turku School of Economics, Turku, Finland
jasiko@utu.fi

We have entered into an era of new colonialism: data colonialism, which has normalized the exploitation of human's through the collected data form individual. However, we should resist against this phenomenon building societies based on total algorithmic control, where we are lessening the humans to a mere resource for economic purposes. Resisting, however, does not mean the rejection of data use and collection, but rejecting current data practises [1]. We need a more humanistic approach to see the people behind this "information machine" that has forgotten human as an end to themselves and treats them as means only – an issue that Kant sees as the central threat to ethical actions in his second formulation of the categorical imperative [2]. The humanistic approach needs to be implemented through the whole society including individual, business, and governmental level to make a difference.

In their paper, Vildjiounaite et al. presented that the technology – IoT-Based Team Barometer – can be used in companies to improve the wellbeing of people in their working life. This paper presents well, that technology can be used for good, even though there may also be unexpected effects and not so desirable consequences, found by some papers in this track.

Heimo and Kimppa consider the problem that our freedom is limited more and more by the non-democratic way of corporate actors in their paper "Public Fora and Freedom of Expression." Corporations create products such as social media, used by the masses, while corporates gain in the situation where they can decide how their products are used and what kind of information can be shared via their products. We have ended the situation of private censorship that already is limiting the freedom of speech by fear of corporations, governmental actors, and other members of society. Likewise, this private censorship is a black box; we do not know what is censored, how it is censored, and why it is censored. This can hardly be in line with the values of our society.

Vuorinen and Bergroth analyze the other timely phenomenon of self-tracking as an illustration of the transition from Michel Foucault's discipline society to the Gilles Deleuze's control society in their paper which shows the less positive side of the technology. The following quotation is well descriptive of the differences between discipline society and control society:

> "*In the discipline society, the institutions were the spaces of transformation. In the control society, the space of transformation is within the self and is mediated by the codes of control.*"

The problematic issue that Vuorinen and Bergoths note in the paper is that self-tracking creates an image of ourselves that seems to be fragmented and has built-in imperative to examine our performance. This leads towards a troublesome worldview where everything (or everyone) needs to be evaluated and improved, and thus constantly are seen as deficient, as there is always something to be improved upon, and we do not meet the expectations laid upon us.

Rantanen and Koskinen focus on what the opinions and values of individuals have towards the data economy in their paper. It seems that people desire to gain control over their personal data, but also demand more clear and honest communication from the practitioners of data industry. It seems current practices on the data economy are not transparent or truly person-centric from an individual viewpoint.

Papers of this track illustrate that we need to give attention to the kind of society which is built upon current data use. Are we going to end up with data colonialism and lessening humans for economic profit? Or is there a possibility of finding a new path where ethical values and wellbeing of people are reached not only as an end for themselves but also as a means for resilient business likewise?

References

1. Couldry, N., Mejias, U.A.: Data colonialism: rethinking big data's relation to the contemporary subject. Television & New Media, **20**(4), 336–349 (2019)
2. Kant, I.: Groundwork of the metaphysics of morals (1785)

Humans of the European Data Economy Ecosystem - What Do They Demand from a Fair Data Economy?

Minna M. Rantanen(✉) and Jani Koskinen

Turku School of Economics, University of Turku, Turku, Finland
{minna.m.rantanen, jasiko}@utu.fi

Abstract. Personal data has become a commodity that can be used to create more value and growth in business within the data economy ecosystems. Although individuals are part of these ecosystems, they and their needs and demands are often neglected in the research and practices. In this study, we thematically analyse the demands of Europeans from Finland, France, Germany and the Netherlands (n = 4,792). The results show that the Europeans demand more transparent communication and ways to be active parts of the European data economy in the order it to be fair. We suggest that there is a need for more transparent deliberation in order to grant the humans of the European data economy what they want so that it can be fair.

Keywords: Data economy · Ecosystems · Europe · Human-centric · Demands · Thematic analysis

1 Introduction

Data - especially personal data - have become a commodity that can be used to create more value and growth in business. This development has lead to the development of data economy ecosystems which rely on trading and analysing personal data. Data economy ecosystems are networks that are formed by different actors using data as their main source. Actors and stakeholders of these ecosystems are indirectly and directly connected. Actions of the actors within the network are guided by official and unofficial rules such as legislation and norms [1].

In the past, there has been a lot of discussion and research about the different forms of data economy ecosystems. Big data, open data, governmental data, small data, and personal data can all have their economies, but can and often are connected in some form or way [2]. Whereas different types of data and technical solutions of data economy ecosystems have had attention [3], there is still a lack of a more holistic view.

However, this more holistic view has still left the individuals in the mere role of data subjects [4]. Thus, they are not seen as active actors of the data economy ecosystems but as sources of different kinds of data that can then be used to create more business value. Individuals, however, are not in an inactive role in data economies due to the pervasiveness of technology that allows monitoring, measuring and analysing personal data [5, 6].

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 327–339, 2020.
https://doi.org/10.1007/978-3-030-62803-1_26

People also have become increasingly aware that tech companies collect and distribute their data with third-party data brokers and advertisers. People have become more aware that their data has value and that they should have the right to control them [7]. Data is also seen more than ever as the property of the individuals - although legally it rarely is the case [8].

Personal data analytics hold promises for many benefits, but it has been argued that the current personal data economies are also enabling a new form of digital enslavement since they diminish the autonomy of individuals and societies at large [9]. Thus, there can be larger impacts than just endangered privacy of an individual. For example, in recent years micro-targeting of social media users has been used for influencing opinions, especially in politics [10].

Thus, individuals are often both sources and targets of the data economy ecosystems, although they are rarely seen as active actors of data economy ecosystems. This does not seem fair. However, fairness often is a subjective judgement when there are no established norms about what is right or wrong. Thus, we should consider what is seen as fair in the context of data economy ecosystems by focusing on individuals. Hence, in this paper, we focus on the individuals and their demands from fair data economy ecosystems. Acknowledging individuals' needs is crucial to the future of data economy ecosystems since viable personal data economy ecosystems rely on the cooperation of individuals and organisations.

In this paper, we focus on the empirical research of the demands and needs of the individuals in the context of the fair data economy ecosystem. We limit this research on Europe since the European Union is attempting to enforce the human-centric perspective as part of their data economy ecosystem. [11, 12]. Thus, or research question is:

RQ: What do the Europeans demand from a fair data economy ecosystem?

To answer this question, we analysed the open responses of a survey regarding a fair data economy conducted by TNS Kantar for Suomen itsenäisyysrahasto (henceforth Sitra). The data was collected from Finland, France, Germany, and the Netherlands. In total, the survey received 8,004 responses. In this paper, we analyse 4,792 open answers regarding fair data use and minimum requirements for fair data label using thematic analysis [13] to find out what people would demand from data economy ecosystems.

This study contributes to the emerging field of human-centred data economy ecosystems and provides insight into the needs of the individuals through their demands. It also gives an insight into the Europeans view on what a fair data economy ecosystem would be like and paves the way for future research about fairer data economy ecosystems.

The rest of this study is structured as follows. The next section shortly introduces related research about the European data economy. In the third section, the research process is presented, and it is followed by the results section. In section five, the results are discussed further. Finally, conclusions are presented in section six.

2 Background

The data economy is the area of growing interests in research and business fields. Likewise, the governments are giving more and more attention to the data economy and noted the need for regulation and policies for it. Especially, European Union has given strong focus on data economy and use of information in the Europeans single market strategy [14], data portability in GDPR [12], and free flow of non-personal data in the Regulation (EU) 2018/1807 [11].

We are focusing on EU as it has been the forerunner in its aims and strategies for the human-centric data economy. European Commission [15] stated in the data strategy that:

> "*The objective of the European data strategy is to make sure the EU becomes a role model and a leader for a society empowered by data. For this, it aims at setting up a true European data space, a single market for data, to unlock unused data, allowing it to flow freely within the European Union and across sectors for the benefit of businesses, researchers and public administrations. Citizens, businesses and organisations should be empowered to make better decisions based on insights gleaned from non-personal data. That data should be available to all, whether public or private, start-up or giant.*"

Despite the hype, the data economy is still a new research field that lacks the consistent terminology and does not have established the clear boundaries what belongs to the data economy. Thus, we need to define, what we mean by data economy and what it is this context.

Data economy as a concept is misleading since it is often used to describe a system of humans and technologies rather than an abstract economy. Thus ecosystem metaphor is often used to describe its ever-changing form and balance. Oliveira et al. [4] describe data ecosystems as *"socio-technical complex networks in which actors interact and collaborate with each other to find, archive, publish, consume, or reuse data as well as to foster innovation, create value, and support new businesses"*. This definition reveals the multitude of actions in a data economy and is our used definition here.

However, in this paper, we focus our research on personal data and use of those in the data economy and thus do not consider non-personal data. By personal data, we refer to "*any information that relates to an identified or identifiable living individual*" as defined by the European Commission [16]. The data is not personal if it is anonymous and cannot be traced to individual anymore. However, if the person can be re-identified, it remains personal data. This is problematic as in many cases anonymous data can be traced to individual and thus should be considered as personal data [17]. However, this is out of the scope of this paper and thus not considered further. It must be noted that we use the term personal *data* since it is commonly used, although it would be more accurate to talk about personal information.

Gathering personal data should rely on people's willingness to disclose them. If data is not shared, then it cannot be analysed, and the potential value is not actualised. Thus, we need to understand the individuals in the context of personal data economy ecosystems.

In this paper, we consider this issue from the perspective of fairness. Fairness is the quality or state of being fair, especially referring to impartial treatment [18]. However, what constitutes as fair treatment in a data economy ecosystem is biased by the focus on the business side of data economy ecosystems. Thus, there is a need to evaluate what the individuals whose data is used judge as fair. In other words, what they demand from a fair data economy ecosystem. Thus, to develop and govern a fair data economy ecosystem, there is a need to understand the perspective of individuals.

3 Research Process

Empirical material for this research was collected by Kantar TNS for Sitra. They conducted an online questionnaire in Finland, France, Germany, and the Netherlands in late 2018. The questionnaire aimed to clarify the use of digital services and attitudes towards the collection and use of personal data amongst Europeans. Focus on the questionnaire was on how individuals experience the potential use of their data from the perspective of data protection and privacy.

The survey included four sections: 1) background information, 2) rights in relation to data, and attitudes towards terms of use and privacy settings, 3) trust towards service providers and increasing trust, and 4) disclosure of information and its management and the concept of fair data service. The survey included 23 questions, of which two were open questions. Additionally, one of the questions included a possibility to clarify the answer. In total, there were 8,004 respondents who answered the open questions 4,792 times.

Since we focus on demands on towards a fair data economy, we focused our research questions about this particular topic. Demands for fair data economy were addressed in following open questions (n = 4,792). The questions are:

Q21: *Service providers collect a lot of data of you. In your opinion, how should this data be managed for you to feel that it is fair for you?*

Q23: *If services that use personal data would have a "fair data" label, what would be the minimum requirement for it?*

This study continues the work previously done by the authors. Rantanen [19] studied the values of the Europeans in the context of the fair data economy. She noted that in Finland, France, Germany and Netherlands, individuals value autonomy, protection their privacy, security, transparency, trustworthiness, benevolence, and justice. During this analysis, it became apparent that there are cross-cutting demands for a fair data economy in this data set.

In this paper, we focus on demands towards fair data economy by using qualitative analysis of the open answers. Respondents are identified in the original set by country and an ordinal. These identifications are utilised in this analysis as well. The full data set is available: https://www.sitra.fi/en/publications/use-digital-services/.

The research process followed the basic steps of thematic analysis [13]. First, the authors familiarised with the data set and then identified and coded the themes in each answer. After initial coding of the data set, the authors discussed the codings and similar themes were combined. Nvivo Pro 12 was used in the coding.

4 Results

Thematic analysis resulted in seven distinguished themes:

- User's control over data and data sharing
- Transparency and being informed
- Security
- Trust and fairness
- Compensation or benefits for users
- Supervision and rules
- Negative attitudes towards data collection and data economy

It is noteworthy that these themes are partly overlapping. In the following subsections, these themes are presented and analysed in further detail.

4.1 User's Control over Data and Data Sharing

User's control over data and data sharing are often seen as qualities that are demanded from a fair data economy. Respondents from all countries highlighted that they should consent to the data collection or at least have some power over data sharing.

> *"To ask the user beforehand for their informed consent (and every time it is useful), to clearly explain how they are used and what for, to clearly lay out which actions to take to amend or delete them, to mention how long they are stored for."* (Q23, FR365)

As it becomes apparent from the answer above, there is a demand for not only asking for permission to collect and use data but also to have control over it, and it should be service providers' responsibility to explain how it could be done clearly. Some respondents elaborated the idea that there should be mechanisms beyond consent:

> *"Data security and appropriate use of data is primary. Similarly, information about who is collecting and what data about it should be easily available, what it is used for and how I can influence the volume and nature of the data collected with my own choices and have it erased if I want."* (Q21, FI610)

Erasing the information and correcting it when necessary, naturally requires a possibility to see what is collected. Respondents did not limit the power over data sharing to the services providers that they are in direct contact with, but also stated that they should also decide whether or not the data are shared or sold to third-parties.

> *"It should guarantee that the consumer has access to editing their own data, erasure, etc. It should also guarantee who it discloses data to, and to my mind, it might also be good to have automatic approval as an option, as well as manual approval, so that you can decide which third parties you want your data to be disclosed to."* (Q23, FI474)

Thus, the is a demand for not only to mechanisms of viewing and controlling data but also transparency trough out the data economy ecosystems. Respondent GER251 summaries this demand as follows:

> *"It [fair data collection] should happen in a transparent way and I should have the opportunity at any time to undertake changes or deletions. And the provider must always specifically get my permission if he wants to pass on the data to third parties."* (Q21, GER251)

Being able to give informed consent for any data sharing, the possibility to edit and remove information, supervision of personal data and its correctness tie seem to tie into preserving one's autonomy and privacy in a data economy ecosystem. It is also apparent that the individuals see their information as their property and thus, feel that they should be able to control it.

4.2 Transparency and Being Informed

Whereas the previous theme was about the need to gain control, this theme is about being aware of what is happening to the personal data in a data economy ecosystem. Naturally, this theme overlaps with the previous one by being an enabler of power. However, the responses that expressed a demand for transparency and being informed were common. Especially in France, the need to being informed was more often mentioned than a need to control data.

> *"That each provider collecting our data asks for our opinion and explains clearly and simply what they are being used for."* (Q21, FR253)

Many of the respondents wished that the purpose of data collection were more comprehensible in general and expressed their frustration towards current ways of communicating:

> *"Open and clear, that is, saying in plain language which data and for what purpose it will be used. Very many data protection approval policies are complicated legal jargon that you have to read many times to really get the essence. These should be expressed more clearly and concisely."* (Q21, FI60)

Clearness and easiness of being aware of the personal data that is collected were also often seen as something that is not currently seen as sufficient or fair. Many of the respondents wished for better usability in the ways of obtaining their data:

> *"First, currently, the statutory information about the use of your own data is often ultimately difficult and bureaucratic to obtain. This should be made considerably easier. Second, the fact that I have to review long lists of different purposes of data use and amount of advertising on each site I visit is far from easy and user friendly. It feels like it was intentionally made as difficult as possible so that as many as possible would just automatically accept everything. A central, neutral and non-commercial service in which I can specify at least broader guidelines on what I want to allow in terms of the use of my data would make it a lot easier."* (Q21, FI166)

Informing users about changes is also seen as part of transparency. Thus, transparency is seen as a continuous action, not just static statements that are occasionally approved:

> *"Open and transparent towards the clients, that at all times you can gain insight into your data and can remove them, that they never share your information with third parties without your personal consent, that they don't just change the conditions without informing you about it."* (Q23, NE269)

Thus, transparency and being informed are inherently linked with understandable information about personal data use. This information is crucial in making decisions about whether not people consent to give their information to the service providers and thus, in the use of data economy ecosystems.

4.3 Security

Respondents from all the countries highlighted security and/or privacy as a feature of the fair data economy. It is clear that data security—protecting data from unauthorised access and data corruption—is and should be sufficient in any technological solution. Many of the respondents just stated that the requirement for the fair data economy or fair data label is that the system is secure.

Anonymity was often seen as a means to security, although there was little variance in opinions about what kind of information should be anonymous. Type of information that should be anonymous varied a lot from names, phone numbers, and addresses to medical and financial data.

> *"To commit to store them securely, to not share any sensitive information (address, phone number, medical and financial data for instance) and to really delete any data deleted by the user."* (Q23, FR381)

Also confidentiality and not sharing the data forward intentionally or unintentionally were often mentioned.

> *"Confidential data is not disclosed to other parties without permission, and no access for inappropriate people."* (Q21, FI318)

Some respondents also demanded absolute security without any possibility that data is leaked, which unfortunately is an impossible promise. However, some took into account the possibility of information leaks and called for honest communication in these cases.

Wishes for restrictions to the times that the data is kept and where and by whom they are analysed were also commonly seen as a security measurement. Many requested that the data should be stored only limited time, and it should have a clear purpose.

> *"[D]ata should be kept for a limited time, a minimum period of time for specific reasons, then regularly deleted."* (Q21, FR171)

Some of the respondents note the possibility of human errors as part of security, which in some cases, also means that people would only want their data to be handled via technical solutions, such as artificial intelligence. In contradiction, others stated that their personal data should only be handled by human beings. In general, answers incorporating security were vague and showcase that the understatement about data security varies a lot despite it being seen as an important aspect of the data economy.

4.4 Trust and Fairness

Trust, fairness and other issues concerning responsibility were often mentioned in the answers. Trust and trustworthiness were generally seen as keeping one's promises. Promises and knowing that they are kept require transparency since without it cannot be known what is promised.

> *"[Being] trustworthy is keeping promises and not amending the terms and conditions in the middle of everything."* (Q23, FI271)

Adjectives used to describe fair data management were plenty. Honest, ethical, moral, respectful and appropriate were often used without further explanations:

> *"To sell information ethically and not first and foremost to make money."* (Q23, FR457)

> *"With the greatest respect. Not for the services, apart from the usually obligatory need to give my data."* (Q21, NE560)

Some respondents highlighted the fair treatment of the employees in data economies. Likely, a fair data label was intuitively linked to the fair trade label. Nevertheless, fair treatment of all parties – including employees – was seen as an important part of the fair data economy.

> *"No child labour, no wages that are abnormally low, employees being treated well."* (Q23, NE223)

Answers within this theme describe how the data economy should be to be trustworthy and fair. However, the answers give little insight into how these aspects can be actualised. It seems that the Europeans want data economy ecosystems to be just, but how the just behaviour can be incorporated into the data economy ecosystems and communicated to users is still fuzzy. Nevertheless, people demand ways to assess the trustworthiness and fairness of the practices. This is not possible without transparent communication.

4.5 Compensation or Benefits for Users

The respondents were generally aware that their data has value and requested that they should also benefit from enclosing it. Some requested monetary compensations, such as a respondent from Germany:

> *"I want to be informed every time MY data earns money, and additionally, I would like 50\% of the money earned with my data transferred to my account!"* (Q21, GER196)

Most often, the respondents did, however, request other kinds of benefits than money. Benefits suggested varied greatly from premiums to improving quality of life to better personalisation.

> *"Data should be managed to know me better, and therefore, help improve my quality of life."* (Q21, FR87)

Better personalisation of advertisement was also often seen as a possible benefit for disclosing data. However, current personalisation was seen insufficient.

> *"I benefit personally, for example, personalised offers and advertisements. While the targeting of advertisements is possible, I still get car and nappy advertisements even though I don't have a driver's licence or children."* (Q21, FI233)

Respondents also noted services that collect data should be free of charge and that a fair data economy should not demand monetary investments from the users. However, using freeness as a way to justify collecting personal data and benefiting from it was not seen as sufficient practice:

> *"It is fine if money is earned with the data. It is not okay to make out you are offering a free service."* (Q21, GER239)

Although the majority of respondents mentioning benefits demanded benefits for themselves, some highlighted the common good instead of the hedonism. These respondents seemed to feel that it is fair to use their data if it benefits societies and/or humankind instead of seeing data merely as a business resource. For example, scientific research was often mentioned.

> *"To make the world a better place but not for your own interests."* (Q21, NE498)

Thus, it is clear that people understand the value of their data and want something in return for disclosing them. Despite the varying ideas about those benefits, it seems again that the benefits should be transparently and honestly communicated to the people.

4.6 Supervision and Rules

Some respondents highlighted the supervision and/or obedience of rules as the main issues of a fair data economy. These respondents noted that the actors should follow the laws and regulations. Also, several respondents noted that the institution supervising a data economy ecosystem should be an independent instance, such as a non-profit organisation and/or government.

> *"Checks by a government agency that is accountable to the politicians, for a customers' council, and in the public through the media."* (Q21, NE68)

> *"To have a central supervising organisation appointed by the government. It has branches that are specialised in data collection for healthcare, terrorism and other security, as well as statistics about education. These should supervise the private service providers that have to work according to strict guidelines."* (Q21, NE482)

Some respondents also noted that there should be sanctions when someone is not obeying the agreed rules.

> *"Hefty sanctions and compensation to users if it is found that misuse has taken place."* (Q23, FI92)

> *"[A]nyone who messes up is out, and it is published."* (Q23, GER7)

These kinds of answers showcase that the respondents have at least some trust in legal systems that regulate data economies. However, it must be many respondents described the ideal situation and not the current one. For these legal systems to provide the justice that the respondents describe, there is still a lot of work to be done. Again, it is apparent that supervision and compliance require openness and transparency from legislators and organisations practising in data economies.

4.7 Negative Attitudes Towards Data Collection and Data Economy

Although the majority of the answers were solution-centred and somewhat positive, some of the respondents were strictly against any data collection or pessimistic about the possibility of fair data economy ecosystems.

> *"It would be fairest if you were to not collect and use any data at all."* (Q21, GER262)

Some also showed pessimistic or negative attitudes towards the fair data label, mainly because they did not feel that it would have any effect on the data economies.

> *"It's all empty words. Empty words. It's about nothing. It all doesn't happen so fairly and it will not become so in the future either. Strategic considerations combined with wanting to make a profit, and that's how it will likely remain."* (Q23, NE175)

> *"This is a sham anyway. Nobody is prevented from having lucrative trade with data if there is a profit in it for him."* (Q23, GER479)

> *"Labels are not worth anything. Either they are fake or they are simply bought by companies who can afford to."* (Q23, FR177)

These negative and pessimistic attitudes are no surprise, but make it clear that there is distrust towards data economies. This lack of trust poses a challenge for data economies since, without it, it is likely that people will not take part in data economy ecosystems. Thus, there is a need to regain this trust, which could be achievable trough incorporating previous demands to the practices of data economy ecosystems.

5 Discussion

5.1 Findings and Implications

The main findings of this research are summarised in Table 1. As it can be noted there are two central issues that connect the demands behind the themes: transparent communication and respective active role of the individuals.

Transparency of data economy ecosystems is a central demand both in itself and as an enabler of the other demands. Without transparent communication, people cannot have true control over their data, be genuinely informed, aware of security, able to form trust, be aware of benefits or make assessments about the trustworthiness or obedience. Thus, transparency in all communications is needed, and it should be comprehensible to the users.

The other central demand is that people seem to want ways to be more involved and active in data economy ecosystems. They want to control their data, know what is happening in the data economy ecosystems, benefit from this participation and even participate in the supervision of these systems.

