Andrea Kő · Enrico Francesconi ·
Gabriele Kotsis · A Min Tjoa ·
Ismail Khalil (Eds.)

LNCS 13429

Electronic Government and the Information Systems Perspective

11th International Conference, EGOVIS 2022
Vienna, Austria, August 22–24, 2022
Proceedings

Springer

Lecture Notes in Computer Science 13429

More information about this series at https://link.springer.com/bookseries/558

Andrea Kő · Enrico Francesconi ·
Gabriele Kotsis · A Min Tjoa ·
Ismail Khalil (Eds.)

Electronic Government and the Information Systems Perspective

11th International Conference, EGOVIS 2022 Vienna, Austria, August 22–24, 2022 Proceedings

Springer

Editors
Andrea Kő
Corvinus University of Budapest
Budapest, Hungary

Enrico Francesconi
National Research Council
Florence, Italy

Gabriele Kotsis
Johannes Kepler Univversität Linz
Linz, Oberösterreich, Austria

A Min Tjoa
Vienna University of Technology
Vienna, Austria

Ismail Khalil
Johannes Kepler University of Linz
Linz, Austria

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-031-12672-7 ISBN 978-3-031-12673-4 (eBook)
https://doi.org/10.1007/978-3-031-12673-4

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

Preface

The 11th International Conference on Electronic Government and the Information Systems Perspective, EGOVIS 2022, took place in Vienna, Austria, during August 22–24. The conference was collocated with the 33rd International Conference on Database and Expert Systems Applications (DEXA 2022) and associated events.

The international conference cycle EGOVIS focuses on information systems and ICT aspects of e-government. Information systems are a core enabler for e-government/governance in all dimensions: e-administration, e-democracy, e-participation, and e-voting. EGOVIS 2022 brought together experts from academia, public administration, and industry to discuss e-government and e-democracy from different perspectives and disciplines, i.e. technology, policy and/or governance, and public administration.

The Program Committee accepted 11 papers from recent research fields such as artificial intelligence, machine learning, smart solutions, and semantic technologies for services in the public sector. Beyond theoretical contributions, the papers cover e-government practical applications and experiences.

This proceedings is organized into three sections according to the conference sessions: e-Government theoretical background, semantic technologies and legal issues, and artificial intelligence and machine learning in e-Government context.

The chairs of the Program Committee wish to thank all the reviewers for their valuable work; the reviews raised several research questions to discuss at the conference. We would like to thank Ismail Khalil for the administrative support and for stimulating us in proper scheduling.

We hope for pleasant and beneficial learning experiences for the readers. We also hope that the discussion between researchers will continue after the conference and contribute to building a global community in the field of e-government.

August 2022

Enrico Francesconi
Andrea Kő

Organization

Program Committee Chairs

Andrea Kő	Corvinus University of Budapest, Hungary
Enrico Francesconi	National Research Council of Italy

Steering Committee

Gabriele Kotsis	Johannes Kepler University Linz, Austria
A Min Tjoa	TU Wien, Austria
Robert Wille	Software Competence Center Hagenberg, Austria
Bernhard Moser	Software Competence Center Hagenberg, Austria
Alfred Taudes	Vienna University of Economics and Business and Austrian Blockchain Center, Austria
Ismail Khalil	Johannes Kepler University Linz, Austria

Program Committee

Luis Alvarez Sabucedo	Universidade de Vigo, Spain
Alejandra Cechich	Universidad Nacional del Comahue, Argentina
Wojciech Cellary	WSB University in Poznan, Poland
Wichian Chutimaskul	King Mongkut's University of Technology Thonburi, Thailand
Peter Cruickshank	Edinburgh Napier University, UK
Vytautas Čyras	Vilnius University, Lithuania
Ivan Futo	Multilogic Ltd, Hungary
Fernando Galindo	University of Zaragoza, Spain
Francisco Javier García Marco	Universidad de Zaragoza, Spain
Henning Sten Hansen	Aalborg University, Denmark
Christos Kalloniatis	University of the Aegean, Greece
Nikos Karacapilidis	University of Patras. Greece
Evangelia Kavakli	University of the Aegean, Greece
Herbert Leitold	Secure Information Technology Center – Austria (A-SIT), Austria
Javier Nogueras Iso	Universidad de Zaragoza, Spain
Aljosa Pasic	Atos Origin, Spain
Aires Rover	Universidade Federal de Santa Catarina, Brazil
A Min Tjoa	TU Wien, Austria

Costas Vassilakis	University of the Peloponnese, Greece
Christopher Wills	Caris Research Ltd, UK
Chien-Chih Yu	National Chengchi University, Taiwan

Organizers

Contents

e-Government Theoretical Background

Notions of Fairness in Automated Decision Making: An Interdisciplinary Approach to Open Issues

Yasaman Yousefi(✉)

Department of Legal Studies, University of Bologna, Bologna, Italy
yasaman.yousefi3@unibo.it

Abstract. Artificial Intelligence (AI) systems share complex characteristics including opacity, that often do not allow for transparent reasoning behind a given decision. As the use of Machine Leaning (ML) systems is exponentially increasing in decision-making contexts, not being able to understand why and how decisions were made, raises concerns regarding possible discriminatory outcomes that are not in line with the shared fundamental values. However, mitigating (human) discrimination through the application of the concept of fairness in ML systems leaves room for further studies in the field. This work gives an overview of the problem of discrimination in Automated Decision-Making (ADM) and assesses the existing literature for possible legal and technical solutions to defining fairness in ML systems.

Keywords: Automated Decision-Making · Machine Learning · Fairness · AI ethics

1 Introduction

Every comma of contemporary society is influenced by modern technologies and automated decisions through AI systems[1]. AI systems shape people's lives in important instances through automated decisions. ADM refers to instances where "a conclusion is reached without any human involvement". This automation can come from the application of a specific and limited algorithm, or from "algorithms using complex AI supported by ML [1]". Automated Decision Systems (ADS) are increasingly used in legal fields, especially in criminal justice [53], and public administration [51]; loan, mortgage, and social benefits are allocated using ADM [37]; and there is an exponential growth in using ML and ADM in the healthcare sector [35]. Relying on technologies in decision-making contexts can result in more efficient outcomes, and 'potentially fairer' decisions

[1] AI systems are those that manifest intelligent behaviour and take actions with some degree of autonomy to achieve certain goals [25]. Machine Learning techniques are used to train AI systems to take data as input and use algorithms to output predictions [1].