Table 1. Summary of the demands

Theme	Demands for a fair data economy
Control over data and data sharing	• Power over data sharing and content • Preserving privacy • Informed and dynamic consent • Possibility to control the content (requires access) • Easy-to-use control mechanisms
Transparency and being informed	• Transparency of processes • Ability to see what is collected • Comprehensible and on-going communication
Security	• Secure services • Confidentiality, anonymity • Transparent data management • Better understanding of security
Trust and fairness	• Trustworthiness of service providers • Clear promises that are kept • Fair treatment of all parties
Compensation and benefits	• Concrete benefits for disclosing data • Common good • Clear and honest communication about the benefits
Supervision and rules	• Rules and laws are obeyed • Supervision of the system by an independent institution • Sanctions if rules are broken
Negative attitudes	• Trust needs to be gained back • More holistic approach than a label needed

Currently, communication is seen as insufficient and in the worst-case scenario as empty promises that do not provoke any trust. Similarly, ways to influence the data collection and sharing can be limited to either accepting the incomprehensible end-user agreements or not using the services. Some cases you may alter the personal settings and control how information is used. However, this is not made user-friendly and laymans actually cannot gain meaningful and accurate picture of how their information is used and by whom.

Thus, in order to give the Europeans genuine ways to actively participate in the data economy ecosystems, we need to incorporate deliberative, participatory approach in data economy ecosystems and their practices. This approach should support the individual's possibility to influence the processes and practices of the European data economy ecosystems. Of course, there is a difference between intention to participate and actual participation. Nevertheless, there should be a possibility to participate and have an influence if one wishes so.

The deliberative approach is also needed to fight against the negative attitude towards data economy that is most case based on the justified lack of trust and missing possibilities to affect how data economy ecosystems work. People can gain more control over the use of their information by increased knowledge and improved possibilities to affect if deliberative decision making is integrated into data economy ecosystems.

5.2 Limitations and Future Research

Naturally, there are some limitations to this research. First, the survey was conducted in fairly similar countries, which could have given a too unified picture of the demands of the Europeans. Generalising these results to apply to all Europeans is not possible. Thus, more similar research about the demands of individuals should be conducted in other countries as well to see if there is a pattern.

Second, the answers were often rather vague and thus, left room for interpretations. Because that, there is a need to conduct more in-depth empirical research about the topic to gain a more comprehensive picture of the demands and needs towards fair data economy ecosystems.

Nevertheless, this research paves the way for future research about human-centric data economy ecosystems. As the development of the human-centric European data economy ecosystem is actualising, there is also a need for more research on the individuals and their part in it. Additionally, there is a need for finding deliberative practices for human-centric development. Thus, this topic opens avenues for the plurality of topics that should be studied in order to achieve a fair data economy ecosystem.

5.3 Conclusions

This paper studied the demands of the individuals in regard to fair data economy ecosystems. From the 4,792 responses, we found seven distinguished themes that revealed that the respondents have the willingness to be more active actors in order to data economy ecosystems to be fair. The respondents expressed, for example, a demand to gain control over their personal data in theses ecosystems, but also demanded more clear and honest communication from the practitioners.

It is clear, that the current practices on the data economy ecosystems are not transparent or allow enough power to the individuals. In order to achieve a fair data economy ecosystem, there should be more transparent deliberation about the justified practices of data economy ecosystems. In other words, we should pay attention to the humans of data economy ecosystems.

References

1. Koskinen, J., Knaapi-Junnila, S., Rantanen, M.M.: What if we had fair – people-centred – data economy ecosystems? In: Proceedings of IEEE Smart World Conference 2019 (2019)
2. Thinyane, M.: Small data and sustainable development—individuals at the center of data-driven societies. In: 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K), pp. 1–8. IEEE (2017)

3. Oliveira, M.I.S., Lóscio, B.F.: What is a data ecosystem? In: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, pp. 74:1–74:9. ACM, New York (2018). https://doi.org/10.1145/3209281.3209335
4. Oliveira, M.I.S., de Fátima Barros Lima, G., Lóscio, B.F.: Investigations into data ecosystems: a systematic mapping study. Knowl. Inf. Syst. **61**(2), 589–630 (2019). https://doi.org/10.1007/s10115-018-1323-6
5. Lammi, M., Pantzar, M.: The data economy: How technological change has altered the role of the citizen-consumer. Technol. Soc. **59**, 101157 (2019)
6. Birchall, C.: 'Data. gov-in-a-box' Delimiting transparency. Eur. J. Soc. Theory **18**, 185–202 (2015)
7. van Ooijen, I., Vrabec, H.U.: Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. J. Consum. Policy **42**, 91–107 (2019)
8. Igo, S.E.: Me and my data. Hist. Stud. Nat. Sci. **48**, 616–626 (2018)
9. Chisnall, M.: Digital slavery, time for abolition? Policy Stud. 1–19 (2020). https://doi.org/10.1080/01442872.2020.1724926
10. Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., Serrano, J.C.M.: Social media and microtargeting: political data processing and the consequences for Germany. Big Data Soc. **5**, 2053951718811844 (2018). https://doi.org/10.1177/2053951718811844
11. Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union. Official Journal of the European Union. L305, 59–68 (2018)
12. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. L119, 1–88 (2016)
13. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**, 77–101 (2006)
14. European Commission: Building a European data economy. European Commission (2019)
15. European Commission: Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions - A European Strategy for Data, Brussels (2020)
16. European Commission: What is personal data? European Commission (2020)
17. Andrew, J., Baker, M.: The general data protection regulation in the age of surveillance capitalism. J. Bus. Ethics (2019). https://doi.org/10.1007/s10551-019-04239-z
18. Fairness. https://www.merriam-webster.com/dictionary/fairness
19. Rantanen, M.M.: Towards ethical guidelines for fair data economy – thematic analysis of values of Europeans. In: Proceedings of the Third Seminar on Technology Ethics 2019, CEUR-WS, pp. 43–54 (2019)

Public Fora and Freedom of Expression

Olli I. Heimo(✉) and Kai K. Kimppa(✉)

University of Turku, Turku, Finland
{olli.heimo,kai.kimppa}@utu.fi

Abstract. This paper focuses on freedom of expression in the age of digitalization specifically on social media public fora. As the current political climate worldwide is polarising and both democracies and dictatorships are tightening their grip on ‘hate speech’ and other ‘unwanted’ discourse, we as people should focus on what is being done with our freedom of expression. As it is obvious that not all expression can be free, e.g. child pornography, we should focus on both what expression we should limit, but also how: who should have the power to limit the expression and how should they do it? The argument of the paper is that the current corporation-lead black-box censorship in social media creates a danger for the freedom of expression as the corporations, often foreign for most people, seem to use the platform to affect the functioning of the democracies’ legislation processes. Thus, if we value our democracies, the censorship should be conducted in a justified and as open as possible manner and never without a sound reason.

Keywords: Freedom of expression · Social media · Ethics · Public fora · Censorship

1 Introduction

In this paper we discuss freedom of expression. Freedom of expression is a crucial part of a working democratic society as the citizen’s require the ability to share and receive information, express their needs and wants, and discuss on how to develop the society to be a better place for them and everyone else. (C.f. e.g. [1]) In this paper freedom of expression is examined from the viewpoint of political philosophy with the implications on who is limiting it, how, and why, not from the perspective of the law, as we are proposing basis for new law, not trying to explain how the current legislation functions. It is important to notice that the discussion presented here is not only about freedom of speech, but of expression. An expression of art – for example a modern dance against the totalitarian regime of the People’s Republic of China – can and must be seen as valuable as verbally pronounced critique against the said ‘republic’. When this ability for expression, or receiving the expression is limited, it should be done with utmost care due the consequences. As the society can be seen as a struggle between tyranny and liberty (cf. [1]), freedom of expression is a tool to keep the tyrannical tendencies in governance at bay.

In the current era of digitalization, expression of opinion has been turning increasingly towards social media and Internet in general. Internet has become a hotbed for political discussion. Whereas the President of the United States used to declare

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 340–350, 2020.
https://doi.org/10.1007/978-3-030-62803-1_27

information through press conferences, the trend has been turning more and more towards Twitter. Since 2017 Tweets made by The President of the United States have been considered official statements [2]. And the President of the United States is not alone in this, but government actors and politicians all around the world, as well as corporations, celebrities, and private citizens announce their actions, express their opinions, and communicate with each other through social media. Digitalization has made an era where a child from a developing country can comment US president's post – and be heard; possibly. And a child from Sweden is already having an on-going discussion (of sorts) with US president; and is heard.

In social media though, not all the participants are equal, centralization is a clear issue. When checking the statistics in Fig. 1, it is clear that social media platforms differ in their user-base.

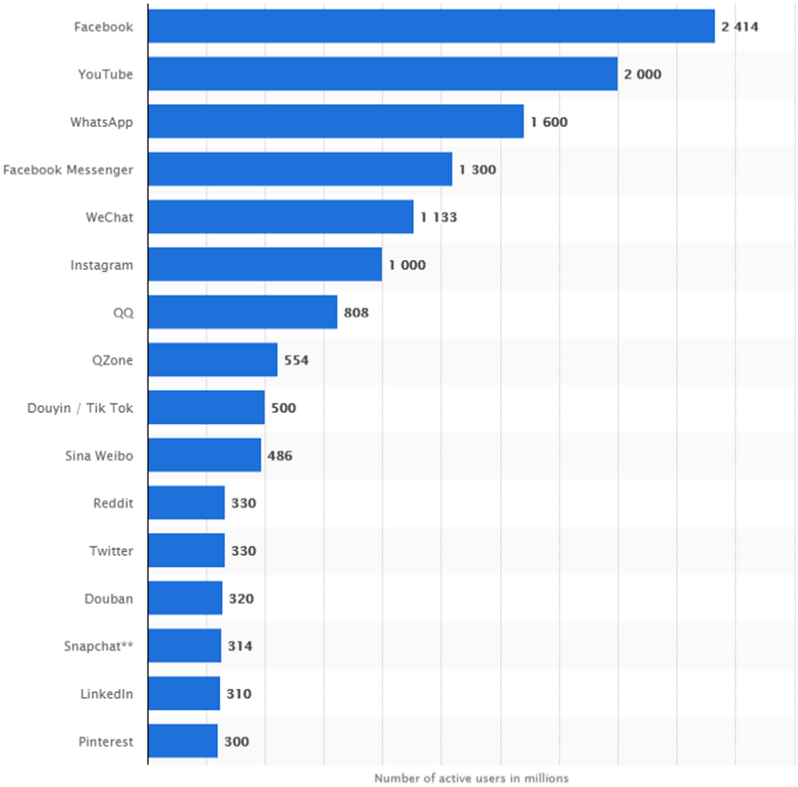


Fig. 1. Active users by social media platform [3]

The impact of video platform sizes is somewhat hard due the issue to calculate, but as an estimate Statista, a German online portal for statistics, declares the seven biggest video sharing sites in United States to be YouTube (90% of Internet users reached), Facebook (60%), Instagram (by Facebook) (35%), Twitter (21%), Snapchat (18%), CNN (17%) and Fox News (16%) [4], the last two being the only sites to have an editor-in-chief. Globally, compared to the second best video-only sharing site Vimeo, YouTube has over 2 billion users log in and watch videos on YouTube every month, whereas Vimeo has just 973,000 registered users in total [5].

Hence it is clear that in both video sharing and overall social media, centralization of the Internet is true. Moreover, one could clearly argue that with that amount of centralized information sharing, it would make quite an impact socially if someone were to decide on what gets to be shared and watched, and by whom.

2 Censorship

Censorship has traditionally been connected to state censorship. This is due to the state having been the only power capable of ordering large scale pre-censorship and punishment for publishing materials seen to be worth stopping from circulation amongst the people. Of course that has led to self-censorship (or publishing under a pen name or entirely anonymously) already in these situations, as the potential authors (or painters or musicians) have typically been quite aware of what might happen were they found creating – or even in possession – of forbidden materials; this ranging from fines through prison sentences or deportation all the way to being burned as a heretic. These days, however, a new form of major censorship has become an even more pressing problem, as the power of the state as a censor has waned, especially in liberal democracies; namely censorship by the corporations.

The Internet opened the possibilities for us to express ourselves enormously; anyone could build a web page, send messages to message boards, talk in chat software – and as of late, do all these in one platform or another such as Facebook, Twitter, LinkedIn, Instagram or similar, either internationally or regionally. However, as these platforms have centralized, and in many cases there are no optional platforms, but rather a monopoly on specific areas has become the norm, Twitter for short messages, Facebook for longer discussions, Instagram for pictures (owned by Facebook) or WhatsApp for replacing text messages and small groups (also owned by Facebook), YouTube for video (owned by Google) etc., their community standards and wild interpretations of laws on different countries or regions are threatening freedom of expression. Not always the freedom to say what one thinks, of course you can go on your own web page and say things, but rather the freedom to receive communications, as more often than not, those aforementioned webpages are just not visited by many, if any. The same is true if one tries to migrate to a smaller, less used platform, as most of ones friends or followers are unlikely to do so. Thus, the point of censoring has also shifted from the state (although it is still indirectly there) to the corporations, which are practically fully outside of our democratic control.

Merriam-Webster dictionary defines censoring being "to examine in order to suppress or delete anything considered objectionable". Censorship itself however can be done in several different ways. In this paper we examine three simple ways to limit the freedom of expression:

1. to prevent or suppress the expression to reach the audience,
2. to delete the expression after the expression has reached the audience, and
3. to prevent or suppress the publication of the expression before it is expressed.

Prevention or suppression of the expression to reach the audience requires active censoring and control over the media the expression can be delivered upon. Hence, only big actors – e.g. governments and large corporations – can effectively do it. Typically there is also co-operation (voluntary or involuntary) with the public and the private sector.

Deletion of the expression is of course the most inefficient way to censor because the expression, when published, can be copied easily and to fully censor an expression all the copies must be gathered to the deletion (e.g. late music video Knebel by Lindemann which was censored at YouTube, of which the uncensored version is available at e.g. Vimeo). This, however often leads to less access to the content, as it must be presented on alternative media platforms, which are typically not as easily censored, but also less used in general and also less visible in search engines; "if it is not on the first page of a Google search, it does not exist", as the saying goes.

Third and most efficient method of censorship*, preventing or suppressing the publication of the expression in the before it is even expressed* works with the fear of consequences tied with the act of publishing an expression deemed 'unpublishable'. These consequences include (but are not limited to) legal punishment, social stigmatization, or denying certain services or privileges.

Whereas all these alone can be effective by themselves, the most efficient method seems to be to combine these methods, for example actively censoring certain content, deleting all copies mentioning that content, and retaliating against those who have a copy or spread it. Destroying most copies of a book is not nearly as effective as destroying most copies and making the possession of the book a crime.

To emphasize the unwillingness to publish an expression even more effective is to make it unclear *where the limits of acceptable expression are*. When the boundaries are clear, it is easy to understand whether the publication of the said expression legal. As an example, it can be made quite explicitly clear where an expression is child pornography, as Interpol report [6] states:

> *"Child pornography is the consequence of the exploitation or sexual abuse perpetrated against a child. It can be defined as any means of depicting or promoting sexual abuse of a child, including print and/or audio, centred on sex acts or the genital organs of children"*[1]

Moreover, if one is not sure is one's piece of expression child pornography when compared to this statement, one should clearly not publish – or make! – it in the first place; if not for the sake of legal consequences, then for the sake of the children and their rights. But *when the definition lacks clarity* and there are no clear precedents available, the act of self-censorship called 'chilling effect' is likely. In these cases the publisher of the expression does not have a clear idea whether or not a certain

[1] It is worth to note however that there is discussion on virtually created child pornography and whether it should be banned or not, as there are no directly harmed parties if it is virtually created. Indirect harm may be strong enough for a case, but there is no consensus on this as of yet, and some countries (e.g. Japan) allow creation and possession of such material, where others (e.g. Sweden) do not. This discussion, and especially the ethical implications around it, however is one the authors do not wish to pursue further due the social stigmatisation and the chilling effect created around the subject.

expression is illegal and therefore – just for the safety of it – refrains from publishing it. Or, as Open Net Initiative [7] web pages explain:

> *Another common and effective strategy to limit exposure to Internet content is by encouraging self-censorship both in browsing habits and in choosing content to post online. This may take place through the threat of legal action, the promotion of social norms, or informal methods of intimidation. Arrest and detention related to Internet offenses, or on unrelated charges, have been used in many instances to induce compliance with Internet content restrictions. In many cases, the content restrictions are neither spoken nor written. The perception that the government is engaged in the surveillance and monitoring of Internet activity, whether accurate or not, provides another strong incentive to avoid posting material or visiting sites that might draw the attention of authorities.*

In the current climate of 'hate speech' – either legally mandated or limited by the platform owners – more and more people self-sensor even thoughts and ideas that either are not 'hate speech', but can understood as such by fringe elements, or in fear of being a target of hate speech, such as doxing or threats (see e.g. [8]). At times these fears are of course also justified. As a banal example either being extremely frustrated or under the influence of mind altering substances it is likely not a good idea to publish ones thoughts, as it can too easily end up being hate speech, such as inciting or even promising to participate in actual violence (see e.g.[2]. This, however is not a justification to censor oneself from opinions which only differ from those of other discussants, yet, the chilling effect causes people to self-censor these comments as well, especially in environments where "wrong think" is punished either directly at the site or even outside the site by attempts to dox the person with "wrong" views, or affect their private or professional life.

Historical expressions quite acceptable at the time are also in the line of attack now that the legislation has changed; especially since the use of language has changed to add previously more-or-less neutral words to the category of ones now unacceptable. (See e.g. [8].) Therefore it seems that as the morals shift with the time, we are expected to start censoring our history – or at least warn people about the history which they should be aware. E.g. in Finland, a long-standing member of parliament and ex-minister from Christian Democrat Party is at the moment being investigated for publishing bible quotations in 2004 [9] according to a law which came into force 2011 [10].

As the legislation is often after crude language, chilling effect affects mainly people with lower education since they lack the linguistic tools to express themselves in formalized manner. This results in a problem with equality before the court as the critique towards different policies and groups of people often coincide with criminal activities or moral questions [8]. Whereas an educated person can argue the increased immorality, criminality, etc. of a certain group in an eloquent and thus also in a legal manner, the same idea delivered with lesser skills in wording will lead to blatant and illegal conclusion. For example:

[2] http://web.archive.org/web/20061021052645/thorntree.lonelyplanet.com/messagepost.cfm?postaction=reply&catid=32&threadid=776884&messid=6606542&STARTPAGE=1&parentid=0&from=1&showall=true)

1. "The increased amount of immigration will very probably raise the levels of criminal activity in the area due to socio-economical differences in the population groups."
2. "The immigrants will soon be robbing and raping us!"

Mill [1] points out that:

> *The peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error.*

Therefore, according to Mill [1], freedom of expression is also about being able to receive dissenting views and other expressions deemed uncomfortable by some.

Reference withheld for blind review [8] also point out that:

> *When people start censoring themselves to not conduct "hate speech", which is defined differently in different societies, anything can be determined to be hate speech. In Turkey critiquing government policies or supporting the Kurds can be considered hate speech; in Finland critiquing current immigration practices – or defending them – can be hate speech; in Russia promoting gay rights is hate speech – in the Netherlands critiquing them is hate speech.*

Therefore it seems that the legislation – not ethics – defines hate speech and legislation is observably culturally dependent, whereas ethics at least ought not be. Although ethics ought to define what is good and what is evil, what is right and what is wrong, and laws ought to reflect this, we know from empirical observation this is not always the case. Many laws in many countries are clearly unethical, examples abound from cruel and unusual punishment to punishing people about choices that affect only themselves; and of course, in the history of the humankind there is plethora of unjust laws. Let us see to it, that hate speech legislation does not become one of those laws.

3 Public Versus Private – Public Fora and Private Space

While the most efficient censorship done so far is by the government, private sector censorship has gained ground within the growth of the Internet due to the centralization of the Internet as discussed in chapter 1 (see also [11]). With this centralization the 'big players' have taken the key roles of delivering information, some by generating a lot of content, and others by creating a public platform – *a public forum* – for the user-generated content. Whereas content-creators such as Springer, Netflix or Washington Post thrive from the expressions **they produce or publish**, the public forum providers rely on the **users to express themselves** and then benefiting from the advertising revenue, monthly fees, and private information provided by the large user-base to be sold for various purposes, including but not limited to advertisers. This is a key difference. Some of these platforms, e.g. YouTube, reward the users from generating the revenue whereas others, e.g. Twitter or Facebook, do not. This also plays a role in the discussion.

As it is seen as normal in Western society that a traditional media – radio, newspapers, other print media, television, video renting companies etc. can choose what expressions they wish to publish and promote, the new media and the Internet has been seen as a ‘wild west’ of expression – anything goes, as long as it is not illegal. With the centralization and perhaps ‘stabilization’ of the Internet, many of the new media platforms are mainly benefiting from user-generated content to their various fora. The publisher has an editor-in-chief who is responsible for the content published by the publisher, in the public forum the responsibility currently lies with the users doing the publishing. Some limitations do apply however, as the forum upkeep is in some countries required to remove illegal content from the server within a reasonable time.

This difference has been noted by the US Appeals court, which in decision Knight First Amendment Institute vs. Trump 2017[3] noted that “[d]efendants’ blocking of the Individual Plaintiffs from the @realDonaldTrump account violates the First Amendment because it imposes a viewpoint-based restriction on the Individual Plaintiffs’ participation in a public forum.” In this case the president (a public sector actor) used Twitter’s (a private sector actor) platform (a public forum) function to limit the freedom of expression.

4 What Is Happening?

Various issues arise from the concentration of social media in very few hands and the polarisation of the users within these fora. These include (but are likely not limited in) demonetization, ending in bubbles which strengthen one’s world view without ever challenging it, hiding content from searches and not promoting it even though the user has expressed their wish to see all the content by the provider.

Demonetization is a complex phenomenon. First, it cuts the publishers’ income, when they are acquiring income via new media content monetization. Most of the money one receives from the content comes within the first few days. Thus, if one is demonetized, one practically always loses the main part of the potential income with the content even if it is remonetized later [12]. Secondly, in many cases (e.g. YouTube), one’s content is also ‘unlisted’, when demonetized so that the user must know where to find the content, and the content is not promoted [13]. From the perspective of censorship, this is a form of light-censorship, as unlisted media does not produce information about new content from the producer to the audience, neither to those subscribing to the videos or those hoping to get them as suggestions, and especially not to those who want to value their privacy by not logging into the media platform, for example, those using browser privacy-modes or deleting their browser history and cookies, or those using publicly accessible computers, such as library computers or computers in Internet cafes. Therefore, only those seeking new content from the producers by clicking their content directly or searching with exact keywords will find the unlisted content, while other, competing content is made easily visible before, during, and after viewing videos.

[3] Case 1:17-cv-05205 Document 1, available at https://assets.documentcloud.org/documents/3892179/2017-07-11-Knight-Institute-Trump-Twitter.pdf, accessed 17.2.2020.

If it is altogether too easy to choose one content over another as it is clearly the case here, it will bias the information received and thus might bias the world-view of at least some portion of people and therefore produces not only Internet bubbles where people do not see certain views, or only those accepted by the platform provider, but also previously mentioned ‘light censorship’ where the competing opinions cannot compete within “the marketplace of ideas”, as it was called by Justice Wendell already in 1919. Moreover, when those committed or acquainted with certain perspectives notice that information supporting their world view is being even slightly censored, it creates frustration, anger, and in the end, polarization, pushing them more deeply in their bubble and others against them. If the idea is to not polarize people but to get people to work together and find compromises, this kind of behaviour is contradictory to the intent. In the end it only strengthens the defensive lines and suppresses the will of the groups to understand each other.

In a public forum one should not use (at least closed) AI, or a black box version of an algorithm. We need to have public officials who verify which algorithms are used and how they actually work – they need to produce the same result with the same input; and it needs to be understood how. This is relevant to the democratic development of society as a whole, and the freedom of different ideas to compete in society, so that the best ideas can actually be found. If the algorithm is indeed a ‘black box’ or the rules apply differently to different people, there is a concern of treating people differently and thus promoting racism, sexism, or other forms of oppression. If the results change over time or due to the algorithm ‘learning’ what the user searches, the results will indeed bias the view of the viewer. Many – the authors included – try to avoid this in various ways, for example pressing “like” on interesting, rather than agreed with content, using aforementioned privacy modes and even other computers and computers at different places to see content that could raise new ideas; but unfortunately most people are unlikely to follow such practices to get a wider understanding of the topics, and rather end up in the bubbles the algorithms push them into. It is also questionable whether this even works in the long run, especially if the algorithms use IP address as an identifier; in this case even privacy mode might not help – or the content offered could be related to whatever those who use the same IP are interested in, not what the person searching for information is interested in.

One could argue that these are private entrepreneurs not the public sector. So why cannot a private actor choose what is published in their fora, with their money, with their name? The answer is twofold: first, it is about the authorship and its responsibility. Whereas a publisher of a magazine, television broadcast, video rental, or a WordPress web page is responsible on what is published, in a public forum the responsibility and accountability lies with those creating the content and not those providing the forum for publishing it. Hence when on a discussion board, whether it be an Internet forum or a public space in the physical world, the responsibility of what people publish is by those who upload the content, not those who supply the space for discussion. We should treat an owner of a YouTube account, Facebook account, Twitter account or any other new media account as the editor in chief. Interestingly, the US legislation might just support this (see e.g. [13]), but this is yet to be seen. Secondly, as the centralization of the Internet becomes more prevalent [14], the control over discussion will be moving more and more to the hands of the few when looking from the perspective of content producers. If only those few get a say on what is discussed, the consequences may be quite dystopian. As the control of the expression falls in the hands of few and if there are little rules dictating on how that control over others can be used, the one in control has quite free hands on what ideas and expressions are being discussed and promoted in the society as a whole. The idea of freedom of expression does not specify the participant ruling the consequences or applying the censorship, quite the contrary.

As the world is clearly moving to a state where wealth is more and more in the hands of the few, those few, even if they cannot now, will soon be able to acquire all the media platforms and thus dominate what we are discussing. And it can be argued they already can, as especially Google and Facebook (or their respective parent organizations) are concentrating this power in their hands. As Mill [1] already noted, to protect our freedom of expression now and in the future we must allow the freedom of expression forms we are not comfortable with and not allow anyone, neither the governments nor the corporations to use our differences of opinion in matters of immigration, religion, political views, economics nor anything else similar to justify their control over the public opinion. The stakes are just too high. And to protect us from tyranny of the few, we need freedom of expression.

If one has received substandard service in a government office, there are typically ways to find out why. In the Internet platforms, the censored typically get either no explanation, or such a vague explanation that it means nothing in relation to the perceived offence (see e.g. [15]). There are numerous examples on how public forum providers have censored political discussion. For a current example a Finnish Facebook group mensroom (miestenhuone) with over 50 000 members ($\sim$1% of all Finns) was censored probably due the promotion of a legal initiative to legalise cannabis [16] without any warning nor given information to the administrators of said group[4]. (see also [17]) Neither the users nor the moderators of the group got any information that they have been removed from the group, that the group had been dissolved, or any explanation why. The legislation initiative (also requiring over 50 000 supporters, just by coincidence) was accepted for government processing [17] and will be delivered for discussion in the Parliament of Finland. The initiative did not support the use of said drug, but to remove punishment of using it due to increased harm to the users caused by the current legislation. Facebook authorities seem to have interpreted this however as a pro-drug campaign even though it clearly was a discussion about the social consequences and the functioning of a justice system in a modern society. Hence it is clear that the policies of Facebook do not follow the legal system and juridis-ethical discourse in Finland, but instead try to influence that discourse with a moralistic mind-set. The censorship of course did not concern only the discourse on cannabis and its state of legality, but of all the discussions in the said group, which all were permanently deleted from the public due to this one legally sound ‘mistake [16]. It should be clear that legislation initiatives should be open for discussion and if they are denied by the reason that what is promoted to be legalised is illegal and hence should be not discussed it is hard point on the legal discussion. Moreover if the discussion is about the ethics it is clear where the platform controllers - in this case facebook - has a moral stand. As the

[4] Not clear due to no communication from the Facebook officials, but the timing fits.

discussion in this case - the case of legalising recreational use of marijuana - is indeed legal in many countries and areas, should this be legal discussion is clearly denying should this be a legal discussion is clearly a moral standpoint.

It seems, then, that Facebook acts in the legislative process by hindering (and possibly in other cases promoting?) the legislations it chooses to be morally sound. The question is why Finns or any other group of people accepts this? Moreover, why is this not an issue in more general sense on who gets to choose what is discussed, promoted, and censored?

5 Conclusions

As stated above, the Facebook group was possibly completely censored due to a single link to a government web-page where citizens were able to influence the legislation of their country. However, because Facebook, YouTube, Twitter, or any other public forum providers are responsible in answering the accusations or questions arising from their censorship, this is not certain. The first step towards more open society would be to make them accountable on the methods and decisions behind their censorship.

Whereas clear rules, open decision making, and public discussion are methods on developing an open society where people can without fear publish their opinion, learn from each other, and participate in the decision-making process concerning them and people around them, obfuscated rules, black-box (perhaps AI-based?) decision making processes, and fear of arbitrary punishment by powerful corporations, government actors, and other members of society alike, are all likely to lead to a society where only some opinions are accepted.

Freedom of expression is a tool that upholds democracy and civil liberties. Without freedom of expression we cannot function as a democracy but are limited by those limiting our expression. We should be careful about the government limiting our expression, and we should not give that power to private corporations either. These are powerful tools in the wrong hands. The public fora should remain public spaces for discussion and development of the society, not private spaces where someone with capital, political power, or loud enough supporters can choose what is discussed and how.

References

1. Mill, J.S.: On liberty. In: Courtney, W.L., L.L.D (ed.) With an Introduction. The Walter Scott Publishing Co., Ltd., London and Felling-on-Tyne, New York and Melbourne (1869, 2011). https://www.gutenberg.org/files/34901/34901-h/34901-h.htm. Accessed 17 Feb 2020
2. CNN: White House: Trump's tweets are 'official statements' (2017). https://edition.cnn.com/2017/06/06/politics/trump-tweets-official-statements/index.html. Accessed 17 Feb 2020
3. Statista: Global Social Networks Ranked by Number of Users 2020. Published by J. Clement, 14 February 2020. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

4. Statista: Leading online video platforms in the United States in November 2018, based on reach (2018). https://www.statista.com/statistics/266201/us-market-share-of-leading-Internet-video-portals/. Accessed 17 Feb 2020
5. Techradar: YouTube vs Vimeo. https://www.techradar.com/news/youtube-vs-vimeo. Accessed 17 Feb 2020
6. UNESCO: Sexual abuse of children on the Internet: a new challenge for Interpol, UNESCO (1999) https://unesdoc.unesco.org/ark:/48223/pf0000114734. Accessed 17 Feb 2020
7. OpenNet Initiative, About Filtering. https://opennet.net/about-filtering. Accessed 17 Feb 2020
8. Heimo, O.I., Naskali, J., Kimppa, K.K.: Hate speech recognition AI – a new method for censorship? In: ETHICOMP 2018, Poland, 24–26 September (2018)
9. Räsänen, P.: MIEHEKSI JA NAISEKSI HÄN HEIDÄT LOI - Homosuhteet haastavat kristillisen ihmiskäsityksen (2004). https://www.lhpk.fi/julkaisut/aamuntahdet/29_mieheksijanaiseksi.pdf. Accessed 17 Feb 2020
10. HE 317/2010 vp, Hallituksen esitys eduskunnalle Euroopan neuvoston tietoverkkorikollisuutta koskevan yleissopimuksen lisäpöytäkirjan, joka koskee tietojärjestelmien välityksellä tehtyjen luonteeltaan rasististen ja muukalaisvihamielisten tekojen kriminalisointia, hyväksymisestä ja laiksi sen lainsäädännön alaan kuuluvien määräysten voimaansaattamisesta sekä laeiksi rikoslain ja tietoyhteiskunnan palvelujen tarjoamisesta annetun lain 15 §:n muuttamisesta. https://www.eduskunta.fi/FI/Vaski/sivut/trip.aspx?triptype=ValtiopaivaAsiat&docid=he+317/2010. Accessed 17 Feb 2020
11. Hackernoon: The evolution of the internet, from decentralized to centralized (2018). https://hackernoon.com/the-evolution-of-the-Internet-from-decentralized-to-centralized-3e2fa65898f5. Accessed 17 Feb 2020
12. Business Insider 'Like you've been fired from your job': YouTubers have lost thousands of dollars after their channels were mistakenly demonetized for months (2019) https://www.businessinsider.com/youtubers-entire-channels-can-get-mistakenly-demonetized-for-months-2019-8?r=US&IR=T. Accessed 17 Feb 2020
13. Prager University vs. Google Inc. and YouTube, LLC.: Case 5:17-cv-06064-LHK Document 1, United States District Court, Northern District of California (2017)
14. Agre, P.E.: P2P and the promise of internet equality. Commun. ACM **46**(2), 39–42 (2003)
15. CGP Grey: YouTube vs Grey: a ballad of accidental suspension, 24 November 2019. https://www.youtube.com/watch?v=DIssymQvrbU. Accessed 3 Dec 2019
16. Iltalehti: Suosittu Facebook-ryhmä Miestenhuone poistui yllättäen – yleisö ymmällään (2019). https://www.iltalehti.fi/digiuutiset/a/fb68adb2-eca4-48f8-a088-52ff1359f8f5. Accessed 17 Feb 2020
17. Kansalaisaloite: Kansalaisaloite Kannabiksen käytön rangaistavuuden poistamiseksi (2019). https://www.kansalaisaloite.fi/fi/aloite/4360. Accessed 17 Feb 2020

Self-tracking, Power, and the Transition from Discipline to Control

Jukka Vuorinen[1](✉) and Harley Bergroth[2]

[1] Information Systems Science, University of Turku, 20014 Turku, Finland
juanvu@utu.fi

[2] Department of Social Research, University of Turku, 20014 Turku, Finland
harley.bergroth@utu.fi

Abstract. This paper examines self-tracking as illustrative of the transition from Michel Foucault's discipline society to the Gilles Deleuze's control society. These two forms of societies both relate centrally to the organisation and (re)production of power relations, but they organise space and time in different ways. Self-tracking refers to data-driven practices of self-monitoring by digital devices, and the practice is here treated as an individual subjective activity, in which the self subjects itself with the help of a self-tracking device. Importantly, our claim is that this subjectivation takes place in the broader context of the control society and its increasingly data-driven character, in which traditional institutional discipline is being replaced by in principle unbounded regimes of (self-)control.

Keywords: Self-tracking · Data · Power · Control · Discipline · Society

1 Introduction

Personal everyday use of self-tracking technologies such as activity trackers has proliferated in recent years across Euro-American societies. Self-tracking devices come in different forms, the most popular wearable devices being wristbands and rings and self-tracking technologies including applications such as sport trackers for mobile phones (Berg 2017; Crawford, Lingel and Karppi 2015; Lupton 2016). These devices observe different bodily functions, such as movement and heart rate, and use these bodily functions as input data for the device. Data is then algorithmically turned into various forms of continuous information flows that relate to the health, well-being and performance of the user. For example, many wearable self-tracking devices (e.g. Fitbit) allow their user to track the quality of daily activity and sleep based on the level of (non-)movement of the body and display this information in numbers and graphs.

Although self-tracking devices most clearly pertain to the obvious questions of physical condition, health and well-being – which can be seen as self-evidently desirable features of human life in cultural contexts that encourage self-responsibility and e.g. preventive and proactive self-care – it could be questioned whether coupling oneself with such devices ultimately enables "better" lives or increased self-awareness. In any case, self-tracking literally pertains to the socially constructed "self", and this paper argues that the relation to the self is a phenomenon that always emerges with and

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 351–360, 2020.
https://doi.org/10.1007/978-3-030-62803-1_28

in relation to the interactions and form of the techno-political social context. Furthermore, self-tracking is a practice of power in the sense that a self-tracking device is a technology through which the self is *subjected* to monitoring. In self-tracking, the self becomes the studied object of activity and well-being. In the course of this paper, self-tracking is thus treated as a method of control that implies the presence of power relations. To simplify, an individual action of self-tracking relates to and emerges through the societal and discursive contexts in which it takes place.

This article does not aim to map the current socio-political contexts and discourses that promote and propagate proactive self-tracking but rather focuses on the theoretical question of what kinds of power regimes self-tracking as an everyday data-driven practice of self-monitoring relates to. More specifically, we map the concepts of the "discipline society" (Foucault 1995) and the "control society" (Deleuze 1992) and employ them in order to analyse self-tracking as a regime of power. We pay particular attention to the different concepts of time and space in relation to discipline and control.

First, the Foucauldian idea of power is introduced, as he developed the idea in relation to the concept of discipline. We then highlight the differences between Foucauldian discipline and Deleuze's (1992) control society and apply these terms to the employment of self-tracking technologies. Lastly, we make the point that self-tracking may be observed as a practice of transition from discipline to control.

2 Discipline Society and Control Society

2.1 Foucault's Concept of Power

Michel Foucault (1926–1984) was a French philosopher who was interested in the systems of thought that are actualised through discourses. He examined power (explicitly and implicitly) in his many writings (e.g. Foucault 1990, 1995, 2002, 2003). In the early stages of his academic career, Foucault's concept of power was negative in the sense that Foucault analysed power in restrictive and repressive terms. For example, the confinement of mental patients can be seen as a restrictive act that confines freedom (Foucault 2003). However, later in his career, Foucault (1982, 1990, 1995) started to emphasise the positive and productive aspects of power. In a positive sense, power does not hinder, restrict, repress but rather works in constructive and productive ways: it produces activity, supports and enables its subjects. In other words, for Foucault power suggests, proposes, encourages and praises, rather than restricts, bans or forbids. For example, there is a positive aspect in the power of social media platforms. Social media networks encourage their users to create updates and to upload new content. Many health and well-being related technologies and apps, such as Sports Tracker, likewise provide the possibility of sharing the user's results on social media networks, and as such work through positive power by connecting users to digital means and spaces for sharing. Sharing in turn re-establishes the position of the social media platform as a social space. Importantly, social media platforms only survive because of their users: without the users and their will to share, social media would no longer thrive (Vuorinen et al. 2020). Although successful tracking applications and social media platforms require users and sharing, the form of power they apply is not based on negative but on positive terms. The platforms need to attract users, which they do through positive means.

Foucault's concept of power has another notable feature. For Foucault (1990), the productivity of power means that power always creates an outcome and activity. For example, when a fitness enthusiast employs a workout programme, the programme suggests specific exercises for the enthusiast, creating a special schedule for training which means the programme suggests what to do and when to exercise. Essentially, the programme executes a form of positive power. In milder terms, there is a power relation between the subject (the exerciser) and the programme, and this relation produces activity. Yet another way to put this is that the fitness enthusiast subjects the self to the programme. In training, the programme is actualised.

The power of a programme stems from a number of different sources. The workout programme is visible and tactile in itself (whether it is an app on a mobile phone or a written notebook) but it is also perceivable in terms of its presence and effects. In gym training, every repetition that the programme requires can be felt as muscle burn when the exercises of the programme are carried out. However, there are more aspects and actors connected to the programme, which are not present in such an obvious way and thus not easily perceivable. For example, workout programmes emerge from expert discourses on health and well-being. A Foucauldian "statement" such as "gym exercise is good for the health" implies power at the level of discourses, in Foucault's (1990, 2013) terminology. In such a case, the workout programme is a messenger of healthy life, an instrument towards a good life. In other words, it emerges from a certain combination of statements (Foucault 2013). However, there are different discourses of what is considered "rational" and desirable and thus what the goal of the programme is. For example, if the programme is designed for a bodybuilder, the programme might be a messenger of how to build body mass and how to reduce body fat. In this case, the programme has become a messenger of aesthetics, as bodybuilding is about being visually impressive. If the workout programme is created for a sumo wrestler, the rationale is presumably different in terms of body fat. The point is that there are different discourses – health, well-being, sumo wrestling, bodybuilding – in relation to which different sets of rules for training can emerge. Nevertheless, the discourse – a formation of statements, in Foucault's (2013) terms – always carries a power dimension that can be actualised in different ways. As emphasised above, Foucauldian power is doing; the relation comes into being as it is enacted. For example, as the programme is carried out, the power dimension is carried out.

Foucault (e.g. 1990, 2013) examines how different types of power actualise as different practices. The training programme is an example of such actualisation. A self-tracking application and device is another. The devices suggest and alert. They produce knowledge about their user (cf. Foucault 1982). Foucault's (2003) initial and implicit power concept resides at the level of institutions and the emergence of their subjects. Certain types of subject emerge under particular discursive formations that, for example, define madness or other forms of abnormality (Foucault 1995). All the institutions treat the subjects in a specified way by focusing on specific problematizations but monasteries, schools, military barracks, factories, hospitals, and prisons also have similar ways of organising their subjects (Foucault 1995; also Deleuze 1992). Each of these institutions bases their activity on discipline and they all are establishments for mass populations, which are separated as individual cases and placed in confined spaces for observation (Foucault 1995; also 2002). In each establishment, the subjects have imposed on them the institutional schedules, activity, and flows of time (see Foucault 1980). The internalised individuals become subjected to the practices of the institution; they become subjects of the power system (Foucault 1980, 1995). A student, a soldier, a worker, a patient, a prisoner are all examples of subjects. They are subjected to the system that uses disciplinary methods to obtain the desired outcome. These institutions are also places of transformation in which the subject is corrected, healed, educated – made an obedient body. (Foucault 1995; Deleuze 1992) More specifically, Foucault (1995) focuses on how measures of correction are used to normalise subjects and how subjects' bodies are made docile.

For Foucault (1995), power is a two-way arrow: in terms of activity, which is a product of power, there is always the possibility of resistance. Moreover, power relations are everywhere. More precisely, power forms a dimension through which the relation is partly carried out (Foucault 1980). Thus, there is no place outside power relations, because practices, suggestions, and proposals are everywhere. At the same time, there is the possibility of resistance. For example, a self-tracking device might give an idle or inactivity alert in order to push the user to become more active. There is the possibility of resistance because the alert and the tracker device can be ignored, or at least its orders can be questioned or disobeyed at specific points in time. Moreover, the device could be jettisoned because of the alerts.

2.2 From the Society of Discipline Towards the Society of Control

In Gilles Deleuze's (1992) interpretation, Michel Foucault's observations and concepts of power pertain mainly to the discipline society, which has its own characteristic features. For Deleuze (1992), the discipline society is a fading form of power which in modern, increasingly data-driven societies is becoming replaced by what he calls the control society. In the control society, power is actualised in a different manner than in the discipline society. One of the essential characteristics of the discipline society, which Deleuze (1992) emphasises, is the analysis of the individual in relation to the mass. A medical check-up is a helpful example of this: in a medical check-up, individual cases are compared. The comparison is carried out through different attributes and markers. For example, what is the normal (healthy) level of diastolic blood pressure? In answering such a question of normalcy, the Gaussian (or normal) curve provides an answer as a means of defining "normal" or "healthy" (Canguilhem 2012, also Foucault 1995). In the discipline society, the individual is invisible when their relation to the mass is within the normal level, close enough to the average. However, when the variance is great, it draws attention. Efforts are made to normalise abnormally high or low blood pressure results. The power of the normal – the power of the mass – prevails.

Deleuze (1992) notes that power in the discipline society works through institutional spaces. As mentioned above, the discipline society organises spaces (e.g. schools, hospitals and factories) and takes the individual under its gaze in such carefully designed facilities (Foucault 1995). Foucault (1980, 1995, 2013) emphasises that individuals become subjects through different sets of knowledge, e.g. medical knowledge. For example, a cancer diagnosis makes the individual a cancer patient that is to be treated in a certain way (e.g. operated on, observed, and medicated). In other words, medical knowledge guides how the subjects should be treated. Thus, power and knowledge are intertwined (Foucault 1995). A power relation requires knowledge. In turn, knowledge creates a power relation with its subject. Foucault is interested in how such subjects emerge through grids of specification (Foucault 2013).

In addition to this subjection, there is another important feature that comes with the spaces of power. Namely, the spaces of the discipline society are finite in scope (Deleuze 1992). The institutions of the discipline society are places of transformation through which individual transformations can become complete and, in a sense, ready. For example, a hospital releases a patient when they are considered healthy (enough). A student earns a degree in the school system designating that the student is no longer a pupil and has become capable of carrying out a certain task (e.g. that of a medical doctor or a psychologist). Importantly, there is an end point or, if not an end, a point of exit.

In Deleuze's (1992) control society, things unfold in a different way. Essentially, normalcy and the individual's relation to the mass (the main axis of the discipline society) is no longer the first dimension of measurement; the relation to the "code" replaces it. Machines in sites of work are different in the control society. Code-running machines replace the slicing and cutting machines of the factories. Today, Deleuze's vision is apparent in terms of data driven technologies. The power of the algorithm has taken over, as algorithms organise our worlds and experiences (Introna 2016). Algorithms are an inseparable part of the consumer world. In addition, they are a crucial part of politics. The data-driven society nurtures algorithms. However, the power of algorithms is invisible and fluid. Algorithms do suggest, guide, encourage and produce. They generate activity and possess agency (cf. Latour 2005). This kind of power pertains to activity trackers. We live with and on algorithms, and algorithms work with and on individuals.

Algorithms can make guesses about the user. For example, the suggestions of YouTube or Netflix for the next video are guesses about which video would attract the user. Algorithmic power lies in suggestions and proposals, not in forcing. It is as if the user is given a quasi-choice: yes, no, this or that. The "quasi" part in the suggestion is based on exclusion: the suggestion excludes millions of choices and offers a few from which to choose. In this way the algorithm enables and empowers the user. However, the power of the user is very limited. The suggestions are based on big data but the data is not analysed merely in terms of gaussian curve (normal choices and abnormal choice). Rather, there are a number of different profiles constructed. A learning algorithm draws conclusions from the smallest inputs (user activity).

Algorithms work by dividing users into sets and categories. For example, age, sex, location and interest in different topics are obvious attributes. As Deleuze (1992) noted, codes create *dividuals* as they *di*vide the individual in a number of dimensions. In other words, through the algorithms we become slices and pieces which are then examined and put together as *a profile* that is a sellable product, a commodity. A sellable digital profile is capitalism encapsulated. The divisive logic of the dividual relates closely to self-tracking practices, in which human bodies and selves become divided into ever more nuanced bodily functionalities, which are displayed as data points at specific points in time as well as over temporally extended trajectories, and in which a certain type of often health-related profiling pertains to the self (see Bergroth and Helén 2020).

The control society also differs from the discipline society in terms of space. There are no separate and enclosed spaces; spaces overlap and penetrate each other. Of course, data and information always require a material space (Blanchette 2011). A concrete storage space, such as an SSD-drive, is needed for bits. However, mobile accessibility nullifies the meaning of a particular space in terms of user location. For example, a cloud service is invisible beyond its interface (e.g. a mobile application), no matter where the interface is used. Different data spaces and services can be accessed through multiple gateways. Thus, a user can take different roles, and occupy and change different social positions through technology (Carter and Grover 2015). In the control society, users are divided and assembled in virtual spaces through IT artefacts. For example, a person can check their blood test results, deal with work-related emails, and send their partner love-heart emojis, all in a single location though a single machine. Different data spaces and different user roles actualise on a single concrete spot. In the control society the roles of the self are not stable and directional as they were in the discipline society. Generally, Instagram posts and tweets fade away; the roles of subject have to be constantly re-established. There are no single spaces but a multiplicity of places.