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 3–17, 2022.
https://doi.org/10.1007/978-3-031-12673-4_1

than human decisions which might be subject to subconscious prejudices and biases [29]. The reliance on new technologies in important sectors of decision-making effects us individually and collectively, leading to a series of debates on ethical AI and its surrounding principles, fairness being one of them[2].

Automated decisions are made using AI techniques that rely on historical data. Unlike some cases of human bias, ADM usually discriminates unintentionally. In the case of biased input data, a decision might lead to the perpetuation of existing patterns of unfairness and discrimination in the society [5], which often brings along harmful outcomes for minority groups [29]. Further, with incomplete or unreliable data, an ADM might under-represent or over-represent certain groups [28]. It is also critical to put the right algorithm in use, as ML models can perform differently for each task [45].

With automated decisions, we face more abstract and subtle ways of discrimination in comparison with the human discrimination that is usually based on the existing stereotypes and prejudices, or discriminatory norms in societies. Thus, automated discrimination is more difficult to perceive, and the established legal remedies to detect and fix (human) discrimination are not enough as they can be disrupted and outdated [23]. Computational techniques are addressing aspects of these concerns through discrimination-aware data mining (DADM) and fairness, accountability and transparency machine learning (FATML) [45]. Reducing such biases is not only a technical challenge, but a legal and social wake-up call.

The main question of this paper relates to the perspectives we can adopt in mitigating open discriminatory issues related to ADM. Beginning by studying these open issues, the first section will give an overview of the problem, putting together concepts of discrimination and fairness. In the second chapter, the study will investigate open issues in applying fairness to ADM, where several notions of fairness will be analyzed to assess which can be the best.

Further, the human impact on unfairness in ADM, as well as the social and legal problems arising from unfair ADM will be studied in the third chapter. The question of unfairness is a valid concern also when we use symbolic AI solutions, where the rules are created through human intervention instead of relying on correlations between inputs and outputs [26]. Fairness and explainability solutions have been proposed through symbolic AI solutions that will be briefly studied in the last section. However, the main focus of this paper remains on unfair ML decision-making. Eventually, the possible legal and technical solutions to fairness in ADS will be assessed. This work is a literature review that contributes to the intersection of technical and legal solutions to fair automated decisions.

[2] It should be noted that other principles such as the principle of transparency and explainability, reliability and safety, accountability, and other novel principles such as explicability [18] and knowability [33], are equally important fields of study. However, the focus of this article is mainly on the notions of the principle of fairness in ML systems, and due to the limited scope of the paper, there is not enough room to discuss these principles individually.

2 Fairness and Discrimination

Possibly the most common way of conceptualizing fairness in the realm of ADM is relating this concept with that of non-discrimination and differential treatment [24]. The legally established principles of non-discrimination are based on traditionally defined criteria such as ethnicity, race, language, gender, religion, or political views. Algorithmic discriminatory outputs are based on less clear categories and previously unpredicted attributes. Therefore, automated discrimination might fall outside the personal and material scope of the relevant laws except for social groups that have traditionally been targets of marginalization. ADM can create new grounds for discrimination that are not protected by the law, but are ethically unjust. They do not directly violate any norm, but they are an obstacle to the fulfillment of equality goals set in legislation. Any biased practices arising from new categories of discrimination might slip through the loopholes of existing non-discriminatory laws [41].

With the use of modern technologies in decision-making contexts, establishing prima facie discrimination presents a challenge. Automated discrimination is more abstract and unintuitive, subtle, and intangible, unlike traditional forms of discrimination [30]. These characteristics make the detection and proving of discrimination difficult due to the lack of a 'comparative element'. People with similar educational backgrounds might be offered different job opportunities on the basis of a characteristic that is not protected by equality laws. Traditionally speaking, one can compare and realize a sort of disadvantage in this scenario. In the automated world, if a certain job advertisement is not shown to a given group of people that have a certain characteristic, they would not even know they are being disadvantaged and even if they realize it, proving that discrimination has happened would be difficult. Furthermore, victims will not be able to prove their case without access to the data and the algorithms, which is usually impossible [23].

EU non-discrimination laws generally prohibit direct and indirect discrimination in Article 21 of the EU Charter of Fundamental Rights, and Article 14 of the European Convention on Human Rights (ECHR). Direct discrimination occurs when an individual is treated differently based on their membership in a protected social group; discrimination based on gender or sexual orientation is an example of direct discrimination. On the other hand, indirect discrimination happens when provisions or practices that appear to be neutral, have a disproportionate and disadvantaging effect on a protected group. For instance, if a job's policy requires that staff must work on Saturdays, it can in a way disadvantage Jewish staff who observe the Sabbath. Moreover, a third type of discrimination or multi-dimensional discrimination can also be mentioned when discrimination is based on more than one characteristic protected by non-discrimination law. An example of multi-dimensional discrimination is a Muslim black woman who has been treated differently for her belonging to 'Muslim', 'black' and 'woman' groups [47].

There are not enough relevant case laws that address indirect discrimination. However, by detecting indirect discrimination, there is hope for reducing

systematic and structural inequalities that are hidden in the normal practices of societies. To prove the occurrence of direct or indirect discrimination, it should be demonstrated that firstly, certain harm has happened or is likely to happen; secondly, the harm is significantly manifest within a protected group of people; and finally, that it is disproportionate. Wachter et al. argue that when it comes to multi-dimensional discrimination, there is a lack of clear guidance and jurisprudence, seen also in judgments of the European Court of Justice. Therefore, there is room for addressing discrimination based on multiple reasons in the development of automated systems [47].

Making case-specific normative choices that reflect the political, legal, and social dimensions of every case, is a necessary element for courts when dealing with discrimination. This explains that fairness is deeply contextual. In fact, 'substantive fairness' safeguards such as the right to non-discrimination and other fundamental rights, as well as guaranteeing 'procedural fairness', for example, due process and transparency, should be taken into account by decision-makers, such as judges and politicians, and are a part of their training [24]. Watcher et al. suggest that automated systems must be able to replicate the judiciary attitude to "contextual equality" in case of their application in legal and ethical decision-making, although this is a difficult task [47].