The user space of the control society has become complex – or multidimensional – compared to the individual's life in the discipline society. In the control society, space is not organised in such a way that a user could merely pass through it (Deleuze 1992). The spaces of the control society are virtual, such that a user lingers, jumps, hovers, and disappears in them. For example, with a mobile phone that tracks step count, the user can take a quick look at their step count. However, after the glance, the user can then a switch to a different space, a different mode, such as social media. Importantly, these spaces of the control society – a step counter or social media sites – cannot be passed through; the user is stuck with them. They flow past the user, offering no other exit than completely quitting the service. There is no graduation ahead. The stillness of death would stop the counter, but it would not necessarily stop all social media services (e.g. Facebook), as the profile can become an "in memoriam profile". The institutions of the discipline society provided a chance for the individual to become ready and complete. In the control society, there is no end in sight. Rather, the algorithms of the control society seem to borrow our attention for a second and let us go (in order to observe us more). But soon they demand the return of the user's attention. For example, in self-tracking, there is no end in terms of tracking the step count (Bergroth and Vuorinen 2019). The usual step count goal is 10000 *daily* steps. If the user achieves this goal of 10000 steps in a day, the device can send haptic vibrations in order to signal that the goal has been achieved. However, the goal is set for a single day and the counter resets at midnight. The spaces of the control society do not contain direct paths. These spaces are about transformation, the fluxes of crossing currents pulling the self in multiple directions. They are places for the dividual in the control society.

3 Self-tracking and the Control Society

In self-tracking practices the self is divided by itself. This means that the self monitors itself, examines itself. The self appears simultaneously as subject and object. There is the assessing side of the self, that examines and observes. In addition, there is the other side of the self, that is the object of assessment (Bergroth and Vuorinen 2019). Put simply, examination by the self focuses on the self. This process is based on the relation to self (Foucault 2012a). Power is aimed at itself by the self (Deleuze 1988, 103). In this way, the individual has made itself a dividual; the divided individual.

Foucault (2012a) describes the way people form relations to themselves and observe the self as technologies of the self. When an individual – the subject – seeks to obtain a new way of acting, or a correct way of acting, the self has to be subjected to the self. To achieve this subjectivation, technologies of the self are applied. For example, if a consumer wants to get in better shape, a self-tracking device can be acquired. However, there are alternative ways of getting in better shape. The consumer might consider a gym membership, for instance. Furthermore, there are different ways of using self-tracking devices, just as there are a number of ways to exercise at the gym. In each way of exercising, there is a method, which is seen to improve the physical shape. However, there are many ways to carry out the method. In each case, a relation to the self is created.

In ethical terms, this implies the question of a "good life": how does the self treat itself as it seeks to create a good life? So in addition to the relation to the self, there needs to be a scale or code by which the subjects can subject themselves. In the example above, the code is good/poor shape. However, the meanings of good and poor shape are constituted in multiple ways, as seen above.

In the case of fitness or more generally well-being, self-tracking devices provide such a pre-set code to which an individual can subject themselves, through which they can create a relation to the self. Depending on the tracking device, different traits can be monitored. Each trait, in Deleuze's (1992) terms, would be a dividual feature. Trackers can examine, for example, the number of steps taken, the number of floors climbed, active hours, and/or time and quality of sleep. In addition, one of the central features in contemporary self-tracking devices is the heart rate monitor. Trackers can also be used to assess diet in terms of calories and macros (i.e. fats, carbs, and protein). Every single one of these monitored features is an example of a division made in the individual. They constitute a divided individual; a whole self sliced into smaller parts. If in the discipline society spaces were sliced and divided in order to create certain subjects, the control society applies the principle of slicing to the individual: one's bodily functions and life. Eating, sleeping, being awake, the beating of one's heart all become inputs of life and health for trackers; the noises of life are sent as a message to the tracker device that – as a space of transformation – returns them as an assessment of and for the self. Furthermore, in the control society, though these inputs are mere indicators of "healthy" or "unhealthy", they easily become a forceful part of everyday life, as the pre-set, algorithmic code of self-assessment run in the background of everyday actions, providing an in principle endless dataflows on oneself based on which one can modify one's behaviour.

The tracker is a way of creating a relation to the self. It provides an axis by which the self can assess itself. With its alarm features, a tracker device can push the individual to move. Yet there is a side that it cannot touch: the relation between the assessing self and the tracker device. The choice exists. It can be ignored, treated as a nuisance or a gimmick. This is the untouchable part.

4 Discussion

In Ancient Greece, there was the well-known notion of "gnothi seauton": know thyself. These are words of ethical imperative, very familiar in the therapeutic assemblages of today (Salmenniemi et al. 2020), encouraging self-related knowledge production. In order to be a good individual, one should strive to know (more about) oneself. However, there is also another practice that Foucault (2012b) is interested in, namely "epimeleia heautou": care of the self. This too implies an imperative: the instruction to take care of oneself. Without going deeper into Ancient Greek practices (see Foucault 2012a, b), it should be noted that in self-tracking the two concepts come together. The tracker constantly provides information about the self by displaying bodily functions. Heart rate, (in)activity, exercise, sleep, calories burned, and step counts are provided in visual, numeric and sometimes haptic forms. However, in doing this, self-tracking practices are also about slicing the self into small pieces in providing information. Self-tracking is "dividualising", as it divides individual activity and the seemingly whole self into slices by enabling ever more nuanced monitoring of separate functionalities of the body (see also Bergroth and Helén 2020). Furthermore, this is a feature of the control society because self-tracking enables knowledge about and care for the self via in principle infinite monitoring of the processes of life through data. In this sense, self-tracking is not bound to a certain place or institution and therefore there is no "natural" end to it, no exit point in time. Self-tracking cannot be "completed". Any goals are mere waymarkers. Tracking may begin but in principle it does not end. Even if the tracking device is abandoned, it can still trouble the ex-user by its haunting presence as an internalised demand or as a mode of relating to oneself (Bergroth and Vuorinen 2019). Thus, it is not a mere device of knowing oneself – it is a demanding collection of imperatives on both knowing and taking care of oneself: sleep; exercise; be active enough; keep going.

Yet neither the imperative to take care of yourself nor that to know yourself are solely internalised by the subject as would be the case in the discipline society. Everyday technologies of the self play a crucial role in how knowing about and caring for the self are partly externalised in the control society. For example, self-tracking devices provide categories of what to watch for, what to observe, either through algorithmically generated health guidance or by providing data streams of specific functionalities of the body. They provide an accompanying code to one's everyday life, through which the self can constantly be subjected even when one is not making an effort in self-monitoring in the sense of paying constant attention to oneself. Codes of self-tracking still run in the background when the devices are worn. Moreover, self-tracking is unrelenting in the sense that the user never graduates, never becomes "ready". There is no end, just continuous cycles. In this sense, the self-tracking device is not an institution of the discipline society but a materialisation of a code of the control society that follows the subject everywhere and is accessible all the time. The physical or institutional place does not matter the way it does in the discipline society. In the control society, place disappears and the code (in the self-tracking device) is with the user at all times and in all places.

5 Conclusion

In this article we have related contemporary digital self-tracking practices and their data-driven character to Foucauldian and Deleuzian ideas of the functioning of power regimes in society; the discipline society and the control society, respectively. We have argued that, as a practice, self-tracking intertwines the notions of knowing about the self and caring for the self, in the process enacting regimes of (self-)control in which care becomes enabled by continuous data streams on the self. As such, self-tracking practices divert from the principles of the discipline society in which subjectivities are produced within institutional contexts. Self-tracking seems more aligned with the principles of the control society, the key tenets of which are the division of seemingly indivisible individuals into ever more nuanced parameters and functionalities, and the logic of ongoing transformation in which the subject of self-tracking is never complete (see also Bergroth and Helén 2020). In the discipline society, the institutions were the spaces of transformation. In the control society, the space of transformation is within the self and is mediated by the codes of control.

References

Berg, M.: Making sense with sensors: self-tracking and the temporalities of wellbeing. Digit. Health **3**, 1–11 (2017)

Bergroth, H., Vuorinen, J.: Towards the ontology of becoming in self-tracking research. In: Kurosu, M. (ed.) HCII 2019. LNCS, vol. 11566, pp. 270–287. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22646-6_19

Bergroth, H., Helén, I.: The datafication of therapeutic life management: assembling the self in control society. In: Salmenniemi, S., Nurmi, J., Perheentupa, I., Bergroth, H. (eds.) Assembling Therapeutics: Cultures, Politics and Materiality, pp. 107–123 (2020)

Blanchette, J.F.: A material history of bits. J. Am. Soc. Inf. Sci. Technol. **62**(6), 1042–1057 (2011)

Canguilhem, G.: On the Normal and the Pathological, vol. 3. Springer, Heidelberg (2012). https://doi.org/10.1007/978-94-009-9853-7

Carter, M., Grover, V.: Me, my self, and I(T) conceptualizing information technology identity and its implications. MIS Q. **39**(4), 931–958 (2015)

Crawford, K., Lingel, J., Karppi, T.: Our metrics, ourselves: a hundred years of self-tracking from the weight scale to the wrist wearable device. Eur. J. Cult. Stud. **18**(4–5), 479–496 (2015)

Deleuze, G.: Foucault. University of Minnesota Press, Minneapolis (1988)

Deleuze, G.: Postscript on the societies of control. October, **59**, 3–7 (1992)

Foucault, M.: Power/Knowledge: Selected Interviews and Other Writings 1972–1977. Harvester Press, London (1980)

Foucault, M.: The subject and power. Crit. Inq. **8**(4), 777–795 (1982)

Foucault, M.: The History of Sexuality: An Introduction, vol. 1. Vintage Books, New York (1990)

Foucault, M.: The Birth of the Clinic. Routledge, London (2002)

Foucault, M.: Madness and Civilization. Routledge, London (2003)

Foucault, M.: Discipline and Punish: The Birth of the Prison. Vintage Books, New York (1995)

Foucault, M.: The History of Sexuality: The Use of Pleasure, vol. 2. Vintage Books, New York (2012a)

Foucault, M.: The History of Sexuality: The Care of the Self, vol. 3. Vintage Books, New York (2012b)

Foucault, M.: Archaeology of Knowledge. Routledge, London (2013)

Introna, L.D.: Algorithms, governance, and governmentality: on governing academic writing. Sci. Technol. Hum. Values **41**(1), 17–49 (2016)

Latour, B.: Reassembling the Social: An Introduction to Actor-Network-Theory. Oxford University Press, Oxford (2005)

Lupton, D.: The diverse domains of quantified selves: self-tracking modes and dataveillance. Econ. Soc. **45**(1), 101–122 (2016)

Salmenniemi, S., Nurmi, J., Perheentupa, I., Bergroth, H. (eds.): Assembling Therapeutics: Cultures, Politics and Materiality. Routledge, London (2020)

Vuorinen, J., Koivula, A., Koiranen, I.: The Confidence in social media platforms and private messaging. In: Meiselwitz, G. (ed.) HCII 2020. LNCS, vol. 12194, pp. 669–682. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-49570-1_48

Towards Social Enterprise with Internet of Office Desks

Elena Vildjiounaite(✉), Johanna Kallio, Julia Kantorovitch,
Vesa Kyllönen, Pauli Räsänen, and Jussi Ronkainen

VTT Technical Research Centre of Finland, Kaitovayla 1, 90570 Oulu, Finland
{Elena.Vildjiounaite, Johanna.Kallio,
Julia.Kantorovitch, Vesa.Kyllonen, Pauli.Rasanen,
Jussi.Ronkainen}@vtt.fi

Abstract. Social enterprises are organisations, combining profit making with support of their employees and environment. Employee stress is harmful for both society (due to increased risk of mental health disorders and cardiovascular diseases) and for enterprise performance (due to increased risk of presenteeism, employee turnover and early retirement). This work aims at helping enterprises to improve employee wellbeing. To this end, we introduce a concept of IoT-based privacy-aware "team barometer" and present a first study into using inexpensive PIR (passive infrared) motion detection sensors in such barometers. The study was conducted as follows: first, we deployed IoT system in real offices and collected employee data in the course of everyday work during several months. Second, we developed a machine learning method to classify human conditions on the basis of collected PIR data. In the tests, this method recognised employees' stress with 80% accuracy and dissatisfaction with indoor environmental quality - with 75% accuracy. Third, we integrated stress detection results into a "team barometer" and conducted interviews of line managers. Interview results suggest that the proposed IoT-based team barometer can be beneficial for both employees and enterprises because of its potential to discover and mitigate workplace problems notably faster than with current practice to use periodic surveys.

Keywords: Stress detection · Team barometer · PIR sensors

1 Introduction

Recent social, economic and political challenges lead to the rise of so-called "social enterprises": organisations, combining revenue growth and profit-making with respect and support of their employees, environment and stakeholder network [1]. This is not a matter of altruism: initiatives that have positive social impact are necessary for sustaining and growing businesses [2] and for attracting and retaining critical workers [1].

Reducing employee stress serves both goals of social enterprises: first, around 50% of all lost working days have some links with work stress [3], which is costly for society; and second, significant correlations exist between workplace stress, organisational commitment and organisational performance [4]. Currently, employers invest

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 361–374, 2020.
https://doi.org/10.1007/978-3-030-62803-1_29

into reducing employee stress, but mainly by providing coaching/relaxation services. These services, unfortunately, do not eliminate underlying problems: work overload, time pressure, lack of work variety, lack of control, unsatisfactory environmental conditions [5]. Indeed, a recent review of interventions to promote work participation of older workers (i.e., experienced workers over 45 years old) concluded that only combination of health services and work modifications improves work participation; there is not enough evidence to recommend health services alone [6].

Work modifications require involvement of managers; thus, managers should be aware of the problems. Typically, employee satisfaction/dissatisfaction with workplace culture, own tasks, workload etc. is assessed via surveys, and individual answers to these surveys are aggregated into so-called "team barometers": for example, they report average values of multiple answers. Surveys are usually infrequent, whereas 29% of new hires quit in 90 days: 45% of them - because day-to-day role was not what they expected, and 28% because of unsatisfactory company culture [7]. New tools to detect problems earlier are emerging, but they are based on explicit reporting of team members, for example, via email [8].

Modern technologies, such as Internet of Things (IoT) and data science, have a potential to detect problems in time and unobtrusively for the employees. Therefore technologies can help social enterprises to improve employee wellbeing and health. This study introduces the concept of IoT based "team barometers", where IoT is used to collect human and environmental data, and intelligent algorithms are employed to assess human conditions on the basis of these data.

Technology-assisted assessment of human conditions is a relatively new research area; it started from analysis of video and physiological data. At work, however, video cameras rise privacy concerns. Physiological sensors are costly, and their data are affected by physical activities: in a recent large-scale real-life study stress detection accuracy was as good as random guess for 38% of test subjects [9]. Thus real-life stress detectors often utilise behavioural data, mainly, obtained from mobile phones [10] (e.g., application usage, locations etc.). This approach does not require any extra gadgets, but data collection may quickly drain phone battery [11].

In-office sensors do not require maintenance (e.g., charging) efforts from the monitored employees, and in-office motion detectors, such as depth cameras and PIR sensors, are usually well accepted [12, 13]. Motion features are also indicative of stress [12, 14, 15]. Currently, however, PIR sensors are mainly used for energy savings, whereas needs of users within the building are often ignored [13]. Furthermore, recently it was proposed to employ IoT to classify employees as high performers vs. low performers [16]. One of the features, derived from PIR data for the classification, was "higher performers spend more time at work during weekends". This approach can be dangerous for employee health and detrimental for employers in the long term.

This work, on the contrary, presents a study into using IoT and PIR sensors for improving work satisfaction and wellbeing of employees. This work is organised as follows: Sect. 2 presents a concept of IoT-based team barometer. Section 3 describes IoT setup, used for collecting human data in the course of everyday office work. Section 4 presents a methodology to assess wellbeing of employees on the basis of collected data, along with the accuracies of data analysis results. Section 5 describes a methodology to aggregate results into team barometer and focus group study with line managers, conducted for evaluation of the proposed PIR-based team barometer.

The main contribution of this paper is introducing concept of IoT-based team barometer. To the best of our knowledge, this is the first real life study into using PIR sensors for assessing human stress and satisfaction with environmental conditions, and the first work, presenting a focus group study with IoT-based team barometer data.

2 IoT-Based Team Barometer

An overview of the proposed sensor data analysis-based team barometer is presented in Fig. 1. The barometer relies on IoT nodes in the office cubicles to collect data, and on intelligent algorithms to assess individual conditions and to aggregate them into team states. Aggregated results are then visualised.

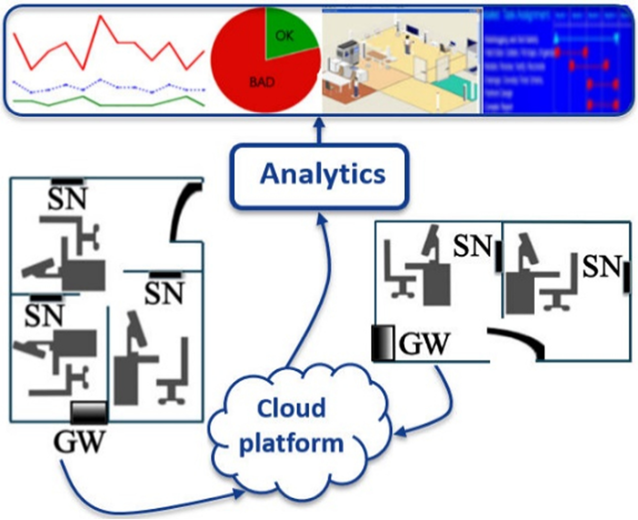


Fig. 1. System overview: SN - sensor node; GW - gateway.

In this work, we recognise conditions of office workers on the basis of behavioural data, acquired from PIR sensors, and we aggregate recognition results over multiple individuals (to avoid privacy problems) and over time (e.g., couple of weeks). Aggregation over time is performed, first, because long-lasting problems are more likely to result in negative consequences for the individual (e.g., burnout) and for the organisation (e.g., resignation of a talented employee) than a short-term problem. Second, aggregation over time increases accuracy: for example, in [12] accuracy of classifying each day as stressful vs. normal was less than 70%, but accuracy of classifying months exceeded 90%. Last but not least, managers are also humans, they can be also stressed and hence do not want too detailed information.

3 Data Collection

3.1 IoT Setup

IoT-based system was deployed in our working premises to continuously collect sensor data. Figure 1 presents examples of IoT setup in three-person office and in two-person office, and Fig. 2 presents a photo of an office cubicle and location of a sensor node (red arrow points there). In each cubicle, sensor nodes were positioned 1.1 metres above the floor level in a vicinity to a seating position (1–2 m distance). This way, each sensor node was monitoring a single individual, but motion detection was robust to changes in his/her positions inside the office cubicle: for example, Fig. 2 shows that the monitored subject had two computers and was moving between them.

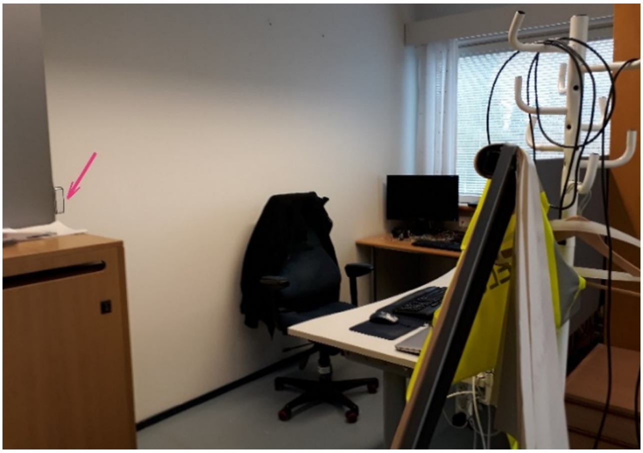


Fig. 2. Sensor placement in the office cubicle (Color figure online)

In this study, we use Tiny Sensor Node [17] with Panasonic EKMB1301113K PIR sensor [18]. It acquires 12 samples per minute; every single sample can take either value 0 (no motion) or value 1 (motion detected). Tiny Sensor Node aggregates these values over one minute (aggregated values range from 0 to 12) and sends them over Bluetooth Low Energy to the data acquisition gateway application, running on a Raspberry Pi computer. The gateways send data to Microsoft Azure cloud platform via MQTT protocol over TCP/IP.

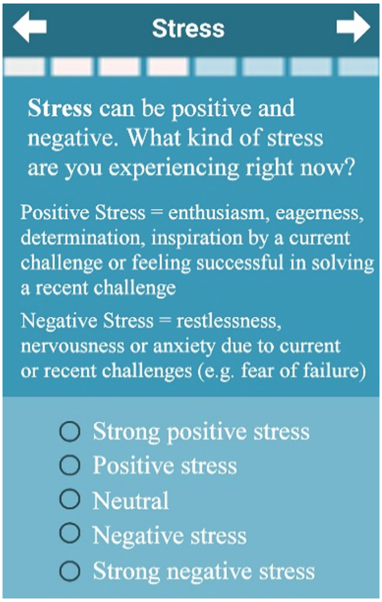
Data retrieval from Azure utilises a published API for Microsoft Azure Table Storage. We implemented a client application, based on Cosmos DB API, which provided an interface to TableStorage via a module, called TableService [19]. Data were retrieved periodically in batches.

3.2 Collection of Self-reports

Self-reports of the monitored subjects were collected on 5-point Likert scale via mobile phone application, developed for Android phones. Figure 3 presents a screenshot of stress-related part of the application. Other questions, used in this study, were formulated as follows:

- How productively did you work today? Options to answer: (a) much more productively than usually; (b) more productively than usually; (c) as usually; (d) less productively than usually; (e) much less productively than usually.
- How would you estimate air freshness at the moment? Options to answer: (a) very good; (b) good; (c) acceptable; (d) bad; (e) very bad.
- How would you estimate temperature at the moment? Options to answer: (a) cold; (b) cool; (c) neutral; (d) warm; (e) hot.

Questions regarding air freshness and thermal comfort were asked two times on each workday (morning and afternoon); questions about stress and productivity were only asked in the afternoon. Answers were stored in MongoDB and retrieved via REST API.



Stress can be positive and negative. What kind of stress are you experiencing right now?

Positive Stress = enthusiasm, eagerness, determination, inspiration by a current challenge or feeling successful in solving a recent challenge

Negative Stress = restlessness, nervousness or anxiety due to current or recent challenges (e.g. fear of failure)

○ Strong positive stress
○ Positive stress
○ Neutral
○ Negative stress
○ Strong negative stress

Fig. 3. Screenshot of stress-related part of self-reporting application.

3.3 Dataset

For this study 30 subjects were recruited, but in the middle of data collection part of the subjects was relocated because renovation started in their offices. The renovation started because of unsatisfactory air quality, which is a good sign: it shows that aggregated opinions of employees can influence working conditions. Unfortunately, relocations and several other drop-outs (e.g., one subject fell ill, another one started to work on PhD elsewhere etc.) reduced number of subjects, whom from self-reports were obtained in sufficient quantities: for analysis of thermal comfort and air quality we used data of 15 subjects (they were mainly middle age researchers, 7 females and 8 males). For analysis of stress and productivity perceptions we used data of 6 persons (2 females and 4 males) because stress- and productivity-related questions were asked only once per day; hence, data size for these questions was approximately twice smaller.

4 Data Analysis

Stress-induced motion changes are distinct to every individual (for example, some individuals sit still during difficult tasks, and some fidget [15]), and person-specific models of stressed behaviour usually achieve 20% higher accuracies than "one-fits-all" models [14, 15, 20]. Perception of thermal comfort and air quality, as well as physical reactions at them, are also person-dependent. Therefore, we employed person-specific models in this study. To date, the majority of existing stress detectors were trained in fully supervised way [10], and thermal comfort models too [21]. Thus in this study we also used supervised classifier - SVM (Support Vector Machines). As class labels we used subjective perceptions of the test subjects because it is a common practice: it is not possible to obtain objective measures of human conditions in everyday work. We have chosen SVM because human behaviours vary a lot, self-reports may contain mistakes, and SVM is a noise-robust classifier, capable of learning from relatively small datasets. We used SVM implementation in scikit-learn Python library. We used sigmoid kernel with its default parameters, but with twice higher penalty for misclassifying examples of "bad" class because there were fewer examples of this class in the data.

For feature extraction we used tsfresh Python library; it extracts over 300 data features (for descriptions see [22]). Feature selection was performed by calculating correlations between every feature and respective class labels (e.g., correlation between "energy" feature values and stress labels, correlation between "energy" feature values and productivity labels etc.). Thus we selected best features for each problem separately.

Although self-reports were collected on 5-point Likert scale, their quantity was sufficient only to train 2-class classifier. Hence in this study we distinguished between "bad" and "OK" conditions. For stress detection "bad" class included "negative" and "very negative" self-reports, while "OK" class included "neutral", "positive" and "very positive" self-reports. Similarly, for productivity classification "bad" class included "much less productively than usually" and "less productively than usually" answers, while "OK" class - all other answers. For air/temperature perception "bad" class included "bad" and "very bad" answers; "OK" class included all other answers. This approach resulted in the following percentages of answers of "bad" class in the data: stress - 28%; productivity - 18%; air freshness - 45%; thermal comfort - 51%.

Then, SVM was trained to map sensor data into one of the two classes of personal perception: for example, to classify stress self-report into either "bad" or "OK" class.

4.1 Experimental Protocol

Since data were not abundant, for each test subject we used leave-one-self-report-out protocol: first, one self-report and the corresponding sensor data were selected as test data, and all remaining data were used for feature selection and SVM training. Then this process was repeated for all self-reports of this person, and then the same procedure was repeated for all other subjects. For each subject, we trained a separate SVM model for each question. Sensor data features were extracted from time windows, preceding each self-report in training data (we used time window length two hours).

For testing, we extracted selected features from the test data, classified each test sample with SVM and compared SVM output with the corresponding self-report. We estimated test accuracy according to the following measures:

$$Total\,Accuracy = \frac{NcorrectBad + NcorrectOK}{N\,bad + N\,OK} \tag{1}$$

$$True\,Bad = \frac{NcorrectBad}{N\,bad} \tag{2}$$

$$True\,OK = \frac{NcorrectOK}{N\,OK} \tag{3}$$

In the Eqs. (1), (2) and (3) *NcorrectBad* is the number of correctly classified "bad" answers; *NcorrectOK* is the number of correctly classified "OK" answers, *N bad* and *N OK* are numbers of "bad" and "OK" answers, respectively.

4.2 Classification Accuracies

Table 1 presents classification accuracies, obtained from PIR data.

Table 1. Classification accuracy.

Problem	Total accuracy	True OK (True Negative)	True bad (True Positive)
Stress	0.80	0.86	0.60
Thermal comfort	0.75	0.72	0.80
Air freshness	0.75	0.70	0.80
Productivity	0.70	0.92	0.25

Table 1 shows that the employed classification approach results in reasonably high accuracies (considering coarseness of PIR-based motion data) for all questions except for productivity. The proposed approach was not able to detect practically any case of perceived low productivity. We consider this result good, too, because it means that the employer cannot use such system for assessing employee performance; instead, the employer should invest efforts into keeping the employees healthy, happy and motivated.

5 Focus Group Study

5.1 Visualisations of Data Analysis Results

Figure 4 presents one of the proposed team barometer views: a plot of weekly team states for 12 weeks, created using stress detection results of all test subjects (for some subjects we did not have data for longer time period). Vertical axis in Fig. 4 presents a stress score, calculated as follows: first we obtained stress detection result for each day as described in the previous section, then calculated a week score of each subject according to Eq. (4), and then averaged the scores over the subjects.

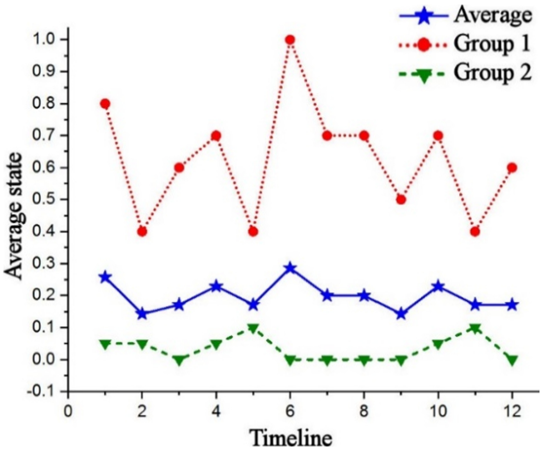


Fig. 4. Stress detection results, averaged over weekly intervals and test subjects. “Average” plot presents average over all test subjects; “Group 1” and “Group 2” plots present averages over the most stressed one/third of the subjects and the least stressed one/third of the subjects respectively

$$Score = \frac{1}{N}\sum\nolimits_{i=1}^{N} ClassifierDecision_i \tag{4}$$

In Eq. (4), N is number of days in the corresponding time interval, for which classifier decision was obtained. $ClassifierDecision_i$ is SVM output for test sample number i. It takes value 1 if SVM classifiers the day into “bad” class, and value 0 if SVM classifiers the day into “OK” class. Accordingly, average score of a time interval ranges from 0 (all days were classified as “OK” days) to 1 (all days were classified as “bad” days).

Horizontal axis in Fig. 4 presents interval number. In Fig. 4, interval is one week, but we don’t display it in the figure legend because for the focus group study we needed unspecified time interval.

Figure 4 shows that subjects in the two groups experienced stress at different times (e.g., results for weeks 5 and 11 are negatively correlated), possibly due to differences in work tasks. Figure 4 also shows that Group 1 experienced higher stress, either due to differences in work tasks or personalities or both.

Figure 5 presents stress detection results, averaged over two-weeks-long time intervals, and corresponding averages of self-reports for two test subjects. They were selected because they were very different (Person X was stressed much more often than Person Y) and because stress detection errors for them look impressive. For Person X, Total Accuracy is 0.70 (True OK = 0.42; True Bad = 0.80); for Person Y, nearly none of self-reported stress cases were detected.

Average score for each interval in Fig. 5 was calculated according to Eq. (4), and average label was calculated according to Eq. (5).

$$Label = \frac{1}{N}\sum\nolimits_{i=1}^{N} BinaryLabel_i \tag{5}$$

In Eq. (5), N is number of days in the corresponding time interval, for which classifier decision was obtained, and $BinaryLabel_i$ is a discretised self-report for sample i. It takes value 1 if "bad" or "very bad" stress was reported, and value 0 otherwise. Accordingly, average labels for biweekly period range from 0 (all days were reported as "OK" days) to 1 (all days were reported as "bad" days).

Figure 5 shows that the proposed PIR-based stress detector over-estimated number of stressful days of Person X and under-estimated number of stressful days of Person Y. The main reason is great variety of tasks and motion behaviours: e.g., classifier results were influenced by absences from the office and by numbers of visitors (communication with them typically involves more motion than typing on a computer). Detection errors for Person Y were also due to his/her imbalanced data: number of examples of "OK" class notably exceeded number of examples of "bad" class. Nevertheless the overall conditions were estimated reasonably well.

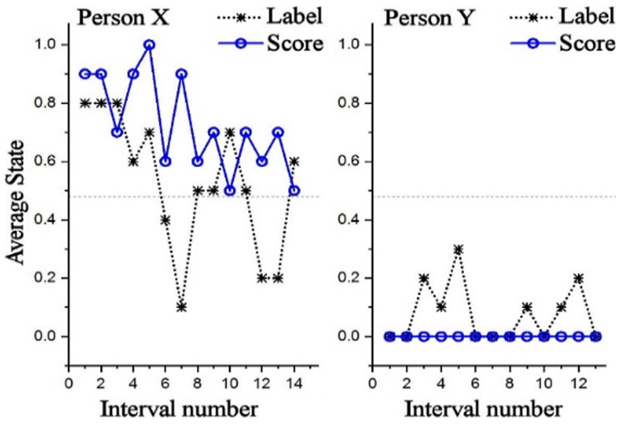


Fig. 5. Stress classification results and self-reports for two test subjects, averaged over biweekly intervals. Dotted line separates difficult periods from easier ones: if a person is stressed on more than 50% of days, this period is definitely not easy.

For person X Fig. 5 presents results for 14 biweekly intervals, which corresponds to 140 days in the office; for person Y - for 13 biweekly intervals, which corresponds to 130 days in the office.

5.2 Study Protocol

We conducted face-to-face interviews (as a free-form discussion) with 8 individuals: one representative of a trade union and 7 subjects who held managing positions for at least a year. Each interview lasted about 30 min. Four of the interviewed persons had about 30 subordinates; three of them had 100 and more subordinates. In the beginning of each interview we described IoT solution and Fig. 4. We explained that Fig. 4 is a graph of stress detection results in an organisational unit, presenting (1) an average over all monitored subjects; (2) the most stressed subgroup; (3) the least stressed subgroup. Then we asked the following questions:

Q1: let's assume that this system is working in your unit; what would you do when you see these results?

Q2: how would you see forming of “group 1”, “group 2” etc.: for example, split can be by stress levels, by work tasks/positions/organisational units, or in some other ways?

Q3: what kind of timeline (interval size) do you envision: weekly, bi-weekly, monthly, anything else?

Q4: we can make similar graphs for thermal comfort and comfort with air quality, what would you do with them?

Then we presented Fig. 5, explained how scores and labels were calculated and pointed out the discrepancies between scores and labels. After that, we asked:

Q5: as you see, the system is not accurate, what do you think about its inaccuracy, and how would you take it into account?

5.3 Results

Answering to Q1, all managers started from workload distribution and task assignment problem, but they approached it from different directions. Their answers largely fall into two groups: “happier workers work better” and “unhappy workers do not work well”. That is, managers in the first group believe that attention to conditions of their subordinates would pay off in a long term in a form of higher motivation and improved workability. Managers in the second group prioritise current organisational performance. They want to give every task to a person who is the best fit for this task, and they believe that an already overloaded/stressed person cannot be “the best fit”.

Both types of managers stated that they are ready to help stressed persons by contributing themselves to a difficult task or by temporarily allocating more resources to this task, but managers of the second type were more inclined to delegate the problem of long-lasting stress to occupational health therapists than managers of the first type.

Managers of the first type appeared to be well prepared to deal with anonymous graphs: they all stated “if the system signals a trouble, I'll go and talk to people”. These managers were also interested to find stress reasons: they suggested augmenting team barometer with contextual information, such as various deadlines, company announcements etc., to improve planning (e.g., to start certain tasks earlier). In addition, a high-level manager said that if some group is constantly more stressed than the others, he would investigate whether leadership style causes problems.

Managers of the second type were more interested in getting lists of persons who are "able to take a new task now" and "who needs help now". This approach is difficult (if possible at all) to realise in a privacy-safe way. These managers were not expecting subordinates to work in any condition, however. Thus team barometer could be designed also to recommend stressed individuals to seek managers' support.

Answering to Q2, managers of the first type approved team split as "worst stressed group" vs. "least stressed group" and emphasised importance of checking conditions of "the worst stressed" group periodically - to see trends. They also said that if a new "worst stressed group" will form meanwhile, team barometer should also show it. Managers of both types suggested to split groups also by work tasks: they said that this split would help to provide timely support to "the task in trouble". Other suggestions were (1) to split groups according to work challenges (when work requires more novelty/creativity, it needs more attention from a manager) and (2) to split groups by demographic factors, like age. One manager, however, said that since he anyway would need to go and talk to people, any kind of indication that problem exists would suffice.

Regarding Q3, the majority of the managers had similar opinions: if stress lasts longer than a month, it is at least a first checkpoint. Being stressed for a month before a major deadline is a common case, but after that trends need to be monitored. And if stress lasts longer than two-three months, then manager's help would be definitely needed. Managers could not say, however, how exactly timeline in team barometer should look like (e.g., week, two weeks or a month) - they suggested testing different views in long-term use. Managers of the second type were more confident: they expressed interest in checking team barometer on a weekly basis in cases when tasks are more challenging than usually or when stressed group is likely to include people whom they don't know well.

Answering to Q4, most of the managers stated that it's easy to move the employees around until everybody is happy: for example, employees can be co-located based on similar preferences for thermal comfort and ventilation.

Answering to Q5, managers of the first type said "technology is just an indicator, I'll need to go and talk anyway", and "tools, based on explicit reporting, are not highly accurate either". Managers of the second type wanted to get as accurate actionable information as possible at least with respect to readiness to take a new task, but nobody expected any barometer to be 100% accurate.

Representative of a trade union said that team barometer could help to discover workload distribution problems and leadership problems: if for example three persons in a team have high stress level for more than a month, it is a reason to go and talk to the team. He also said that he would check average team states on monthly basis unless somebody comes and complains; then more frequent checks would be needed.

6 Conclusion

Skilful, motivated and healthy workforce is a key to enterprise success, that's why recently the term "human capital" started to supersede the term "human resources" and that's why employee wellbeing emerges as a strategic priority of enterprises [23]. This work presents a first study into using IoT for improving wellbeing of office workers.

In this study we deployed inexpensive PIR sensors in real office cubicles, collected motion data of the monitored subjects and developed machine learning method to access subjects' conditions. We consider use of real-life data very important because the majority of previous studies into use of motion data for stress detection took place in the labs. Lab studies typically last only a few hours under assumption that a short-term high mental workload is equal to stress. Furthermore, in lab studies stressful tasks typically alternate by periods of relaxation. In real life, however, tasks can last long, and frequent periods of relaxation are not possible. Therefore stress does not display itself in human behaviour in real life in a same way as in the labs [12].

The main advantage of employing in-office sensors is their unobtrusiveness, and the main drawback - ability to evaluate employees' conditions only when they are in their offices. This is not crucial, however, as the most dangerous stress type is chronic stress: long-lasting stress of low intensity may have equal or greater health impact than short-term high intensity stress [24]. Therefore team barometer does not need to detect stress on daily basis; detecting long-lasting stress would suffice, and interviews with line managers support this approach. Team barometer can employ also other sensor types, if the system is privacy-safe and unobtrusive.

The main limitations of this study are small number of test subjects and that all interviewed managers were from research organisation. Stress is an important health problem also for researchers, however [25]. In addition, growing competition for talented workers and consequent rise of "social enterprise" start influencing work culture in companies [23]. In future, we plan to collect more data and to test team barometer in the course of everyday work in teams of most trustworthy managers, to discover potential privacy threats, to develop privacy safeguards and to give employees control over the system. We also plan to compare various visualisation approaches.

We consider results of this work encouraging for future studies into using IoT to improve work satisfaction, motivation and wellbeing of the workforce because none of the interviewed managers said "stress of my subordinates is not my problem", and nobody expressed any intention to start a blame game based on technology reports. Instead, the managers said that the proposed team barometer is a good indicator of problems and could therefore help to improve work conditions and that they need to talk to their subordinates before taking any action.

References

1. Deloitte Global Human Capital Trends. https://www2.deloitte.com/content/dam/insights/us/articles/HCTrends2018/2018-HCtrends_Rise-of-the-social-enterprise.pdf. Accessed 06 July 2020

2. Success personified in the Fourth Industrial Revolution. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/gx-davos-DI_Success-personified-fourth-industrial-revolution.pdf. Accessed 06 July 2020
3. Cox, T., Griffiths, A., Rial-Gonzáles, E.: Research on Work-Related Stress. European Agency for Safety and Health at Work, Luxembourg (2000)
4. Karacsony, P.: Examining the Relationship Between Workplace Stress and Organizational Commitment, MSIE 2019, pp. 26–30 (2019)
5. Michie, S.: Causes and management of stress at work. Occup. Environ. Med. **59**, 67–72 (2002)
6. Steenstra, I., Cullen, K., Irvin, E., Van Eerd, D.: A systematic review of interventions to promote work participation in older workers. J. Safety Res. **60**, 93–102 (2017)
7. Job Seeker Nation Survey. https://www.jobvite.com/wp-content/uploads/2019/04/2019_Job_Seeker_Nation.pdf. Accessed 06 July 2020
8. Team mood. https://www.teammood.com/en/. Accessed 06 July 2020
9. Smets, E.: Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit. Med. **1**, Article number: 67 (2018)
10. Alberdi, A., Aztiria, A., Basarab, A.: Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review. J. Biomed. Inform. **59**, 49–75 (2016)
11. Muaremi, A., Arnrich, B., Tröster, G.: Towards measuring stress with smartphones and wearable devices during workday and sleep. BioNanoScience **3**(2), 172–183 (2013)
12. Vildjiounaite, E., Huotari, V., Kallio, J., Kyllönen, V., Mäkelä, S.-M., Gimel'farb, G., Unobtrusive assessment of stress of office workers via analysis of their motion trajectories. Pervasive Mob. Comput. **58** (2019)
13. Jia, M., Komeily, A., Wang, Y., Srinivasan, R.S.: Adopting Internet of Things for the development of smart buildings: a review of enabling technologies and applications. Autom. Constr. **101**, 111–126 (2019)
14. Garcia-Ceja, E., Osmani, V., Mayora, O.: Automatic stress detection in working environments from smartphones' accelerometer data: a first step. IEEE J. Biomed. Health Inform. (2016)
15. Koldijk, S., Neerincx, M.A., Kraaij, W.: Detecting work stress in offices by combining unobtrusive sensors. IEEE Trans. Affect. Comput. **9**(2), 227–239 (2018)
16. Mirjafari, S., et al.: Differentiating higher and lower job performers in the workplace using mobile sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **3**(2), 24, Article 37 (2019)
17. VTT Sensor Node. https://www.vttresearch.com/en/ourservices/energy-autonomous-wireless-sensor-networks#search=tiny%20node. Accessed 06 July 2020
18. PIR sensor. https://na.industrial.panasonic.com/products/sensors/sensors-automotive-industrial-applications/lineup/pir-motion-sensor-papirs/series/70516/model/73542. Accessed 06 July 2020
19. Azur table. https://docs.microsoft.com/en-us/python/api/azure-cosmosdb-table/azure.cosmosdb.table.tableservice?view=azure-python. Accessed 06 July 2020
20. Maxhuni, A., Hernandez-Leal, P., Sucar, L.E., Osmani, V., Morales, E.F., Mayora, O.: Stress modelling and prediction in presence of scarce data. J. Biomed. Inform. **63**, 344–356 (2016)
21. Peng, Y., Nagy, Z., Schlüter, A.: Temperature-preference learning with neural networks for occupant-centric building indoor climate controls. Build. Environ. **154**, 296–308 (2019)
22. tsfresh library. https://buildmedia.readthedocs.org/media/pdf/tsfresh/latest/tsfresh.pdf. Accessed 06 July 2020

23. Well-being: A strategy and a responsibility. https://www2.deloitte.com/us/en/insights/focus/human-capital-trends/2018/employee-well-being-programs.html. Accessed 06 July 2020
24. Lamb, S., Kwok, K.C.S.: A longitudinal investigation of work environment stressors on the performance and wellbeing of office workers. Appl. Ergon. **52**, 104–111 (2016)
25. Stewart, C.A., Krefeldt, M.: Your good health is a workforce issue. In: PEARC 2018, Article 4, 8 p. (2018)

Gender, Diversity and ICT

Gender, Diversity and ICT

Sisse Finken[1], Johanna Sefyrin[2], and Charles Ess[3]

[1] IT University of Copenhagen, Denmark
sisf@itu.dk
[2] Linköping University, Sweden
johanna.sefyrin@liu.se
[3] University of Oslo, Norway
c.m.ess@media.uio.no

The call for papers from the IFIP workgroup 9.8, Gender, Diversity and ICT, concerns "Data-work in relation to Gender and Diversity, Work, Educational and Daily Life" in an effort to explore work that unfolds with and around data; that is, work that takes place in the very collection, use, and undertakings of and on data – including the often invisible forms of work (Fisher et al., 2017, Bossen et al., 2019). This called for studies exploring the resources (social, cultural, material, digital), which are mobilized for digital data to serve in work, educational settings, and/or daily life from perspectives of gender and diversity that bring feminist concerns into the analyses and e.g. explore instances of data bias in such particular work (Perez 2019). At the same time, papers on other topics related to gender, diversity, and ICT were welcome. Out of the six papers submitted to WG 9.8, two were accepted. These papers are: "What Can Statistics Tell About the Gender Gap in ICT? Tracing Men and Women's Participation in the ICT Sector Through Numbers" by Morten Simonsen and Hilde G. Corneliussen, and "Silent Data, Active Patients" by Cæcilie Sloth Laursen and Sisse Finken. While these papers differ in the empirical material present and the analytical approaches used, they share in common a gesture towards gender, diversity, and ICT.

The paper by Simonsen and Corneliussen explores statistical material that contributes to an image of an uneven gender balance, both in the ICT work force and in ICT-related education. The paper asks what narratives available statistics can and cannot tell about men and women's participation in ICT related occupations – those occupations which in the available statistics are defined as related to ICT, e.g. ICT professionals such as software and applications developers and analysts. The study focuses on European countries with an emphasis on Norway, and confirms that the gender imbalance in ICT-related occupations remains significant, while suggesting that a continued attention to this issue is important. The statistics are placed in the context of the "Nordic Gender Paradox," the fact that the Nordic countries have a large level of vertical and horizontal gender segregation in educational fields and in the labor market, despite a high degree of gender equality. The study shows that when compared to the general rate of working women in a population, women are underrepresented in ICT-related education and ICT-related occupations in all of the European countries. Further, the paper shows that statistics from Norway display that while women have longer work hours in ICT-related occupations than in average across all occupations, more women than men work part-time in the ICT sector. In addition, women earn less than men for fulltime work. The findings, however, suggest that statistical analyses are not sufficient, and that qualitative research is an important complement and correction to statistical overviews, in particular for identifying factors that alone and in combination contribute to gender inequalities in ICT. The paper concludes with a notion that statistics only make some stories visible, while others remain invisible.

The paper by Sloth Laursen and Finken addresses the call's focus on studies about data work, and concerns how health care patients play an active role in managing their own illnesses: this active role is sustained by access to and use of health data provided by health care authorities through new digital technologies. The empirical material is gathered through ethnographic studies of health care patients' practices of reading health care data and their 'imaginaries' – an STS-related concept – about active involvement in their own health care. The authors explore local versions of sociotechnical imaginaries of the internet in a healthcare context. Inspired by previous studies, the authors discuss the transformative power of sociotechnical imaginaries in terms of how new patient identities become intertwined with a particular version of an internet-based imaginary. Based on this analytical approach, the analysis illustrates how patients become actively engaged when performing data work, when they try to understand the information in a digitalized patient record, for instance through searching the internet for complex medical terms. In conclusion, the study underscores that rather than leading to expected cost reductions and self-service, data access may, in some cases, bring patients to reach out to their hospital for dialogues about the data provided, and hence cause more work for the hospital staff and the patients. This contradicts the political strategies in which data is expected to speak for itself, and illustrates the local imaginaries enacted by the patients in contrast to the imaginaries expressed in political strategies of digitalization. The study points out that patients are required to become active, while data is silent, something which underscores how data is in need of active work before it becomes meaningful for, and can talk to, the patients.