The ability to object to an automated decision (e.g. in Art. 22 GDPR) is restricted, and provoked only when a process has been fully automated, or in the case of decisions with a legal effect and subject to limitations. This limited scope raises the question of "whether there should be additional grounds that entitle people to object to automated decisions, for example, arguments related to dignity, a lack of trust in such systems rendering a balanced decision, or moral objections to the very idea of being subject to ADM [24]"

Even in the most recent proposal of the European Parliament and The Council for harmonizing rules on AI, The Artificial Intelligence Act (AIA), the approach to discrimination and equality is rather general. The lack of association of the concept of fairness in law with discriminatory algorithms to protect vulnerable groups leaves a great room for further discussion and research [12]. The proposal divides AI systems following the potential risks through a pyramidal approach from high to low-risk AI systems, considering fundamental issues such as bias and discrimination. This highly appreciated risk classification fails to give adequate practical solutions to the self-assessment of risks and to ensure the safety of high-risk AI systems that are developed outside the EU. Therefore, it has been suggested that the AIA takes a more flexible and adaptable approach and leaves room for modifications in its risk classification and the areas where risk can be expected [12]. It seems that in practice there are no rights given to individuals that might be subject to unfair decision-making by AI systems. Further, in Article 10 the AIA requires the use of data governance practices that shall concern in specific "examination in view of possible biases". However, the notion of bias is not clarified. In recital 44, the AIA has stated that high-risk AI systems should not become "the source of discrimination prohibited by Union law". Although, the forms of biases prohibited under the existing framework are

not explicitly indicated and there is no mention of how to mitigate algorithmic bias. The AIA does not define the criteria for measuring the quality of data sets, for instance, the fairness of trained ML models [16].

Automated decision-makers may eventually replace or augment human decision-makers. Therefore, the lack of clear models or metrics to determine fairness has become a matter of interest to researchers. As noted before, the protection in the ECHR is rather general in its range of application. European laws provide a wide range of protection concerning discrimination on grounds of race, ethnicity, and gender, especially in terms of employment, occupation, and social advantages. However, this protection is less extensive when it comes to religion, disability, and sexual orientation. In automated discrimination, the challenge of facing new grounds of disadvantage, calls for a clear definition of the concept of fairness. The interesting question is, what notion is the best-if there can even be one-for integrating the law with modern technologies, taking ethics into account.

3 The Fairest Fairness Notion

Helberger et al. mention that the perceptions of fairness may be different between various contexts, cultures, jurisdictions, and even amongst individuals. The principle of fairness is not predefined and does not entail concrete and harmonious definition, as decisions related to this concept are usually judgment calls depending on different factors of each given case [24]. Angelopoulos defines this as 'rational discourse', which relies on the balancing of interests or values [3]. What the EU offers that was mentioned in the previous section, is an established general framework of laws and standards in primary and secondary law for non-discrimination and protection and enhancement of equality. As said before, automated discrimination infringes the existing equality framework in certain circumstances [2], and the general legislative framework does not fully cover the new grounds of discrimination, that might be unknown to us so far.

While the terminology meaning of fairness overlaps with that of the legal concept of discrimination, the notion of fairness is grounded on ethics and contains meanings outside the scope of problems risen in relation to the law. Floridi argues that bioethics resembles digital ethics the most out of all areas of applied ethics in dealing with new forms of agents, patients, and environments [17]. The principle of 'Justice' is one of the four main principles of bioethics by Beauchamp and Childressis [6], which like the other three, adapts well to the new ethical challenges rising in automated decisions [19]. The principle of justice is usually defined as fairness or fair distribution of goods in society, and "matches our common sense of fairness" as Beauchamp and Childressis put it [6]. In the theory of justice that highly inspires the principle of justice in bioethics, Rawls argues that "what justifies a conception of justice is not its being true to an order antecedent and given to us, but its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable

doctrine for us [36]". Binns relates political philosophy to algorithmic fairness. He explains that theoretical notions of algorithmic fairness are influenced by the perception of egalitarianism by people. Algorithmic fairness depends on not only grounds of unequal distribution, but also on how discrimination is traditionally understood [9].

Giovanola and Tiribelli point that there is more to the concept of fairness than non-discrimination and an exclusively distributive dimension. They state that this principle includes features that should be expanded beyond the consideration of biases, including "the consideration of both a distributive and a socio-relational dimension". Their approach to fairness is mostly ethical, grounded on respect for persons both as persons and as particular individuals. They redefine this principle considering three main components that should be added to it: "fair equality of opportunity, equal right to justification, and fair equality of relationship [21,22]".

The Montreal Declaration mentions the importance of the principle of Justice in AI by stating that "the development of AI should promote justice and seek to eliminate all types of discrimination [44]". The European Group on Ethics in Science and New Technologies(EGE), takes a similar approach to justice in its principle, namely 'Justice, equity and solidarity', arguing that AI should "contribute to global justice and equal access to the benefits" of AI technologies, advising also against the risk of bias in data sets used to train AI systems. In its statement, it is the pioneer of the idea of defending against threats to 'solidarity', including "systems of mutual assistance such as in social insurance and healthcare [15,19]". As an ethical principle, justice has been interpreted similarly, with subtle distinctions in the context of AI. Floridi et al. summarize that justice in AI combines "(a) Using AI to correct past wrongs such as eliminating unfair discrimination; (b) Ensuring that the use of AI creates benefits that are shared (or at least shareable); and (c) Preventing the creation of new harms, such as the undermining of existing social structures [19]".

Shin and Park note that the notion of algorithmic fairness can be a complicated one, due to the contextual and subjective definition of fairness. Generally speaking, algorithmic fairness refers to decisions made by algorithms that should not be unjust and discriminatory [40]. Algorithmic bias, under the umbrella of fairness, is an unwanted error "that places privileged groups at a systematic advantage and unprivileged groups at a systematic disadvantage [8]". From a legal point of view, this concept is related to the unfavourable treatment or disadvantage experienced by protected categories of the population by the law either explicitly or implicitly. In computer science, the notion of fairness is based on "various types of fairness delineated by ethicists and philosophers, for example, group, individual, procedural, outcome-based, counter-factual fairness, etc. [20]". We can approach algorithmic fairness through mathematical definitions [46,54], or apply philosophical and ethical concepts of human fairness to algorithms [9,43].

The mathematical definition of fairness is grounded on studying discrimination, also called statistical- or group-level fairness for binary classification.

"Statistical notions require a particular metric -quantifying benefit or harm- to be equal across different social groups (e.g., gender or racial groups) [42]". The criteria of fairness can differ due to different choices for the benefit metric. A widely accepted mathematical definition of fairness is demographic parity, which seeks to even up the percentage of people who are predicted to be positive across different groups, basically with zero difference between the groups getting a positive outcome [11][3]. This rarely happens in reality and highlights the necessity to track down fairness through a more interdisciplinary approach[4].

It is also argued that the literature on fair ML algorithms mainly derives its fairness concept from the legal context of data protection laws. Fairness as mentioned in Article 5 (1) of the General Data Protection Regulation (GDPR) refers to the expectations and effects of processing on the part of data subjects. This principle of purpose limitation focuses on the collection of data only for specified, explicit, and legitimate purposes and does not allow further possessing of data in a way that is not compatible with those purposes, with some exceptions [13]. The focus of computer science on framing substantive fairness in terms of non-discrimination reflects the legal conceptualization of fairness. Therefore, fairness is both a fundamental right (Art. 14 ECHR) and a guaranteed right by various non-discrimination laws prohibiting different treatment of people or groups based on characteristics including race, gender, nationality, sexual preference, political or religious convictions, etc. [24].

As studied in this section, there are different notions of fairness that can turn out to be incompatible with one another. For that, scholars have emphasised the importance of the social context of fairness when assessing algorithmic fairness [29]. Fairness is not a static value that we can fully code in an application; it is the result of a balancing process [24] considering different notions and different factors in every given case. Evidently, people are directly affected by ADM and have to deal with the consequences of discriminatory decisions [24]. It is argued that understanding individuals' perception of fairness is necessary both from an ethical point of view, and to promote trust in the adoption of ADS for society [39].

Formalizing fairness in ADM and ensuring the necessary checks and balances is an important challenge [34], but it is not the main challenge. Helberger et al. state that "designing and implementing ADM systems in a way that humans feel that their dignity is respected and that they are treated appropriately as humans, not numbers" is another equally important challenge [24]. Therefore, approaching ADM from a legal and societal perspective, the focus should not

[3] The aforementioned benefit or harm can be quantified through demographic parity if we consider the classification within a certain group as a benefit or harm; for instance, for university admissions, the admitted group entails a benefit.

[4] Some other examples of the mathematical definition of fairness include: Error Parity [10], equality of False Positive or False Negative rates [52], and False Discovery or Omission rates [27]. For a concentrated study on mathematical notions of fairness refer to [42].

be on fairness in ADM alone, "but the way it is implemented and affects inter-human relationships [24]".

Moreover, there is not only one possible fair outcome in any given decision-making. In some cases, there can be a number of possibilities and different arguments which we should be able to weigh [24,32].

In the study conducted by Helberger et al. people's perspective on AI decision-making has proved to be fairer than humans, accompanied by the belief in the potential and objectivity of ADM. For some people, human decision-making is unreliable, for some, it is necessary that humans are involved in the ADM process. This promising point of view suggests that "algorithmic versus human fairness is not an 'either/or question' ". It is important to see where human fairness meets machine fairness, and cooperate to make fairer decisions. The strength lies in the combination of these notions [24]. The mutual dependency of data and humans calls for this mutual cooperation.

4 Humans Discriminate, Machines Learn!

New technologies can suffer from the same biases that older technologies or their human designers express. Regarding the real impact of automated decisions, one can imagine algorithmic predictions used to guide humans in deciding on immigration status [1], detection of possible welfare fraud as a preventive action by governments [7], and recidivism predictions [48]. Eventually, it is vital to consider the human impact on these decisions in every stage of the development of ADM.

At first glance, it might seem odd to equate a concept such as fairness with AI, as morality presumes the existence of a human being. The modern society displays hope for moral consciousness through Artificial Moral Agents in the future; however, currently, we deal with systems that display autonomy only in a technological and not in a moral sense [50]. This takes us back to challenging the human morality and sense of fairness, on which we can partly blame the unfair automated decisions. Nonetheless, despite the lack of independent moral consciousness, AI systems impact human actors and are typically directed at the attainment of goals with human implications and outcomes. Therefore, when evaluating AI for fairness in an ethical sense, we should consider the fairness of the equivalent human action. This is especially applicable for AI systems that undertake socially important tasks by replacing human judgments, such as in the fields of justice, education, and employment.

Epistemic or ethical concerns that accompany ADM include inconclusive, inscrutable, and misguided evidence, while normative concerns regard unfair outcomes, transformative effects and traceability [31]. A core problem of AI systems is the 'black box' conundrum, because of our inability to explain how algorithms transform their inputs into decisions. Apart from an epistemic and normative concern, this inexplicability of AI raises legal concerns, as it makes it impossible to explain the causal connection and underlying reasoning of algorithmic discrimination. When it comes to human discrimination, there is no such

thing as a 'black box', which makes it possible to raise a legal case [41]. However, the human brain is also obscure and vague when it comes to the decision-making process, in a similar way to that of the black box effect in ML systems. Humans are subject to prejudice and biases, which can be difficult to detect and correct. In this regard, ADM offers prospects to visualize, identify and correct discriminatory biases if trained well to do so and under the right regulatory framework [20].

Human decision-making can be distinguished from that of a machine when we consider that minor shifts can occur in the legal development of a human decision in accordance with the background and moral perspectives of each case. But when algorithms are applied in legal processes, they are not themselves open to this implementation of law [38]. When it comes to proving discrimination in court, the law reveals the necessity of making case-specific normative choices reflecting the political, legal, and social dimensions of every given case. Although it is not an easy task, this judiciary attitude to "contextual equality" is what automated systems must be able to replicate in case of their application in legal and ethical decision-making [47]. Lord Sales mentions the danger of freezing particular relationships in set configurations with set distributions of power that are created by coding algorithms. 'The wider perceptual control which is noticeable as our world becomes increasingly digital also tends to freeze categories of thought along tram-rails written in code'. He believes if not resisted, this can limit imagination and inspiration even for legislative responses to digitisation. This erodes human capacities to change or question power relations. Algorithms can enshrine the unspoken biases of human coders beyond scrutiny and challenge [38]. However, in my humble opinion, it is not as simple as that. Algorithms can indeed have a good side if implied in decision-making by mitigating bias and tackling human discrimination. Even if we are not there yet, I believe relying on ADM poses less risks than relying on human judgment in some cases.