References

1. Bossen, C., Pine, K., Cabitza, F.: Data work in healthcare: an introduction. Health Inf. J. **25**(3), 465–474 (2019)
2. Fisher, J., Crabtree, A., Colley, J., Rodden, T., Costanza, E.: Data Work: How Energy Advisors and Clients Make IoT Data Accountable. In: CSCW, vol. 26, pp. 597–626 (2017)
3. Perez, C.: Invisible Women - Exposing Data Bias in a World Designed for Men. Vintage Publishing (2019)

What Can Statistics Tell About the Gender Gap in ICT? Tracing Men and Women's Participation in the ICT Sector Through Numbers

Morten Simonsen and Hilde G. Corneliussen(✉)

Western Norway Research Institute, Sogndal, Norway
{msi,hgc}@vestforsk.no

Abstract. Which narratives can statistics tell about men and women's participation in ICT? The question is relevant across the western world showing a pattern of more men than women in ICT work. This chapter presents an analysis of available statistics that contribute to an image of women's participation in ICT work and education. The scope of the study is European countries with an emphasis on Norway, however, we also present statistics from OECD. The statistics confirm that the gender imbalance in ICT work is significant, suggesting that monitoring this field is important. The analysis also reveals challenges and gaps in the material, for instance the challenge of finding comparable numbers, a reduced use of gender as a variable in later years, difficulties in identifying the gendered structures of ICT due to a mixture of occupational fields for some of the relevant numbers, while other issues found to be relevant in qualitative studies are not represented in the available statistics. The monitoring of gendered structures of ICT work can be improved by developing statistics that better can capture inequalities and hierarchies. The findings also suggest that qualitative research is an important complement and correction to statistical overviews, in particular for identifying factors that alone and together contribute to gender inequalities in ICT.

Keywords: Statistics · ICT education · ICT sector · ICT work · Women · Gender

1 Introduction

The proportion of women in fields of information and communication technology (ICT) education and work is low across the Western World. This has been documented in qualitative [1–3] as well as quantitative studies [4–6]. In the Nordic Countries, this has been identified as part of a "Nordic Gender Paradox": despite a high degree of gender equality, the Nordic countries experience a high level of vertical and horizontal gender segregation in educational fields and the labor market, particularly notable in fields of ICT [7–9].

While at least 40% of each gender is often used as a goal for (near) "gender balance", the numbers for ICT education and work are lower – in Norway around 24%

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 379–397, 2020.
https://doi.org/10.1007/978-3-030-62803-1_30

and 21%. Research has identified that the lack of gender balance in male dominated environments of ICT creates several challenges for women, including making them appear “out of place” and challenging their feeling of belonging [10, 11]. The low number of women in ICT fields even has a tendency to reproduce a low expectation towards girls’ engagement in ICT [12]. Researchers’ interest for solving the “women and computing problem” is abundant [13] and numerous studies have documented that in order to increase gender equality in fields of ICT, it is vital to increase the proportion of women participating [14], thus, it is also important to monitor the situation closely.

In this paper we explore how the situation for women in ICT can be understood through available statistics. Numbers do not only tell us who participate in the field but can also tell stories about internal hierarchies and work cultures, for instance indicated by gender distribution in working time and salaries [15–17]. Statistics is often used to establish a starting point for understanding or exploring a field. The main research questions we pursue here are: which narratives can statistics tell about women in ICT? And, equally important, which stories cannot be told by available statistics? We approach these questions from our dual background in qualitative research within Feminist Technology Studies [18] that guides our understanding of the field, and in a tradition of engaging statistical material as a tool for identifying trends and tendencies in, for instance, working life [19].

Our use of statistics is explorative and not explanatory as we aim to identify which narratives the available statistics can tell about gender distribution in ICT work, part-time and full-time employment, and in salaries and participation in ICT education. An underlying premise for this work is the recognition that statistics is based on choices informed not only by research, but also by policies. The tables and figures of statistics presented below aim to give an overview of statistics that has been collected by public and national institutions to produce an image of women’s participation in ICT work and education, before we discuss how well the available statistics cover and give insight into what has been identified as important issues within qualitative research in the field.

The scope of this study is European countries with an emphasis on Norway, however, we also present statistics from OECD.

1.1 Navigating the Statistics

Statistics presented in this study come from national statistical agencies in Norway, EU (Eurostat) and OECD as well as from a report from EU’s Institute for Gender Equality [20]. Two code structures are relevant for classification of occupations. One is the NACE code structure developed by the EU[1] which represents “a statistical classification of economic activities in the European Community”[2] used across all member states. The other is ISCO developed by the International Labour Organization,[3] defined

[1] SSB - Norwegian (Statistics Norway).

[2] Eurostat, NACE Rev. 2.

[3] https://www.ilo.org/public/english/bureau/stat/isco/isco08/.
https://www.ilo.org/public/english/bureau/stat/isco/docs/groupdefn08.pdf.

as “a tool for organizing jobs into a clearly defined set of groups according to the tasks and duties undertaken in the job”.[4]

Occupations related to ICT are defined differently in the two structures. Statistics Norway (SSB) uses both structures. For data related to the category ‘occupation’ the ISCO structure is used. For the category ‘industry’ the NACE structure is used.[5] Generally, more data is available for the NACE structure than for ISCO.

2 Men and Women in ICT-Related Occupations in Norway

Table 1[6] shows number of employees in ICT-related occupations in Norway. The major group 25 in the ISCO structure[7] is for ICT Professionals. The table contains information about the subgroup 251 - Software and Applications Developers and Analysts.

Table 1. Number of employees in ICT related occupations (1000’s) (See also https://www.ssb.no/statbank/table/11411/ for salary earners) showing the ISCO subgroups *2512 Software developers* and *2519 Others (Classified as: Software and applications developers and analysts not elsewhere classified.)*

	2011	2012	2013	2014	2015	2016	2017	2018	2019
Men									
2512	11	11	12	11	11	10	10	14	17
2519	7	11	13	16	17	19	18	15	15
Women									
2512	2	2	1	2	3	2	2	3	6
2519	1	2	3	3	4	3	4	5	5
Percentage of women	14	15	14	16	20	15	18	22	26

Note that there were no available statistics for categories 2513 Web and multimedia Developers, 2514 Applications programmers as well as 252 - Database and Network Professionals because their samples are too uncertain to be published according to Statistics Norway.

A total of 43 000 people was employed in group 251 in Norway in 2019. Of these, 26% were women. There is a clear under-representation of women in these occupations, although the percentage share of women has risen by 8% points in 8 years from 2011. As a reference point, the general occupancy rate for women in Norway in 2019 was 48.1% of the total workforce.[8]

[4] ISCO Web site.

[5] https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.

[6] https://www.ssb.no/statbank/table/09792/.

[7] The code structure is based on ISCO-08.

[8] https://www.ssb.no/statbank/table/11153/.

Historically, there are no statistics for 1980–1995 for the occupations defined in Table 1. But there are some clues. From 1982 until 1993, Statistics Norway grouped employees into the category "computer operators". The share of women in this category was 83% in 1982 and 67% in 1993. Of course, this is not the same as computer developers today, since computers in this period also included back office tools operated by clerks. But it is still interesting to see that as computer technology advances, the share of women decreases. There was also a category for employees in "Post- and telecommunication". In the same statistics, number of employees in post services was given and thus telecommunication personnel can be calculated residually. The share of women in telecommunication was 44% in 1982 and 32% in 1995. Again, it seems that when telecommunication technology advances, the share of female employees declines.[9]

Table 2 shows number of salary earners for the same ISCO codes. The total number is lower than in Table 1, since employees also includes self-employed.[10]

Table 2. Number of salary earners in ICT related occupations. (https://www.ssb.no/statbank/table/11411/)

	2015	2016	2017	2018
Men				
2512 Software developers	3351	3500	3769	4155
2513 Web and multimedia developers	149	137	155	180
2514 Application programmers	92	116	149	170
2519 Others	10441	10567	10991	11718
Women				
2512 Software developers	391	405	407	460
2513 Web and multimedia developers	36	44	45	49
2514 Application programmers	34	49	48	51
2519 Others	2868	2775	2904	3076
Total	17362	17593	18468	19859
Women percentage	19%	19%	18%	18%

Table 3 shows median[11] monthly salary for the minor groups 2512 and 2519 distributed on working time (full time or part time), gender and sector for Norway 2018. The sectors are private (including public corporations) and public employed. The table is constructed as percentage of men's corresponding salary and as percentage of maximum salary which was earned by men in private sector. Women earn less than

[9] Statistics Norway, Arbeidsmarkedsunderøkelser 1980–95.

[10] See definitions here https://www.ssb.no/regsys.

[11] The median is used since the average salary level is more sensitive for very high or very low salaries. This means the average may fluctuate more because of more variation even if the median salary (middle income level) has not changed very much.

men measured by the median for full time but not for part time jobs. The female salary for full time workers varies from 93% to 96%. The difference between men and women are smallest where the salary level is lowest.

Table 3. Median monthly salary for women in the ICT sector distributed on working time and sector, in percentages (https://www.ssb.no/statbank/table/11418/)

	Percentages of men's salary	
	Full time	Part time
2512 Software developers		
Sum all sectors (private, state, municipality)	92.7	
Private sector and public corporations	92.6	
2519 Other software and applications developers		
Sum all sectors (private, state, municipality)	92.5	100.9
Private sector and public corporations	91.8	100.0
Municipal administration	96.3	

EU's Institute for Gender Equality (EIGE) has released a study [20] of employees in the ICT sector in different EU countries distributed by gender. The ISCO occupational categories included are 133 Professional managers, ICT professionals defined as sub-group 25 (see above) and ICT technicians defined as subgroup 35. Together, they form the category ICT specialists.

Table 4 shows number of employees in the EIGE definitions in Norway 2018, distributed by gender. All in all, 77 000 people are employed in these categories. Of these, 27 000 worked as ICT professionals as defined by EIGE. This group also includes code 2511 - System analysts and System architects. This group is not included in tables above, accordingly the sum of people employed in subgroup 25 in Table 4 is larger than the groups used above.

Table 4. Number of employees in different ICT subgroups according to ISCO structure and EIGE definitions, by gender, Norway 2018 (https://www.ssb.no/statbank/table/09792/. ISCO codes and definitions http://www.ilo.org/public/english/bureau/stat/isco/docs/d2434.pdf)

	Number of employees (1000's)	Percentage of total per gender	Percentage of total ICT workforce	EU % of total ICT workforce[a]	EU number of employees (1000's)
Men					
1330 Leaders of ICT-units	5	8	6.5		
2511 Systems analysts	16	26	20.8		
2512 Software developers	14	23	18.2		
2519 Software and applications developers and analysts not elsewhere classified	15	25	19.5		
Sum group 25	45	74			
3511 ICT operations technicians	11	18	14.3		
Sum group 35	11	18			
Sum men	61	100	79.2	83.5	7563
Women					
1330 Leaders of ICT-units	1	6	1.3		
2511 Systems analysts	4	25	5.2		
2512 Software developers	3	19	3.9		
2519 Software and applications developers and analysts not elsewhere classified	5	31	6.5		
Sum group 25	12	75			
3511 ICT operations technicians	3	19	3.9		
Sum group 35	3	19			
Sum women	16	100	20.8	16.5	1494
Sum both genders	77				

[a] https://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do, numbers are for EU-28 countries 2013–2020.

Table 4 shows that 21% of the total ICT workforce in Norway 2018 are women. The corresponding figure for EU-28 was 17%. According to NCWIT Scoreboard [21], 26% of all employed in Computer and Mathematical Occupations in USA in 2017 were women. Bailey et al. [22] claim that women occupy 25% of all IT jobs in USA in 2018. These categories are not directly comparable to Table 4 since job definitions may vary. Still they give the impression that women in Norway account for a smaller percentage of employees in the ICT sector than women in USA.

However, the increase of women in ICT jobs are higher in Norway than in USA. In 2011, the percentage of US women in ICT jobs was 25%. Thus, the increase from 2011 to 2018 is only 1% point, while the same figure in Norway was 8% points according to Table 1.

In Norway 2018, a total of 1 268 000 women were employed across all occupations and sectors. The same figure for men was 1 427 000.[12] Based on these figures and Table 4 we can construct Table 5 that shows the percentage of ICT specialists of the total workforce.

Table 5. Percentage of total working force in Norway employed as ICT workers in 2018 according to EIGE definitions

	Women	Men	Total	EU total 2015
Total employed (1000's)	1 268	1 427	2 695	
133 ICT managers	0.08%	0.35%	0.22%	0.1%
25 ICT professional	0.95%	3.15%	2.12%	1.4%
35 ICT technicians	0.24%	0.77%	0.52%	0.8%
Total ICT specialists	1.26%	4.27%	2.86%	2.3%

From Table 5, we observe that the percentage of men is larger than the corresponding percentage of women in each EIGE subgroup. The difference is largest for ICT professionals. Compared to EU figures for both genders, more Norwegian employees in the ICT sector are working as ICT professionals and less as technicians while for managers the figures are more in line with EU average.

According to EIGE 2018, the figure for Norway is almost identical with EU average when both genders are considered. We may also calculate the percentage that female occupancy rate account for in each EIGE subgroup. This is done in Table 6.

Table 6. Percentage of employed in EIGE ICT subgroups in Norway 2018, distributed by gender

ISCO code	Percentage of group		
	Women	Men	Total
133 ICT managers	17	83	100
25 ICT professional	21	79	100
35 ICT technicians	21	79	100
Total ICT specialists	21	79	100

[12] https://www.ssb.no/statbank/table/11153/.

2.1 European Union

Eurostat provides statistics for ICT specialists distributed by gender.[13] From 2011, Eurostat uses ISCO-08, the same code structure used in tables above.[14] In the Eurostat table, some other groups are included that are not included in the analysis above. These are:

- 2152 Electronic engineers
- 2153 Telecommunication engineers
- 2166 Graphic and multimedia designers
- 2356 Information technology trainers
- 2434 ICT sales professionals
- 3114 Electronics engineering technicians
- 7421 Electronics mechanics and servicers
- 7422 ICT installers and servicers

Including these groups, there are 123 800 employed ICT specialists in Norway in 2018. Using this definition for ICT specialists, the share of ICT specialists of total employed in Norway 2018 was 5.1%, slightly above EU average of 4.7%, but well below Finland (8.4%) and Sweden (7.8%). Table 7 shows that of ICT specialists in Norway 2018, 20.3% are women. The largest percentage of women is in Bulgaria with 28.3%, followed by Lithuania and Romania. The lowest percentage is found in Czechia (9.9%) and Hungary (8.5%). The average for EU-28 is 16.5%.

Table 7. ICT specialists (1000's) by gender in EU, EFTA and Turkey 2018 (Eurostat: Employed ICT specialists by sex, EFTA = European Economic Area).

GEO/UNIT	Male		Female		Total employed all sectors[a]	Percent ICT specialists
	1000's	%	1000's	%	1000's	%
EU[b]	7562.6	83.5	1 493.5	16.5	192872	4.7
Belgium	193.2	84.5	35.4	15.5	4112.4	5.6
Bulgaria	68.8	71.7	27.1	28.3	2721.6	3.5
Czechia	196.2	90.1	21.5	9.9	4329.7	5.0
Denmark	98.6	80.7	23.6	19.3	2536.3	4.8
Germany	1349.9	83.2	272.7	16.8	37299.7	4.4
Estonia	29.7	78.2	8.3	21.8	579.9	6.6
Ireland	83.1	81.7	18.7	18.3	1920.5	5.3
Greece	61.2	88.7	7.8	11.3	2542	2.7
Spain	519.5	83.8	100.1	16.2	16363.2	3.8

(*continued*)

[13] https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/DDN-20190513-1.

[14] https://ec.europa.eu/eurostat/cache/metadata/en/isoc_skslf_esms.htm#meta_update1554210955863.

Table 7. (*continued*)

GEO/UNIT	Male		Female		Total employed all sectors[a]	Percent ICT specialists
	1000's	%	1000's	%	1000's	%
France	887.0	83.0	181.3	17.0	23737.5	4.5
Croatia	49.5	85.7	8.2	14.3	1444.2	4.0
Italy	695.7	85.1	121.8	14.9	17650.8	4.6
Cyprus	10.5	81.8	2.3	18.2	347.5	3.7
Latvia	13.0	85.6	2.2	14.4	775	2.0
Lithuania	28.0	74.6	9.5	25.4	1176.8	3.2
Luxembourg	13.7	87.9	1.9	12.1	262	6.0
Hungary	151.5	91.5	14.1	8.5	3982.3	4.2
Malta	9.5	82.6	2.0	17.4	207.6	5.5
Netherlands	396.5	83.4	79.0	16.6	7275.5	6.5
Austria	156.5	81.6	35.2	18.4	3794	5.1
Poland	418.3	86.0	68.0	14.0	12853.8	3.8
Portugal	98.4	85.3	16.9	14.7	3992.4	2.9
Romania	145.4	76.5	44.7	23.5	6456.6	2.9%
Slovenia	32.6	83.7	6.3	16.3	835.9	4.7%
Slovakia	71.7	87.7	10.0	12.3	2177	3.8%
Finland	144.8	79.7	36.9	20.3	2171.7	8.4%
Sweden	274.1	79.1	72.4	20.9	4469.3	7.8%
UK	1 365.8	83.7	265.2	16.3	26857	6.1%
Iceland	6.5	84.5	1.2	15.5	167.9	4.6%
Norway	98.6	79.7	25.2	20.3	2443.9	5.1%
Switzerland	203.2	85.5	34.6	14.5	3943.4	6.0%
Turkey	244.8	89.6	28.5	10.4	19324.2	1.4%

[a] https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsq_eegais&lang=en, Q4 2018.
[b] 28 countries (2013–2020).

Table 8 shows development of percentage of female ICT specialists in EU plus EFTA and Turkey from 2015 to 2018. The table also shows the difference in percentage points over the period and it is sorted so that countries with the most negative development comes first. 14 countries have a negative trend with relatively less women employed as ICT specialists. The most positive development is found in Lithuania with an increase of 5.3% points. The most negative development is found in the neighboring country Latvia with a decrease of −10.3% points. Norway has the fourth most positive trend with an increase of 3.3% points.

According to these definitions, there were 6 100 more women employed in the ICT sector in Norway in 2018 compared to 2015. For all countries in EU, EFTA plus Turkey, about 1.6 million women were employed in the ICT sector in 2018, this is an increase of about 250 000 from 2015. For comparison, 1.1 million more men were employed in the ICT sector over the period with a total of 8.1 million men working in the sector in 2018.

Table 8. Development in percent of female ICT specialists from 2015 to 2018

	2015	2016	2017	2018	Difference 2018–2015
Latvia	24.7	24.8	21.3	14.4	−10.3
Iceland	22.6	21.9	16.3	15.5	−7.1
Romania	27.2	26.3	25.7	23.5	−3.7
Hungary	11.9	13.1	8.9	8.5	−3.4
Croatia	16.6	13.3	13	14.3	−2.3
Finland	22.4	21.9	21.8	20.3	−2.1
Greece	13.2	12.7	10.9	11.3	−1.9
Cyprus	19.7	23	17.4	18.2	−1.5
Ireland	19.7	21.1	20.9	18.3	−1.4
Spain	17.4	15.4	16.1	16.2	−1.2
Turkey	11.6	9.9	10	10.4	−1.2
Portugal	15.3	16.1	14.4	14.7	−0.6
Luxembourg	12.6	13.7	12.5	12.1	−0.5
Switzerland	14.6	14.9	14.9	14.5	−0.1
Czechia	9.9	11.2	9.3	9.9	0
United Kingdom	16.2	16.2	17.6	16.3	0.1
EU-28 (2013–2020)	16.2	16.7	17.2	16.5	0.3
Slovenia	16	17.3	16.1	16.3	0.3
France	16.6	18.1	19.6	17	0.4
Malta	17	12.1	10.2	17.4	0.4
Belgium	15.1	14.1	18.2	15.5	0.4
Germany	16.3	16.6	16.6	16.8	0.5
Poland	13.5	14.5	14.8	14	0.5
Bulgaria	27.7	30.2	26.5	28.3	0.6
Slovakia	11.4	9.2	13.7	12.3	0.9
Denmark	18.4	19.6	19.1	19.3	0.9
Italy	13.8	14.2	16	14.9	1.1
Estonia	20.3	18.7	19.4	21.8	1.5
Sweden	18.9	20.8	20.9	20.9	2
Norway	17	19.4	19.5	20.3	3.3
Netherlands	13	15.6	16.6	16.6	3.6
Austria	14.2	17.2	15.6	18.4	4.2
Lithuania	20.1	24.8	25.7	25.4	5.3

3 ICT and Working Time

Statistics Norway has information on employees in the ICT sector distributed on working time (full or part time) and gender. This information is only available to 2015 and comprises NACE codes 58-63.[15] This means that in addition to the ICT sector, this statistic also contains occupations such as publishing and production of movies, radio and television programmes. Table 9[16] shows that for both genders, there has been a 12% growth in this sector from 2008 to 2015. The growth has been largest for part time men (61%) and slowest for full time women (7,7%). In 2008, the percentage working part time was 3.3 times larger among women than among men, the corresponding figure in 2015 was 2.4.

Table 9. Percentages working full time and part time in ICT and information sector (NACE 58-63) Norway distributed by gender and year

Year	Women employed		Men employed		Both genders, number of employees		
	Full time	Part time	Full time	Part time	Full time	Part time	Sum employed
2008	87%	13%	96%	4%	43974	3303	47277
2009	87%	13%	96%	4%	46187	3484	49671
2010	87%	13%	96%	4%	47949	3698	51647
2011	87%	13%	96%	4%	46446	3635	50081
2012	88%	12%	96%	4%	49413	3645	53058
2013	89%	11%	96%	4%	48409	3212	51621
2014	89%	11%	96%	4%	51186	3608	54794
2015	87%	13%	94%	6%	48864	4267	53131

4 ICT and Salary

Table 10 shows monthly salary for women as a percentage of corresponding male salary in Norway. The table is distributed on working hours in ICT and communication sector from 2008 to 2015. The table shows that women are paid less than men, both as full time and part time employed. The difference is largest for full time employed. In 2008 women on average earned 83% of men's salary as full time employed. In 2015 the figure was 86%. In other words, the gap is closing but not very fast. Among part time employed, women's salary on average was 94% of men's in 2008 while the same figure in 2015 shows a negative trend with 91%.

[15] Subgroup 61 is defined as telecommunication, subgroup 62 as services associated with information technology (programming, system management) and subgroup 63 as information services (data processing, data storing and management of web portals). The last subgroup also includes information services such as news agencies.

[16] https://www.ssb.no/statbank/table/07597/.

Table 10. Monthly salary for employees in ICT and communication sector (NACE 58-63) distributed by gender and working hours Norway. Female percentage of male salary

Year	Women, percentage of men's salary	
	Full time employed	Part time employed
2008	83%	94%
2009	83%	92%
2010	84%	92%
2011	83%	90%
2012	85%	93%
2013	86%	94%
2014	86%	94%
2015	86%	91%

5 ICT and Inconvenient Working Hours

Statistics Norway also provides statistics about inconvenient working hours.[17] This statistics include number of employees in NACE codes 58-63 with inconvenient working hours from 2008 to 2019. These NACE codes include employees working in radio and television broadcasts, motion picture and video production as well as in publishing and news agencies in addition to the ICT sector. This is a source of bias since these occupancies by design (deadlines etc.) have more inconvenient working hours.

Number of employees working inconvenient hours has gone down over the period 2008 to 2019. The biggest reduction in inconvenient working hours from 2008 til 2019 was on Saturdays and Sunday. Most of the inconvenient hours are worked on evenings. About 30% of employees working infrequent on evenings were women in 2019, compared to 24% women working in the ICT sector defined as NACE codes 61, 62 and 63. This shows that working inconvenient hours is common among ICT female workers, but not as common as for other occupancies in the information and communication sector.