Furthermore, algorithms as closed systems cannot capture every potential aspect of resolving human problems that are often nuanced and ambiguous by nature. With the evolution of human law, the open nature of ideas such as justice and fairness provides the possibility for internal criticism of applicable rules. This leaves room for broader values that are not clearly encapsulated in legal algorithms, in order to enter the equation that leads to a final outcome. Lord Sales concludes that at some point in the future, AI systems will be developed to a stage beyond algorithmic systems. Machines will show sensitivity to the application of rules (both arbitrary and contingent) that allow them to navigate issues of equity arising from relevant ethics, human rights and, constitutional considerations. Application of fairness and equity rules or recognition of all circumstances of hard cases where ethical and legal considerations meet, is dependent on pattern recognition. AI is likely to be able to navigate these issues and meet the standards of human judgment. But Lord Sales dryly notes that 'we are not there yet' [38].

5 The Quest for Solutions to Automated Fairness

The first step in the adoption of fairness in ADM is understanding the various causes of discrimination and finding solutions to prevent biases from the design process to the decision-making. Several technical solutions to detecting bias and evaluating ML fairness have been designed that can up to a certain degree be helpful with detecting and mitigating bias from input to output of ML. Examples of such include FairML, IBM AI Fairness 360, and Google What-If Tool[5].

However, these solutions may not suffice. Algorithmic fairness is based on different contextual assumptions and optimizations and that is why a discussion of trade-offs and ethics is needed, where the focus should be on integrating the societal notions of fairness and possible social costs. Research shows that from people's perspective, accuracy can be even more important than equality in a high-risk situation [42]. There are opportunities to improve fairness and accuracy in ADM. From enhancing the overall accuracy by bias detection and looking for possible bugs, to additional training of under-representing data sets to reduce unfair results. This approach to fairness and accuracy can be seen in the study by Buolamwini and Gebru where they measure the accuracy of gender-classification algorithms, focusing on increasing phenotypic and demographic representation in facial recognition data sets [10]. Hacker suggests an integrated vision of anti-discrimination and data protection law to enforce fairness in the digital age [23]. He believes data protection law contains the solution to fairness and states that the GDPR offers a promising path together with the concepts of anti-discrimination laws to combat algorithmic discrimination. "In combination with the instruments of algorithmic fairness, data protection and anti-discrimination law, therefore, provide the tools to implement 'equal treatment by design' [23]". However, it should be considered that non-discrimination law fails to cover new grounds of discrimination brought along with ADM. Further, the GDPR can help uncover the causes of bias and enforce metrics of algorithmic fairness, but fails to offer protection based on the new grounds of discrimination, and counting on it alone, seems too optimistic. The GDPR offers protection through a data-centric approach rather than a human-centric one, except for the protection of minors. Furthermore, as mentioned earlier in Sect. 2, the concept of fairness should entail more than pure non-discrimination and bias elimination [14].

I believe that non of the available technical, legal or ethical solutions alone contain the answer to fairness in ADM. Much more research needs to be done in this direction and in specific for creating a blueprint pathway to ADM fairness. This code of ADM fairness should be applicable to the design process of ML systems, and result in: a) compensating for bias and discrimination in all its forms using technical and legal solutions; b) ensure the individuals' right to justification and explanation; and c) ensure means to hold responsible and accountable either the designers of unfair ML systems, or the human agents involved in an unfair ADM. Fairness in some cases can be translated into equal-

[5] For more see https://towardsdatascience.com/evaluating-machine-learning-models-fairness-and-bias-4ec82512f7c3.

ity, in other cases it entails other principles. Just like the contextual equality approach towards fairness in courts, ML systems should ideally be able to assess and examine a situation and decide what notion of fairness should be applied. This task seems impossible at least now without a human intervention in the process of decision-making. In the endless search for a solution to adapting a notion of fairness in ADM, I noted that algorithmic versus human fairness is not an 'either/or question'. The solution to fairness lies in a mutual cooperation between humans and ML systems, as well as a new combined notion of fairness that covers more than only one strict notion and can be implied circumstantially.

As a solution to fairness in AI systems, Wagner and Garcez postulated symbolic and neural-symbolic solutions to integrate fairness into AI and counter the black box of ML decisions. They believe that combining Explainable Artificial Intelligence (XAI)[6] methods with neural-symbolic approaches leads to the detection of undesired behaviour of a model, and helps reaching explainability and fairness through a model that removes demographic disparities from trained neural networks by using a continual learning method. They demonstrated the effectiveness on addressing undesired gender-based disparities first on simulated data, and then on real-world data in their experiment [49].

The ethical challenges and concerns regarding fairness in automated decisions have been pointed out throughout this paper. It is worth mentioning that automated decision-making can have a good side and might help decrease and mitigate the risks of human bias. Many opportunities come along with algorithmic decision-making that can even help reduce or eliminate discrimination in its direct, indirect and multi-dimensional forms. In terms of obscurity in the decision-making process, human reasoning may be subject to biases, that can be very difficult to detect and correct. In this regard, algorithmic decision-making, if trained correctly, might offer prospects to visualize, identify and correct discriminatory biases under proper regulations. Hacker suggests that algorithmic fairness cannot "address the roots of injustice in society, but it can mitigate some of its outcomes". This is a great advantage of ADM over human decision-making. The "design parameters can be consciously chosen in algorithms [23]".

The technical and legal communities should bring their strengths together to enable "contextual equality" for automated discrimination through practical tools and statistical measures. Designing systems to aid judges in normative decisions by producing statistical evidence, and aid controllers to detect bias and potential discrimination before they happen has been speculated as a better way than correcting discriminatory decision-making through systemic design. This creates an 'early warning system' for automated discrimination with common technical and judiciary standards to assess prima facie discrimination [47].

Non-discrimination has a contextual notion in the EU. Therefore, judicial interpretation plays an important role in decision-making. Defining statistical metrics of fairness and applying them to automated systems is not an adequate approach to encourage the flexibility of equality and fairness in EU law

[6] XAI is AI in which the process of decision-making is understandable by humans unlike the black box effect present in ML techniques [4].

into modern technical tools and measures. Acknowledging this, Wachter et al. propose "Conditional Demographic Disparity" that provides "measurements for making comparisons across protected groups in terms of distribution of outcomes". This measure can help identify disparity and require further investigation and examine both disadvantaged and advantaged groups through the same assessment of the court to detect prima facie discrimination [47]. An interesting topic with legal and technological value can be assessing the mentioned solutions in an exemplary automated decision-making context to choose the best way(s) of introducing fairness to ML systems. Moreover, defining rules of algorithmic procedural fairness is not enough, nor is the protection of human beings for contesting wrong decisions. We should start exploring the acceptability of ADM to people while reducing the deterministic impact of ADM and creating room for resistance and disagreement with automated decisions [24]. The details and concrete requirements for implementation of these procedural safeguards by national regulators should receive more attention in research and policy making.

6 Concluding Remarks

A key challenge for modern societies is the increase of automated decisions in sensitive sectors. Empirical evidence suggests that automated decisions based on algorithmic ML models can be unfair as they might reproduce biases in historical data used to train them. Automated decisions driven by ML threaten to discriminate against legally protected groups or create new sources of unfairness that are unknown to us since they differ from the traditional concept of discrimination. That said, it was also mentioned that if we look at the full side of the glass, ADM can actually mitigate human bias and address discriminatory behaviors and beliefs hidden in the norms of society.

In this article, it was clarified that the existing categories of EU equality and non-discrimination laws do not provide grounds to cover discriminatory algorithmic decision-making. The principle of fairness in automated decision-making was studied through an interdisciplinary approach and by a literature review on different notions of fairness and assessing which notion is the best for implementation in Automated Decisions. It was noted that fairness is not a static value that we can simply code in an ML application; it is the result of a balancing process considering different notions and different factors in every given case. Thus, there is not only one possible fair outcome in all decision-making situations. Therefore, we should not rely on one notion of fairness, or relate this concept to non-discrimination only. Finally, it was concluded that algorithmic versus human fairness is not an 'either/or question'. The solution to fairness lies in the combination of these notions. This mutual cooperation is what can up to a degree replicate the complexity of the contextual equality that lies within the concept of fairness.

Future research can be done on the application of a hypothetical ADM situation by training an algorithm based on existing data sets (even synthetic

ones) and applying different ethical and fairness assessment tools, for putting different notions of fairness or a new notion based on a combination of the existing notions to the test, for the purpose of using ML as a tool to combat bias and discrimination.

References

1. Allen, R., Masters, D.: Artificial Intelligence: the right to protection from discrimination caused by algorithms, machine learning and automated decision-making. ERA Forum **20**(4), 585–598 (2019). https://doi.org/10.1007/s12027-019-00582-w
2. Allen, R., Masters, D.: Regulating for an equal AI: a new role for equality bodies meeting the new challenges to equality and non-discrimination from increased digitisation and the use of artificial intelligence. Equinet (2020)
3. Angelopoulos, C.: Sketching the outline of a ghost: the fair balance between copyright and fundamental rights in intermediary third party liability. Info (2015)
4. Arrieta, A.B., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020)
5. Barocas, S., Hardt, M., Narayanan, A.: Fairness and machine learning (2019). http://www.fairmlbook.org
6. Beauchamp, T.L., Childress, J.F., et al.: Principles of Biomedical Ethics. Oxford University Press, USA (2001)
7. van Bekkum, M., Borgesius, F.Z.: Digital welfare fraud detection and the dutch syri judgment. Eur. J. Soc. Secur. **23**(4), 323–340 (2021)
8. Bellamy, R.K., et al.: AI fairness 360: an extensible toolkit for detecting and mitigating algorithmic bias. IBM J. Res. Dev. **63**(4/5), 1–4 (2019)
9. Binns, R.: Fairness in machine learning: lessons from political philosophy. In: Conference on Fairness, Accountability and Transparency, pp. 149–159. PMLR (2018)
10. Buolamwini, J., Gebru, T.: Gender shades: intersectional accuracy disparities in commercial gender classification. In: Conference on Fairness, Accountability and Transparency, pp. 77–91. PMLR (2018)
11. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.: Fairness through awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, pp. 214–226 (2012)
12. Ebers, M., Hoch, V.R., Rosenkranz, F., Ruschemeier, H., Steinrötter, B.: The european commission's proposal for an artificial intelligence act–a critical assessment by members of the robotics and ai law society (rails). J. **4**(4), 589–603 (2021)
13. European commission: regulation (eu) 2016/679-(general data protection regulation) (2016). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32016R0679. Accessed Mar 2022
14. European commission: regulation (eu) 2016/679-(general data protection regulation) (2016)
15. European commission: ethics of artificial intelligence: statement of the ege is released (2018). https://ec.europa.eu/info/news/ethics-artificial-intelligence-statement-ege-released-2018-apr-24_en. Accessed Mar 2022
16. European commission: proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. com/2021/206 final (2021)
17. Floridi, L.: The Ethics of Information. Oxford University Press (2013)

18. Floridi, L., Cowls, J.: A unified framework of five principles for AI in society. In: Floridi, L. (ed.) Ethics, Governance, and Policies in Artificial Intelligence. PSS, vol. 144, pp. 5–17. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-81907-1_2
19. Floridi, L., et al.: AI 4people-an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Mind. Mach. **28**(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5
20. Gerards, J., Xenidis, R.: Algorithmic discrimination in Europe: challenges and opportunities for gender equality and non-discrimination law. european commission (2021)
21. Giovanola, B., Tiribelli, S.: Beyond bias and discrimination: redefining the ai ethics principle of fairness in healthcare machine-learning algorithms. AI & SOCIETY, pp. 1–15 (2022). https://doi.org/10.1007/s00146-022-01455-6
22. Giovanola, B., Tiribelli, S.: Weapons of moral construction? on the value of fairness in algorithmic decision-making. Ethics Inf. Technol. **24**(1), 1–13 (2022). https://doi.org/10.1007/s10676-022-09622-5
23. Hacker, P.: Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under eu law. Common Mark. Law Rev. **55**(4), 1143–1185 (2018)
24. Helberger, N., Araujo, T., de Vreese, C.H.: Who is the fairest of them all? public attitudes and expectations regarding automated decision-making. Comput. Law Secur. Rev. **39**, 105456 (2020)
25. High level expert group on artificial intelligence: ethics guidelines for trustworthy ai (2019)
26. Hoehndorf, R., Queralt-Rosinach, N., et al.: Data science and symbolic AI: Synergies, challenges and opportunities. Data Sci. **1**(1–2), 27–38 (2017)
27. Kleinberg, J., Ludwig, J., Mullainathan, S., Rambachan, A.: Algorithmic fairness. Aea Papers Proc. **108**, 22–27 (2018)
28. Köchling, A., Wehner, M.C.: Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Bus. Res. **13**(3), 795–848 (2020). https://doi.org/10.1007/s40685-020-00134-w
29. Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P.: Fair, transparent, and accountable algorithmic decision-making processes. Philos. Technol. **31**(4), 611–627 (2018). https://doi.org/10.1007/s13347-017-0279-x
30. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: mapping the debate. Big Data Soc. **3**(2), 2053951716679679 (2016)
31. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: mapping the debate. Big Data Soc. **3**(2), 2053951716679679 (2016). https://doi.org/10.1177/2053951716679679
32. Morison, J., Harkens, A.: Re-engineering justice? robot judges, computerised courts and (semi) automated legal decision-making. Leg. Stud. **39**(4), 618–635 (2019)
33. Palmirani, M., Sapienza, S.: Big data, explanations and knowability ragion pratica. Ragion pratica **2**, 349–364 (2021)
34. Prins, C.: Digital justice. Comput. Law Secur. Rev. **34**(4), 920–923 (2018)
35. Rajkomar, A., Dean, J., Kohane, I.: Machine learning in medicine. N. Engl. J. Med. **380**(14), 1347–1358 (2019)
36. Rawls, J.: A Theory of Justice. Harvard University Press (1971)
37. Sachan, S., Yang, J.B., Xu, D.L., Benavides, D.E., Li, Y.: An explainable ai decision-support-system to automate loan underwriting. Expert Syst. Appl. **144**, 113100 (2020)