6 Education and Employment

6.1 OECD

According to OECD,[18] 3.6% of all master graduates in Norway in 2017 are in the field of ICT defined according to ISCED[19] (level 7). Of all female graduates in Norway same year, 1.3% are in ICT fields while the same figure for men was 6.6%. This means

[17] https://www.ssb.no/statbank/table/09883/tableViewLayout1/.

[18] OECD: Distribution of graduates and entrants by Field, https://stats.oecd.org/Index.aspx?datasetcode=EAG_GRAD_ENTR_FIELD.

[19] UNESCO - ISCED 2011.

that the gender gap for ICT graduates in Norway, defined as male percent minus female, was 5.4% points in 2017.

Figure 1 is based on distribution of gender gap in the ICT education field for all OECD countries. The figure shows ICT share of all graduates among women on the x-axis and the gap between male and female share on the y-axis. Among the Nordic countries, Finland has the greatest gap with 9 while Sweden has 2.4. In Sweden, the percentage of graduates in ICT is the lowest among the Nordic countries when both genders are considered. Estonia has the largest gender gap where the percentage of ICT graduates among males is 9.1% points larger than among females.

Interestingly, there seems to be a trend since the gap is increasing when the female share of graduates in the ICT field is increasing. In other words, when the share of female graduates in the ICT field increases, the corresponding share of men increases even more and the gap between men and women increases, as shown in the figure. This is obviously a feature that should be addressed in further qualitative research.

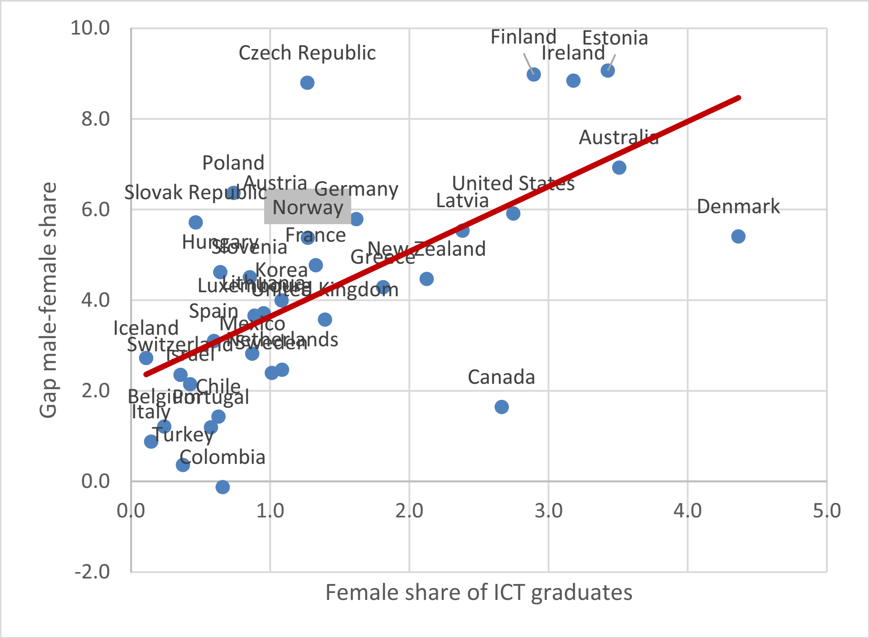


Fig. 1. Female share of graduates in the ICT field vs the gap between male and female shares (Color figure online)

6.2 ICT Education in Norway

The table below shows number of applicants for ICT education in Norway 2008–18 by gender.[20] The table shows that the percentage of female applicants have risen by 6.8%

[20] https://www.samordnaopptak.no/info/om/sokertall/sluttstatistikker/so_sokerstatistikk_2017-sluttrapport.pdf.

points over the whole period. The yearly increase was largest between 2015 and 2017 with an increase of 2.5% points each year.

Table 11. Applicants for ICT education Norway 2008–18 (Data for 2017 and 2018 are taken from statistics documentation in 2018, https://www.samordnaopptak.no/info/om/sokertall/sluttstatistikker/ by gender)

	Women	Men	Total	% Women
2008	341	1617	1958	17.4
2009	305	1617	1922	15.9
2010	359	1678	2037	17.6
2011	375	1923	2298	16.3
2012	441	2189	2630	16.8
2013	450	2297	2747	16.4
2014	500	2476	2976	16.8
2015	609	2887	3496	17.4
2016	756	3039	3795	19.9
2017	1118	3862	4980	22.4
2018	1461	4566	6027	24.2

Table 1 above showed that 22% employed in the ICT sector in Norway in 2018 were women. The percentage of female applicants for ICT education in 2018 is therefore slightly above the percentage of women working in the ICT sector. There are two possible explanations for this. Since the table shows applicants, more women might leave during the education. Or more women than men with ICT education work in other sectors than ICT. Our recent qualitative research among women working with ICT indicates that the second answer should be further explored [23], thus also suggesting that more qualitative research is necessary to complement the statistical data.

6.3 European Union

Eurostat provides information on number and percentage of different genders with ICT education employed anywhere in the economy.[21] The percentage basis is number of people with ICT education employed, not necessarily in the ICT sector.[22]

The table below shows the percentage of women with ICT education in any employment situation. The table also shows the trend from 2014 to 2016 calculated as the change in percentage points of females with ICT education occupied in that time span. The table is sorted so that countries with the most positive trend appear first in the table. The percentage for Norway corresponds roughly to the percentage of women

[21] Eurostat https://ec.europa.eu/eurostat/web/products-datasets/-/isoc_ski_itsex.

[22] "3.3. Coverage - sector Data on persons with ICT education does not use the concept of sectors of economic activities. Persons with ICT education can be employed in any sector or be unemployed."

employed in the ICT sector. This means that most women with ICT education in Norway work in the ICT sector.

Cyprus has the most positive development. Norway and Lithuania are the only countries in addition to Cyprus with a change of more than 7% points. The country with the fourth most positive trend, Portugal, has a change that is less than half of Norway's (Table 12).

Table 12. Women with ICT education employed as percentage of all employed with ICT education with trend between 2014 and 2016

	2014	2015	2016	Difference 2016–2014
Cyprus	25	33	32.6	7.6
Lithuania	12.1	9.3	19.2	7.1
Norway	13.6	21.7	20.7	7.1
Portugal	20.8	24.3	24	3.2
Czechia	9.4	10.8	12	2.6
France	11.1	14.1	13.6	2.5
Spain	18.6	23.3	20.4	1.8
Hungary	13.2	13.8	14.8	1.6
Slovenia	12.2	15	13.2	1
Austria		13.7	14.2	0.5
Belgium	12.5	9.5	12.7	0.2
Germany	13.3	13.3	13.4	0.1
Latvia	19.7		19.5	−0.2
Denmark	16.1	17.3	15.9	−0.2
Switzerland	15	16.5	14.6	−0.4
Poland	11.8	10.8	11.2	−0.6
Malta	18	19	16.9	−1.1
Sweden	26.1	26.3	25	−1.1
Ireland	30.5	27.7	28.8	−1.7
Turkey	25.8	26.4	23.9	−1.9
Romania	30	27	27.7	−2.3
Slovakia	14.5		11.6	−2.9
Italy	21.4	18.9	18.2	−3.2
Netherlands	14.6	17.1	11	−3.6
Iceland		33.6	27.2	−6.4
United Kingdom	24.2	23.7	16.9	−7.3
Greece	36.9	34.2	26.9	−10
Estonia	26.9	19.6	15.6	−11.3
Finland	33.1	29.4	19.9	−13.2
Croatia	22.6	10.3	6.1	−16.5
Bulgaria			29.2	

The table also shows that Cyprus has the highest percentage of women among employees with ICT education in 2016, taking over for Greece and Iceland that was on top in previous years. Bulgaria, Ireland, Malta, and Romania are the countries with the largest percentage after Cyprus. Norway is in 11th place, after Sweden but well ahead of Denmark. The table, however, shows a rapid fluctuation in some countries that might also be caused by changes in the definitions behind the numbers (Table 12).

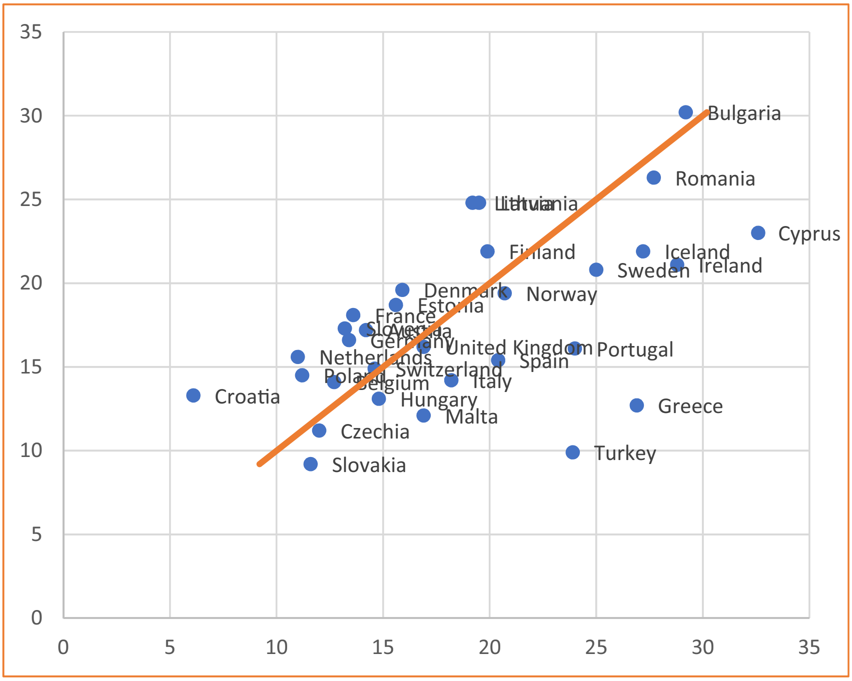


Fig. 2. Women as percentage of all ICT specialists and as percentage of all employed in the ICT sector

Figure 1 shows percentage of women with ICT education vs percentage of women working in the ICT sector in 2016. The red trendline shows the situation if percentage of women employed in ICT sector would be equal to percentage of women with ICT education. In other words, countries where all women with ICT education work in the ICT sector would be on the red line. For countries below the red line, women with ICT education are mostly working outside the ICT sector. For countries above the red line, most women working in the ICT sector do not have ICT education.

Norway is slightly below the red line, meaning many women with ICT education, but not all, is working in the ICT sector. It is worth noting that the percentage of women working in the ICT sector has grown in Norway from 2016 to 2018. This may indicate that Norway would be closer to the red line if 2018 data was used instead of 2016 data. Bulgaria, UK and Switzerland are the countries where the two percentages are most equal. Cyprus has a large percentage of women with ICT education in employment but relatively fewer working in the ICT sector. Greece and Turkey have relatively large percentage of women with ICT education but few of them working in the ICT sector. Lithuania and Latvia have relatively more women working in the ICT sector than women having ICT education.

7 Visible and Invisible Narratives About Women in ICT

The goal of this study was to explore which narratives statistics can and cannot tell about women in ICT, with a main focus on Norway and compared to European countries. The statistical material reviewed here gave us answers to the gender distribution in ICT, showing a clear underrepresentation of women in ICT education and ICT work in all European countries compared to general female occupancy rate. The statistics from Norway showed that while women have longer work hours in ICT than in average across all occupations, more women than men work part-time in ICT. Women also earn less than men for fulltime work in the ICT sector, however, these tables have not included gender since 2016. The percentage of women working inconvenient hours in ICT is difficult to measure, because the available numbers include other occupational groups that are likely to be more affected by inconvenient hours (e.g. news agencies) than the ICT occupations. Thus, while the available statistics provide important insights into trends of women's participation in ICT work, there are also some weaknesses that blur the gendered structures in ICT work.

Another challenge is to find identical and comparable numbers across regions and nations, which is complicated not only due to different code structures, but also different types of measures. Given the complicated nature of the statistics, some of the questions we started with in this project were not possible to answer, in particular numbers combining different factors that each have been identified with gendered patterns, like public vs. private sector, salary, working time, part-time work. Other questions did not have one, but several answers. The question of women's occupational rate in ICT in Norway, for instance, was answered in different ways, showing that out of a total of 123 800 employed as ICT specialists in Norway in 2018, the number of women ranged between 20.3 and 22% (ISCO subgroup 251); female ICT specialists made up 1.26% of the total working force; and women made up 20.7% of all employed with ICT education in 2016.

This last piece of information is important, but not sufficient for measuring the "leaky pipeline" – women leaving ICT work, as it indicates that many, but not all women with an ICT education are employed or work in the ICT sector. Considering that "leakage" of women is recognized as a major problem in other western countries [24], it would be valuable to see statistics on this for Norway. Based on our qualitative research in the field showing that routes to ICT work are still highly gendered in Norway [23], we also miss statistical accounts of the entire "pipeline", not only the "leakage" of women, but equally important, how and when recruitment of women is successful. Monitoring men and women's movements from education to the labor market could help to identify whether the main challenge for increasing gender balance in ICT in Norway is "an input" or "a throughput" problem [25].

Although the different numbers suggest that a careful consideration of statistics is necessary if a precise comparison is the goal, the numbers also document a substantial gender imbalance with more men than women choosing a career in ICT. Even more worrying for the development is the trend that indicates that when the number of women in ICT education increases, the number of men in ICT increases even more, intensifying the gender gap in ICT.

Provided that gender equality including a more gender balanced workforce in ICT is a goal [20], it is critical that the combination of variables of gender and ICT are possible to identify in vital working life statistics. The trend revealed here is unfortunately not only going in the right direction, but rather also showing that with regards to some issues, the accuracy of what statistics can tell about women in ICT is rather reduced. Our exploration of public available statistics, however, confirms that monitoring of gendered structures in the ICT sector, in ICT work and regarding work conditions, can be improved by developing statistics that can show inequalities and hierarchies in the ICT sector and ICT work.

Finally, we came to this task of exploring which stories statistics can tell about women's participation in ICT from qualitative research in this field. The variation in numbers we have seen here indicates that qualitative research is an important complement and correction to statistical overviews, in particular for identifying factors that alone and together contribute to gender inequalities in ICT. Statistics are not pre-given; they are motivated to make visible certain stories while others remain invisible, thus it is critical that national statistics is constantly evaluated and updated in line with new insights from qualitative research.

Acknowledgement. This work was funded by *NORDWIT,* a Nordic Centre of Excellence, studying women in tech-driven careers in the Nordic countries. Read more about Nordwit: https://nordwit.com/.

References

1. Misa, T.J. (ed.): Gender Codes: Why Women are Leaving Computing. IEEE Computer Society and Wiley, Hoboken (2010)
2. Cohoon, J.M., Aspray, W. (eds.): Women and Information Technology. Research on Underrepresentation. MIT Press, Cambridge (2006)
3. Frieze, C., Quesenberry, J.L.: Cracking the Digital Ceiling: Women in Computing Around the World. Cambridge University Press (2019)
4. Charles, M., Bradley, K.: A matter of degrees: female underrepresentation in computer science programs cross-nationally. In: Cohoon, J.M., Aspray, W. (eds.) Women and Information Technology Research on Underrepresentation, pp. 183–203. MIT Press, Cambridge (2006)
5. Vabø, A., Gunnes, H., Tømte, C., Bergene, A.C., Egeland, C.: Kvinner og menns karriereløp i norsk forskning: En tilstandsrapport. Rapport 9/2012, NIFU, Report No.: 8272188201 (2012)
6. Chow, T., Charles, M.: An inegalitarian paradox: on the uneven gendering of computing work around the world. In: Frieze, C., Quesenberry, J.L. (eds.) Cracking the Digital Ceiling: Women in Computing around the World, p. 25 (2019)

7. Roivas, S., Corneliussen, H., Jensen, L., Hansson, A., Mósesdóttir, L.: Meta-analysis of gender and science research–country group report, Nordic countries, vol. 20, no. 02, p. 2016 (2010). Disponibile on line al seguente link: http://www.genderportaleu/sites/default/files/resource_pool/CG-R3_Nordic.pdf
8. Halrynjo, S., Teigen, M. (eds.) Ulik likestilling i arbeidslivet. Gyldendal Akademisk, Oslo (2016)
9. Gunnes, H., Hovdhaugen, E.: Karriereløp i akademia: Statistikkgrunnlag utarbeidet for Komité for integreringstiltak - Kvinner i forskning: NIFU STEP (2008)
10. Corneliussen, H.G.: Gender-Technology Relations: Exploring Stability and Change. Palgrave Macmillan, Basingstoke (2011)
11. Riegle-Crumb, C., Morton, K.: Gendered expectations: examining how peers shape female students' intent to pursue STEM fields. Front. Psychol. **8**(329) (2017). English
12. Corneliussen, H.G., Prøitz, L.: Kids Code in a rural village in Norway: could code clubs be a new arena for increasing girls' digital interest and competence? Inf. Commun. Soc. **19**(1 (Special Issue: Understanding Global Digital Cultures)) (2016)
13. Vitores, A., Gil-Juárez, A.: The trouble with 'women in computing': a critical examination of the deployment of research on the gender gap in computer science. J. Gender Stud. **25**(6), 666–680 (2016)
14. Margolis, J., Fisher, A.: Unlocking the Clubhouse. Women in Computing. MIT Press, Cambridge (2002)
15. Watts, J.H.: 'Allowed into a man's world'. Meanings of work–life balance: perspectives of women civil engineers as 'minority' workers in construction. Gender Work Organ. **16**(1), 37–57 (2009)
16. Belgorodskiy, A., Crump, B., Griffiths, M., Logan, K., Peter, R., Richardson, H.: The gender pay gap in the ICT labour market: comparative experiences from the UK and New Zealand. New Technol. Work Employ. **27**(2), 106–119 (2012)
17. Padavic, I., Ely, R.J., Reid, E.M.: Explaining the persistence of gender inequality: the work–family narrative as a social defense against the 24/7 work culture. Adm. Sci. Q. 0001839219832310 (2019)
18. Bray, F.: Gender and technology. Annu. Rev. Anthropol. **36**, 37–53 (2007)
19. Babbie, E.: The Practice of Social Research Belmont, 12th edn. Wadsworth, USA (2007)
20. EIGE. Women and men in ICT: a chance for better work–life balance - Research note. Luxembourg: EIGE: European Institute for Gender Equality, Publications Office of the European Union (2018)
21. DuBow, W., Pruitt, A.: NCWIT Scorecard: The Status of Women in Technology. NCWIT, Boulder (2018)
22. Bailey, M., Riley, S.: Women in Tech: Unconscious Bias, Parity, and the Path Forward (2018). https://mailchi.mp/57c92dac9f60/2018-women-in-tech-unconscious-bias-report
23. Corneliussen, H.G., Seddighi, G., Dralega, C.A.: Women's experience of role models in IT: landmark women, substitutes, and supporters. In: Helgesen, Ø., Nesset, E., Mustafa, G., Rice, P., Glavee-Geo, R. (eds.) Modeller: Fjordantologien 2019: Universitetsforlaget (2019)
24. Branch, E.H. (ed.): Pathways, Potholes, and the Persistence of Women in Science: Reconsidering the Pipeline. Lexington Books, Lanham (2016)
25. McKinney, V.R., Wilson, D.D., Brooks, N., O'Leary-Kelly, A., Hardgrave, B.: Women and Men in the IT profession. Commun. ACM **51**(2), 81–84 (2008)

Silent Data, Active Patients

Cæcilie Sloth Laursen and Sisse Finken

IT University of Copenhagen, Rued Langgaards Vej 7, Copenhagen S, Denmark
{cael, sisf}@itu.dk

Abstract. With the wake of digital welfare, governments advocate that patients play an active role in managing their own illnesses. This active role is sustained by access to and use of health data provided by health care authorities through new digital technologies. Stepping into an empirical site where patients log in to their own site, 'MyChart', we inquire their practices reading health care data and their imaginaries about active involvement in their own health care. With this, our analysis focuses on *the active patient* and aims to bring forth local imaginaries in an effort to nuance data imaginaries located in political strategies, which relate data access with active partnerships. Within this, we illustrate how patients are active, while data is silent and in need of work before it vocals meaningful for the patients.

Keywords: Digital health care · Active patients · Sociotechnical imaginaries · Data work

1 Introduction

Within healthcare, access to health data is imagined to enable citizens play an active role in their own course of diseases and become active partners. The question remains, however, what it means to be active? When reading the Danish national 'strategy for digital health', which couples data access with patients' engagement in active partnerships, we learn that: "Patients should have access to their own data, in order for them, for instance, to have better opportunities for participating actively in their own treatment" [1][1]. While advocating for access to health data, the strategy spends little time clarifying what an active partnership implies beyond proclaiming more involvement and self-service [2][2]. As such, the notion of *active patients* echoes the agenda of patient participation, which the World Health Organization started advocating in the late '70s [3]. While patient participation is linked with the concepts of patient-centeredness and patient empowerment [3], Lupton relates such discourses and the implementation of digital technologies with assumed economic efficiencies [4].

Drawing on Jasanoff, this article argues that discourses of *active patients* in policy-documents start to form a 'sociotechnical imaginary'. An imaginary enabled by digital solutions allowing data access, while, at the same time, also supportive of and

[1] p. 5.

[2] p. 22.

Published by Springer Nature Switzerland AG 2020
D. Kreps et al. (Eds.): HCC14 2020, IFIP AICT 590, pp. 398–408, 2020.
https://doi.org/10.1007/978-3-030-62803-1_31

sustaining such developments [5]. As we will see below, the imaginary of *active patients* seems to have transgressed onto patients as well.

Taking the patient portal MyChart as an example of a new data infrastructure that provides patients access to their health data, we present empirical examples that illustrate what happens in practice when patients actively engage with their health data through this new digital healthcare solution. Our findings suggest that, while patients welcome the idea of access to their own data, such accessibility can carry with it implications when the data is laborious to understand. Thus, the paper gives examples of how patients perform ‘data work’ [6, 7] and discusses the potential benefits and pitfalls of their data access. As such, our study brings nuance to the notion of the *active patient* alongside reflections on normativities [8] that such ‘imaginary’ [5] produces.

2 Digitalization and Patient Engagement

While we see linkages between data and *active patients* in political strategies concerning digitalization, we have been less fortunate to locate scientific literature concerned with patients’ own conceptualizations of being active patients and their experiences with data access. We have found one recent Swedish survey-study, where patients’ experiences with access to a national electronic health record are explored. In this study, the authors conclude that access makes patients feel more involved in their treatment [9]. However, one of the few existing studies of MyChart concludes that in its present form, the system might not support patients’ active engagement in their treatment, in part because they cannot interpret the information available [10]. Vikkelsø has analyzed four information infrastructures for patient-centered care in Denmark, one of them being online data access. She finds that patients might both need guidance to understand their record, that it can cause worries, and that the language used in the record ought to change [11]. In relation to language usage in patient records, a study of nursing documentation practices in Norway finds that after patients gained access to their record the nurses focused on precise documentation, but with an attentiveness towards avoiding abbreviations and Latin expressions [12]. Similarly, a study of a shared EPR that replaces a client-held record, which pregnant women bring along to visits at general practitioners, midwives, hospitals, etc., found that the women’s demand for completeness of the record challenged the care professionals’ practices [13].

Several studies remind us that the implementation of new technologies often means redistribution of work rather than minimization of work [14–17]. In studies of telemedicine devices Oudshoorn [15] and Andersen [16], for instance, describe how patients become ‘diagnostic agents’ and improve their ‘diagnostic skills’.

From our methodological standpoint, patients are often the “implicated actors” [3] [19] of new digital healthcare solutions. For this reason, it is paramount to investigate what happens in practice as patients take these new technologies into use.