38. Sales, L.: Algorithms, artificial intelligence and the law. Judicial Rev. **25**(1), 46–66 (2020)
39. Schoeffer, J., Machowski, Y., Kuehl, N.: A study on fairness and trust perceptions in automated decision making (2021). arXiv preprint arXiv:2103.04757
40. Shin, D., Park, Y.J.: Role of fairness, accountability, and transparency in algorithmic affordance. Comput. Hum. Behav. **98**, 277–284 (2019)
41. Siapka, A.: The ethical and legal challenges of artificial intelligence: the eu response to biased and discriminatory AI. Available at SSRN 3408773 (2018)
42. Srivastava, M., Heidari, H., Krause, A.: Mathematical notions vs. human perception of fairness: a descriptive approach to fairness for machine learning. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2459–2468 (2019)
43. Starke, C., Baleis, J., Keller, B., Marcinkowski, F.: Fairness perceptions of algorithmic decision-making: a systematic review of the empirical literature (2021). arXiv preprint arXiv:2103.12016
44. University of Montreal: montreal declaration for a responsible development of artificial intelligence (2017). https://www.montrealdeclaration-responsibleai.com/the-declaration. Accessed Mar 2022
45. Veale, M., Binns, R.: Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. Big Data Soc. **4**(2), 2053951717743530 (2017)
46. Verma, S., Rubin, J.: Fairness definitions explained. In: 2018 IEEE/ACM International Workshop on Software Fairness (fairware), pp. 1–7. IEEE (2018)
47. Wachter, S., Mittelstadt, B., Russell, C.: Why fairness cannot be automated: Bridging the gap between eu non-discrimination law and ai. Comput. Law Secur. Rev. **41**, 105567 (2021)
48. Wadsworth, C., Vera, F., Piech, C.: Achieving fairness through adversarial learning: an application to recidivism prediction (2018). arXiv preprint arXiv:1807.00199
49. Wagner, B., d'Avila Garcez, A.: Neural-symbolic integration for fairness in ai. In: CEUR Workshop Proceedings, vol. 2846 (2021)
50. Wallach, W., Allen, C.: Moral Machines: Teaching Robots Right from Wrong. Oxford University Press (2008). https://books.google.it/books?id=tMENFHG4CXcC
51. Wirtz, B.W., Weyerer, J.C., Geyer, C.: Artificial intelligence and the public sector–applications and challenges. Int. J. Public Adm. **42**(7), 596–615 (2019)
52. Zafar, M.B., Valera, I., Gomez Rodriguez, M., Gummadi, K.P.: Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In: Proceedings of the 26th International Conference on World Wide Web, pp. 1171–1180 (2017)
53. Završnik, A.: Criminal justice, artificial intelligence systems, and human rights. ERA Forum **20**(4), 567–583 (2020). https://doi.org/10.1007/s12027-020-00602-0
54. Žliobaitė, I.: Measuring discrimination in algorithmic decision making. Data Min. Knowl. Disc. **31**(4), 1060–1089 (2017). https://doi.org/10.1007/s10618-017-0506-1

Object Relational Mapping Vs. Event-Sourcing: Systematic Review

Pantelelis Michail(✉) and Kalloniatis Christos

Privacy Engineering and Social Informatics Laboratory, Department of Cultural Technology and Communication, University of the Aegean, 81100 Lesvos, Greece
{mpantel,chkallon}@aegean.gr
https://privasi.aegean.gr/

Abstract. Storing information in a software system is challenging, especially in the cloud computing era. Traditional, battle-tested methods, like Object Relational Mapping, seem inappropriate for all cases, but other alternatives feel too complicated to implement. Moreover, new software design methodologies, like Domain-Driven Design, provide alternative modeling tools for software systems, which in contrast to Object-Oriented Design, focus more on Domain-Events and their processing. Additionally, there is an increasing interest in causality when implementing software systems, mainly as a matter of accountability, tracing, and other similar properties, especially in the context of e-Government, in order to support transparency. We are now interested in a system's state history and its current state, and that is no longer the exception. Therefore, this paper focuses on Object Relational Mapping and Event-Sourcing trends for the past years as two alternatives for storing applications data through a systematic literature review. We evaluate how these two alternatives have evolved over the years under the prism of academic literature and discuss our findings, according to modern application development requirements.

Keywords: Object relational mapping · Event sourcing · Command and query responsibility segregation · Domain-Driven Design

1 Introduction

Storing and retrieving information is an essential requirement for any software system. Be it in the file-system as plain files or in some other storage management system, many approaches emerged along with tools and methodologies over the years. However, we rarely use plain files for application data storage in everyday practice, except for configuration purposes. During the '80s, relational database management systems (RDBMS) became mainstream for secure information storage and easy retrieval. With object-oriented design (OOD) methodologies that allow the development of a more complex schema, object relational mapping (ORM) tools help bridge the best of two worlds: Application development using

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 18–31, 2022.
https://doi.org/10.1007/978-3-031-12673-4_2

OOD and information storage using RDBMS. During the '90s, this duo was a defacto choice for mainstream application development. By that time, most complex applications hardly reached the boundaries of an organization, and simple client-server approaches were efficient. Works like Fowler et al. Patterns of Enterprise Application Development (PoEEA) [16], which deals to a great extent with ORM patterns, and Hophe and Wolf in Enterprise Integration Patterns (EIP) [23] which presents mainly messaging patterns, document knowledge, and tools from that era.