[3] While not the focus of this study, clinicians might also be implicated actors, since patients’ increased access to their record and new forms of consultations are likely to change current work practices (see [18]).

3 Sociotechnical Imaginaries and Data Work

The study has been guided by a theoretical positioning within the field of STS. This implies an attentiveness towards the networked entanglements of humans, technologies, institutions, meanings, practices, etc. [20]. Our analysis particularly engages with the concept of ‘imaginaries’ [5, 21], which have been an analytical focus for STS scholars within the past few decades [22]. Jasanoff [5, p. 6] describes sociotechnical imaginaries in the following way:

> “collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology”.

In this paper, we proceed with conceptualizing the notion of *active patients* as a collective imagination of a desired future for digital health care, tightly coupled with developments within information and communications technology (ICT). It is an imaginary producing normatives for social life and social order. Drawing on Jasanoff, Felt [23] explores local versions of sociotechnical imaginaries of the internet in a healthcare context. In her study of Austrian citizens’/patients’ use of the internet, as an information medium, she shows the transformative power of the sociotechnical imaginary by demonstrating how new patient identities become intertwined with a particular version of the imaginary of the internet. She, for instance, explains how the opportunity to get informed through searches on the internet also creates an “obligation to do so” [23][4], which, simultaneously, caters to a practice where patients participate actively in the collective care infrastructure [23]. In line with Felt [23], this study unfolds local versions of the *active patient* by analyzing how Danish patients conceptualize themselves as active, as well as how they engage with health data made available through MyChart through the lenses of ‘data work’ [6]. Data work is “any human activity related to creating, collecting, managing, curating, analyzing, interpreting, and communicating data” [6, p. 466].

4 Methods and Empirical Setting

The patient portal, MyChart, which serves as a hub for the ethnographic study presented here, is part of a larger IT-system or, more specifically, of an Electronic Patient Record (EPR) designed by Epic. The EPR was adapted for the Danish market (and called The Health Platform (red. Sundhedsplatformen) and implemented in the Capital and Zealand regions of Denmark during 2016 and 2017. The Health Platform gathers patients’ data in one place to be shared across hospitals and hospital units in the regions. Further, the Health Platform is designed to better support clinical staff in diagnosing and planning treatment trajectories [24]. Next to supporting clinical staff, the Health Platform is designed with the integrated patient portal MyChart (red. Min Sundhedsplatform), which allows patients to have written communication with the

[4] p. 188.

hospital, to access their health data, including doctor's notes, test results, and upcoming appointments [25]. According to the Capital Region "The Health Platform makes it easier for the patient to play an active role" [25] and thus, the system supports the line set out in the current Danish strategy for digital healthcare, which couples data access with patients' engagement in active partnerships [1, 2].

The paper builds on an ethnographic study [26, 27] conducted by the first author for her master's thesis [18]. The study includes fieldwork at an outpatient clinic in a Copenhagen based hospital on selected days from December 2017–August 2018, visits in homes/workplace of five patients, document analysis of internal documents from the hospital, system descriptions, and political strategies.

The patients were between the ages of 20–60 years old and affiliated with the outpatient clinic, either because they had a chronic disease, or were in the process of being diagnosed. The patients were interviewed about their experiences with MyChart, how they understand the notion of the *active patient*, and their understandings of being active. The patients were also observed while using the patient portal, which sparked reflections about specific functions in the system. The interviews were semi-structured, conducted in Danish, and lasted on average a little over an hour. The field material used in this paper is translated from Danish to English; the same applies to quotes from various strategies and web pages produced by Danish organizations or governmental agencies.

At the time of the data generating, three of the patients were students, and in their early or mid-twenties: Niels[5] had had his diagnosis since he was a child; Emma was recently diagnosed with a chronic condition; and Astrid was in the process of being diagnosed. The fourth patient, Miriam, was also in the process of being diagnosed. She was in her late twenties and worked in the healthcare sector. The fifth patient, Lisbeth, is a 60 years old woman having two chronic conditions. For this reason, she received treatment at two different outpatient clinics.

All fieldnotes, interview transcripts, and documents were coded and analyzed using Situational Analysis [19].

This paper includes empirical data used in [18] as well as material, which did not make it to the final edition of her master thesis. In addition to a new literature survey on *active patient*s and added theoretical readings, new political documents have been gathered for this particular analysis.

5 Silent Data and Active Patients

In situating the *active patient*, we move into the empirical site where patients are interviewed while accessing their health data via MyChart.

[5] All patients appear with fictitious names to protect their privacy.

5.1 Patients as Active Interpreters of Data

As we have seen, access to own health data is imagined to be one of the key features enabling an active partnership. However, while data can be accessed, its meaning might not be attainable and transparent for patients, especially the test results prove challenging to comprehend. An example of this, we find in Niels's reflections when he accesses his blood test results:

> *He clicks on a result* "Oh there I see my values and the standard intervals. Here I am within the interval." *Opens another result* "Here I am in the high end. (…) Here I am also in the high end. Is my result too high or what? Is it dangerous?".

The blood test results in MyChart are presented both with a standard interval and the patient's results for a specific blood value. When Niels accesses his blood test results, he immediately attempts to make sense of and interpret what he is seeing. Niels compares his results to the standard intervals and reacts when they are outside the standard. He is confused whether a result, which is outside or on the verge of the standard interval, is dangerous, and it sparks a degree of concern and him asking if 'it is dangerous?'. During the interview, Niels googled[6] some of the names of the blood values, such as 'Monocytes', to gain information about the different values and their meaning. The actions Niels takes towards interpreting his results can be viewed as 'data work' [6, 13], which he performs in order to make sense of his results. Niels is excited about the access to his health data even though some of the information is difficult to understand. It, for instance, sparked his curiosity that he is able to see the data forming the basis for the physician's assessment.

The other patients in our study also found it interesting to be able to access the same data as the physician. However, they were unsure of how the data was useful to them since they could not understand it. The participant Emma, for instance, said:

> "I was looking at these 40 new test results because [my physician] said to me, 'I will see you in 2 weeks' (…), and until then, I could go and have a look at my test results in here [MyChart]. And I thought that is fine, but then I logged in, and I don't understand anything of what it says."

It is not only the test results that are difficult for the patients to understand. Also, the physicians' notes from consultations are difficult, since they are written in, what Emma calls, 'medical-lingo'. Emma consults her mother, who is a physician, to understand the information on MyChart. Emma imagines that her physician will teach her to interpret her own blood test results. Astrid, in a similar line, imagines that by hearing her physician explain the results during consultations, she will eventually learn to read the results herself. She says:

> "Over several years, I might learn what the results mean. I like that thought. If I will learn, that's another story."

These empirical examples of the patients' engagement with their health data show that while the data is accessible, it is not necessarily understandable for them. It also

[6] This is the term our participants use for performing searches in Google's search engine, why we maintain this phrasing.

shows how data access does not automatically make an *active patient* who can take proactive steps in her/his patient trajectory.

5.2 Data in Need of Work

When Niels is asked if he sees MyChart as encouraging patients to take an active role, he says:

> "Is the activity simply keeping yourself up to date, as what I am doing now, then yes. But it is very few Danes who would read their phosphorus value in the blood test results and think to him or herself: 'Wow I really need some phosphorus'. In this way there is not much activity in it [red. MyChart]. (…) I definitely think the test results should be there, but they might fall short. You need to be more than averagely active in order to use it."

In our view, Niels sums up the issues pertaining to data access as an enabler for being an *active patient*. As it is at present, there is no translation process when data is pulled from the clinicians' system in the Health Platform to the patients' MyChart. That is, the data is the same and it proves difficult for the patients to understand the medical terminology and test results without explanation. In line with this, Haraway [28] reminds us of situatedness of knowledges and knowledge production. Likewise, from a healthcare perspective, Berg [29] points to the context-dependent qualities of the information in the patient record and how it is directed towards clinical personnel in order to serve their work. The particularities of the record becomes very visible once it is put in a different context. As a physician in our study stresses; the patient record is a work tool for him and his colleagues, which is why it is written the way it is[7]. However, with the increased patient access to e-records, the current configuration of the data in patient records might be challenged. Vikkelsø [11][8] states that it requires balancing "to ensure that patient records can simultaneously function as professional decision-making tools, legal document and patient-oriented summaries". However, she also writes that, if patients should be able to understand what is written in the record, the writing style must change [11]. The question is whether this will happen, and if it is possible to be as precise without the medical terminology?

In Miriam's opinion, she has no use of the test results due to the way they are presented. She suggests that the results should be presented with comments:

> "I think there should be some comments attached stating that everything is okay, or that there is a lack of something, or something like that".

Lisbeth, on the other hand, has no expectation of such. She is used to using another health data platform, which has a small dictionary embedded in the website and instead she hopes that a similar dictionary will be integrated into MyChart, allowing her to look up medical terms she doesn't understand.

It seems beneficial to implement some sort of translation process between the data used by clinicians in their daily care work and the data presented to patients. But then again, who would carry out this work and at what cost? Currently, the work is

[7] Fieldwork 13.04.18.

[8] p. 345.

somewhat delegated to patients in the form of data work, which they engage in, in order to make sense of their data. As described, patients try to interpret their health data by looking up test results or unfamiliar words on Google, or they ask relatives to explain their results. The work that patients carry out in an effort to understand the data available can be viewed as cultivating diagnostic skills, as Andersen [16][9] describes:

> "The possibility to remotely question and achieve increased information (…) prescribes patient work of improving diagnostic skills and developing ways to deal with information that is not easily understandable".

The patients' possibility to remotely access their individual health records activates them in performing data work, particularly in terms of interpreting data that is not easy to understand.

If the patients learn to understand the meaning of their test results, is it then desirable to act on the results on their own without consulting their physician? The specificities of what constitutes an active patient remain unclear, but in the following section, we unfold local enactments of *being* an active patient.

5.3 Being Active Does not Necessarily Involve ICT

All patients in our study expressed the importance of being active in their own patient trajectory. As such, one could say that the sociotechnical imaginary of active patients, which has been established by discourses in policy-documents, has transgressed onto citizens. For instance, Miriam says she thinks it is a good idea that people take care of themselves and take responsibility for their treatment to the degree they are able. However, she also stresses that it is important for clinicians to accept and help the ones not capable (Interview with Miriam). Since imaginaries shape social lives and social orders and produce visions of the collective good [5], it is unsurprising that all five patients perceive themselves as active patients. Lisbeth, Niels, and Emma emphasize that one part of being an active patient in one's own trajectory involves asking questions and asking for further explanation from their physicians:

> Lisbeth: "I would say that I am active in my own disease. I ask – until they are on the verge of losing their minds. And I keep asking. And eventually, if I cannot allow myself to ask any more questions, then I wait till next time, and then I'll ask again".

The two other patients, Astrid and Miriam, relate being active to seeking information on their own. Both of them had not received a final diagnosis during the fieldwork, whereby they sought information related to their symptoms and how they could feel better. Astrid, for instance, used Google to look up dietary suggestions, and she asked relatives, with similar issues as hers, for advice.

The above examples show that the patients are active in regard to asking questions and seeking information. However, these local imaginaries and enactments of *active patients* make no linkage with data access and the use of ICT, which are central to the national imaginary of active patients performing different acts of self-service. Instead,

[9] p. 154.

the patients' conceptualization of being *active patients* relates to taking some degree of responsibility in their own trajectory.

Arguably, from what these patients tell, they are active and take responsibility for the management of their illness. Several of them also describe how they engage in discussions with their physicians about their treatment. In our opinion, this constitutes an active partnership. However, it seems to be a partnership that also demands more from the physicians, since the patients ask critical and clarifying questions to the instructions. In addition, as we have seen, the information available at MyChart is not easily understandable, and for this reason, patients can have a need to consult their physicians about the information they have retrieved online. In this way, digital health technologies might change the roles and responsibilities of both patients and clinicians [11, 30].

6 Concluding Discussion

In line with previously mentioned studies on digital health solutions, our research shows how patients become activated [13, 15, 16] when interacting with MyChart. That is, they become activated to perform data work [7], specifically in the interpretation of information in MyChart. We believe that when the data available in MyChart is not easily understandable for patients, it entails some risk. For instance, patients might misinterpret the data, and additionally, the lack of comprehension can provoke worries among some patients. The organization Danish Patients is aware of this risk, but state that even so, the benefits of knowing, counterbalance cases of worry. However, Danish Patients [31] acknowledges the importance of patients having easy access to a physician when questions are present. Thus, rather than leading to cost reductions and self-service, data access may, in some cases, bring patients to reach out to their hospital for dialogues about the data provided. In this way, MyChart constitutes new forms of work for both patients and clinicians [18]. While we acknowledge the record's role as a work tool for the clinicians [11, 29], our study suggests that a modulation of the documentation practices, similar to that of the Norwegian study [12] might be beneficial. Accommodating the needs of both clinicians and patients would be a delicate balancing act [11].

While MyChart and the access to own health data are optional for patients, the system constitutes a new care infrastructure that fuels the sociotechnical imaginary of a future where patients are part of consolidating this infrastructure by way of being active and engaging in data work. In this way, the Danish government's increased focus on digital health solutions and the strategies promoting *active patients* might create "normativities" [8] about what a good patient is – i.e., one who reads and are informed by health data and who performs actions of self-service [2]. As such, the option of accessing health data might create a sense of obligation to be knowledgeable about the information provided. This is similar to the patients in Felt's [23] study who feel obligated to perform internet searches. In addition, others have shown how digitalization initiatives can create "patterns of exclusion" of some citizens [32]. With this we could ask: who might be forgotten in the digital healthcare initiatives and who have access to and can maneuver ICTs? Since patients are the implicated actors in strategies concerned with *active patients*, it is important to explore and voice their local imaginaries and enactments of being active, in particular with respect to the new digital solutions, such as MyChart, which are imagined as enablers of active engagement.

We are not opposed to data access, new digital solutions, and the benefits they can bring. Rather, we invite to careful considerations of normativities that are embedded in the technologies and policy documents as well as considerations of new patient identities and roles for clinicians, which come about with normativties embedded in such. We especially find it imperative to unfold whether access to health data caters to patients in a way where they feel obligated to access it *and* to learn to understand it. More research is necessary in order to further understand how and if patients wish to be involved in their treatment *and* what roles ICTs, data, and patients' maneuverings with such come to play in the very formation of digital care infrastructures. As it is now, while the patients are active, the data is silent.

Acknowledgement. We are thankful to the participants of this study: the patients, physicians, secretaries, and nurses who have opened their homes and workplaces and wanted to share and explain about encounters with this new data infrastructure. We also wish to extend thanks to the two anonymous HCC reviewers.

References

1. Sundheds- og Ældreministeriet: Sundhed i fremtiden juni 2018: Ansvarlig brug af data til gavn for patienten (2018). http://sum.dk/Aktuelt/Publikationer/~/media/Filer%20-%20Publikationer_i_pdf/2018/Sundhed-i-fremtiden-juni-2018/Sundhed-i-fremtiden-juni-2018-2.PDF
2. Sundheds- og Ældreministeriet, Finansministeriet, Danske Regioner, KL. Ét sikkert og sammenhængende sundhedsnetværk for alle - Strategi for digital sundhed 2018-2022 (2018). http://www.regioner.dk/media/6588/digital-sundhed-publikation.pdf
3. Holmström, I., Röing, M.: The relation between patient-centeredness and patient empowerment: a discussion on concepts. Patient Educ. Counseling **79**, 167–172 (2010). https://doi.org/10.1016/j.pec.2009.08.008
4. Lupton, D.: The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc. Theory Health **11**, 256–270 (2013). https://doi.org/10.1057/sth.2013.10
5. Jasanoff, S.: Future imperfect: science, technology, and the imaginations of modernity. In: Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power, pp. 1–33. University of Chicago Press (2015)
6. Bossen, C., Pine, K.H., Cabitza, F.: Data work in healthcare: an introduction. Health Inform. J. **25**, 465–474 (2019). https://doi.org/10.1177/1460458219864730
7. Langstrup, H.: Patient-reported data and the politics of meaningful data work. Health Inform. J. (2018). 1460458218820188. https://doi.org/10.1177/1460458218820188
8. Singleton, V.: Training and resuscitating healthy citizens in the english new public health - normatives in process. In: Asdal, K., Brenna, B., Moser, I. (eds.) Technoscience: the politics of interventions, pp. 221–246. Oslo Academic Press, Unipub Norway, Oslo (2007)
9. Wass, S., Vimarlund, V., Ros, A.: Exploring patients' perceptions of accessing electronic health records: innovation in healthcare. Health Inform. J. **25**, 203–215 (2019). https://doi.org/10.1177/1460458217704258

10. Kensing, F., Lomborg, S., Moring, C.: Evolving relations between the practices of nurses and patients and a new patient portal. European Society for Socially Embedded Technologies (EUSSET) (2017)
11. Vikkelsø, S.: Mobilizing information infrastructure, shaping patient-centred care. Int. J. Public Sector Manag. **23**, 340–352 (2010). https://doi.org/10.1108/09513551011047233
12. Berglind, F.S.: Patient Accessible Electronic Health Records: Impacts on Nursing Documentation Practices at a University Hospital. Studies in Health Technology and Informatics, pp. 14–18 (2018). https://doi.org/10.3233/978-1-61499-872-3-14
13. Winthereik, B.R., Langstrup, H.: When patients care (too much) for information. In: Care in Practice: On Tinkering in Clinics, Homes and Farms, pp. 195–214 (2010)
14. Svenningsen, S.: Den elektroniske patientjournal og medicinsk arbejde: reorganisering af roller, ansvar og risici på sygehuse. Handelshøjskolens, København (2004)
15. Oudshoorn, N.: Diagnosis at a distance: the invisible work of patients and healthcare professionals in cardiac telemonitoring technology. Sociol. Health Illness **30**, 272–288 (2008). https://doi.org/10.1111/j.1467-9566.2007.01032.x
16. Andersen, T.: The participatory patient. In: Proceedings of the 11th Biennial Participatory Design Conference. ACM Press, Sydney, Australia, pp. 151–154 (2010)
17. Finken, S., Mörtberg, C.: Performing elderliness – intra-actions with digital domestic care technologies. In: Kimppa, K., Whitehouse, D., Kuusela, T., Phahlamohlaka, J. (eds.) HCC 2014. IAICT, vol. 431, pp. 307–319. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44208-1_25
18. Laursen, C.S.: My health platform and changing relations of care - an ethnographic study of a new Danish patient portal. Master's thesis, IT University of Copenhagen (2018)
19. Clarke, A.E.: Situational Analysis: Grounded Theory After the Postmodern Turn. Sage Publications, Thousand Oaks (2005)
20. Felt, U., Fouché, R., Miller, C.A., Smith-Doerr, L.: Introduction to the fourth edition of the handbook of science and technology studies. In: Felt, U. (ed.) The Handbook of Science and Technology Studies, 4th edn, pp. 1–26. The MIT Press, Cambridge, Massachusetts (2017)
21. Marcus, G.: Introduction. In: Marcus, G. (ed.) Technoscientific Imaginaries: Conversations, Profiles and Memoirs, pp. 1–9. University of Chicago Press, Chicago (1995)
22. McNeil, M., Arribas-Ayllon, M., Haran, J.: Conceptualizing imaginaries of science, technology, and society. In: Felt, U. (ed.) The handbook of science and technology studies, 4th edn, pp. 435–464. The MIT Press, Cambridge, Massachusetts (2017)
23. Felt, U.: Sociotechnical imaginaries of "the internet", digital health information and the making of citizen-patients. In: Hilgartner, S., Miller, C., Hagendijk, R. (eds.) Science and democracy: making knowledge and making power in the biosciences and beyond. Routledge, London (2015)
24. Steno Diabetes Center Copenhagen.: Nyt IT-system – Sundhedsplatformen (2019). https://www.sdcc.dk/undersoegelse-og-behandling/sundhedsplatformen/Sider/Nyt-IT-system.aspx. Accessed 11 Aug 2019
25. Region H Sundhedsplatformen - én samlet patientjournal. https://www.regionh.dk/sundhedsplatform/om-sundhedsplatformen/Sider/en_samlet_patientjournal.aspx. Accessed 19 Mar 2018
26. Crang, M., Cook, I.: Doing ethnographies, Online version. Geobooks. Durham Research Online, Durham University. United Kingdom, Norwich (1995)
27. Forsythe, D.E.: It's just a matter of common sense: ethnography as invisible work. Comput. Supported Cooperative Work (CSCW) **8**, 127–145 (1999)
28. Haraway, D.: Situated knowledges: the science question in feminism and the privilege of partial perspective. Feminist Stud. **14**, 575 (1988). https://doi.org/10.2307/3178066

29. Berg, M.: Lessons from a dinosaur: mediating is research through an analysis of the medical record. In: Baskerville, R., Stage, J., DeGross, J.I. (eds.) Organizational and Social Perspectives on Information Technology: IFIP TC8 WG8.2 International Working Conference on the Social and Organizational Perspective on Research and Practice in Information Technology 9–11 June 2000, Aalborg, Denmark. Springer US, Boston, MA, 487–504 (2000)
30. Hult, H.V., Hansson, A., Svensson, L, Gellerstedt, M.: Flipped healthcare for better or worse. Health Inform. J. **25**, 587–597 (2019). https://doi.org/10.1177/1460458219833099
31. Danske patienter Journaladgang. In: Danske Patienter. https://danskepatienter.dk/politik/det-mener-vi/journaladgang. Accessed 7 Apr 2019
32. Schou, J., Pors, A.S.: Digital by default? A qualitative study of exclusion in digitalised welfare. Soc. Policy Adm. **53**, 464–477 (2019). https://doi.org/10.1111/spol.12470

Author Index

Baccelli, Michele 64, 77
Bednar, Kathrin 281
Bergroth, Harley 351
Burmeister, Oliver K. 158

Cai, Wenjie 305
Chakrabarti, Sujit Kumar 181
Corneliussen, Hilde G. 379

Davis, Matthew J. 291
De Barros, Martina Jennifer Zucule 233
de Coning, Arno 193
Dresser, Greg 158

Finken, Sisse 398
Fors, Per 291

Galanos, Vassilis 141
Greig, Jenni 158
Grenda, Dennis 171

Heimo, Olli I. 340
Hiratsuka, Mitsuyoshi 64, 77

Ishii, Kaori 5

Jansen van Vuuren, Anna-Marie 207, 248
Jansen van Vuuren, Joey 207, 248
Jayagopi, Dinesh Babu 181

Kakuta, Tokuyasu 86
Kallio, Johanna 361
Kantorovitch, Julia 361
Kimppa, Kai K. 340
Kobayashi, Kazuto 64, 77
Koga, Hiroshi 315
Koskinen, Jani 327
Kurihara, Yusuke 130
Kyllönen, Vesa 361

Laursen, Cæcilie Sloth 398
Leenen, Louise 248

Maruhashi, Toru 100
McKenna, Brad 305
Mischau, Lena 16
Mouton, Francois 193

Nagasako, Tomoko 262
Nakamura, Mariko 41
Nakashima, Mika 29

Ozaki, Aimi 52

Ramluckan, Trishana 233
Rantanen, Minna M. 327
Räsänen, Pauli 361
Rehman, Sabih-Ur 158
Reisel, Mary 141
Ronkainen, Jussi 361

Sereboff, Steven C. 64
Shima, Aki 86
Simonsen, Morten 379
Singh, Shivam Kumar 181

Tuikka, Anne-Marie 171

Ul-Haq, Anwaar 158

Van ‘t Wout, Carien 207
van Niekerk, Brett 233
Vildjiounaite, Elena 361
Vuorinen, Jukka 351

Waizenegger, Lena 305
Warnke, Martin 117
Watanabe, Seiko 222