During the following decades, widespread acceptance of the World Wide Web as a new application development platform dramatically changed software systems' requirements. In this context effectiveness of ORM, along with the issues of Object Relational Impedance Mismatch (ORIM) [47] were criticized. In 2004, Evans proposed Domain-Driven Design (DDD) [15] as guidance for modeling modern applications. This guidance's center was the need to reduce complexity in software systems design by utilizing proper system division according to suggested patterns. Implementation details follow ten years later by Vernon [54]. DDD proposes bounded contexts as a pattern to split cohesive parts of the system under design to reduce complexity. One way for bounded context to interoperate is "messaging": exchanging messages every time a significant event happens in the system. Event Sourcing (ES) and Command Query Responsibility Segregation (CQRS) describe two patterns to implement messaging for information storage, among other usages. Fowler describes Event Sourcing (ES) [17] around 2005 furthermore, CQRS was made famous by Dahan [13] and Young [59] around 2010, while also described by Fowler [18].

ES opens up the possibility of tracking every event in the system under design in a very compact manner. While ORM imprint only the current snapshot of a systems state, event sourcing captures by design the sequence of events that lead to the current state. It is far from trivial to achieve the same result with a relational database; moreover, such an endeavor usually leads to a complex schema and application logic. Temporal data model support, although included in SQL:2011 revision, it is still not widely adopted while also immature implementation wise [28].

At that time, such property was not a requirement for most common applications. On the contrary, the emergence of the Cloud during 2010 dramatically changed the application development landscape. Interconnection between unrelated applications is the rule and out of any notion of control. Information can flow from social networks to e-shops and then into customer relationship management systems. Anyone can use the Cloud, build an application for the Cloud, and integrate with the Cloud. Such an evolution changed the prioritization of requirements dramatically when designing software systems [26]. Qualities like security and privacy must be an integral part of every system. Not as a "byproduct" but "by design." We should reevaluate our information storage modeling tools and data storage mechanisms in this context.

The role of Service-Oriented Architecture (SOA), Event-Driven Architectures (EDA), and the utilization of Enterprise Service Bus (ESB) in e-Government

systems is mentioned by Castellano et al. [9]. A proposal for a cooperative framework for government agencies as early as 2005 highlights the importance of EDA in mitigating the cost and complexity of deploying distributed enterprise applications long before cloud applications became the norm. More recently, Oumkaltoum et al. [38] propose an EDA-based approach for e-Government interoperability and data integration. While both these works utilize events to aid mainly in terms of interoperability, other aspects of e-Government, such as transparency, are expected to be affected positively.

This paper focuses on the history of Object Relational Mapping and Event-Sourcing for the past years as two alternatives for storing applications' data through a systematic literature review. ORMs are well understood and documented but miss dynamic aspects of software systems that are now in demand. ES is a promising solution to overcome such deficiencies.

2 Methods

In this section, we describe the implemented methodology during our research. Our review process follows the PRISMA statement [40]. We conducted our research in December 2021. We only used Scopus (https://www.scopus.com) as our search database since it indexes all well-known sources of our interest and has strict quality criteria for its content. Furthermore, we only focus on academic research in articles and conference publications in this article. Scopus also provides us with handy tools to extract the bibliographic results for further analysis. We summarize eligibility criteria in Table 1.

Table 1. Eligibility criteria.

Inclusion criteria	– Academic journal, conference or workshop, referencing ORM, DDD, or ES – In the subject of Computing – Written in English – Any publication date
Exclusion criteria	– Duplicates – Subject other than Computing – Full-text is not accessible – Books

Our two-step search strategy first identified target publications using the following search string:

```
(TITLE (event AND sourcing) ) OR
(TITLE (object AND relational AND mapping) ) OR
(TITLE (domain AND driven AND design ) ) OR
```

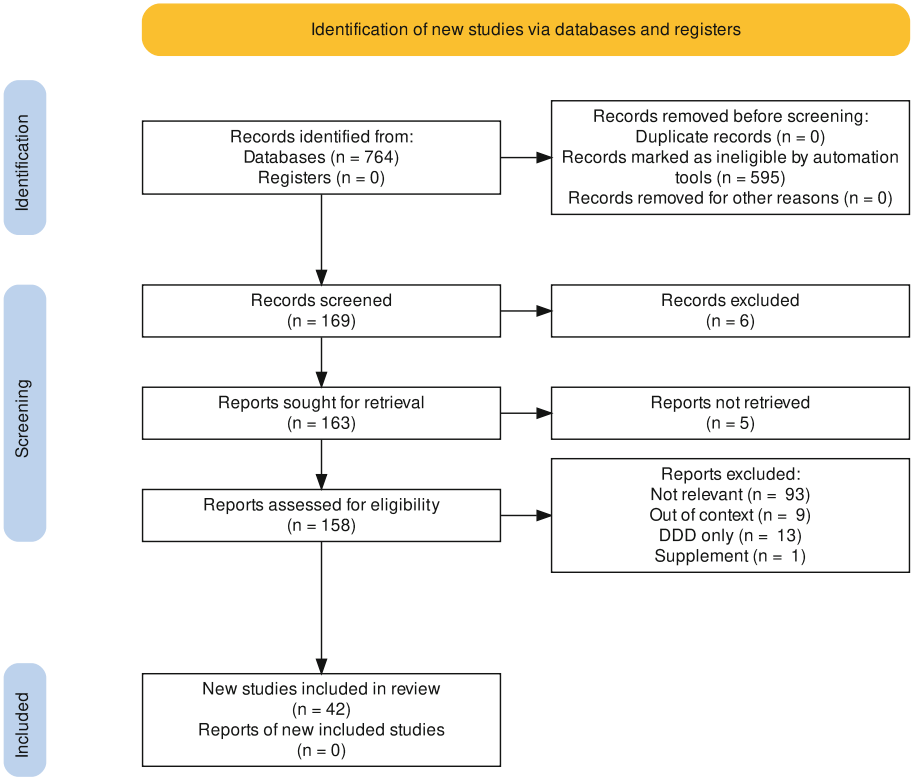

Fig. 1. PRISMA flow diagram.

```
(TITLE (domain AND events) )  OR
((TITLE (cqrs) OR
    TITLE (command AND query AND
            responsibility AND segregation) )
```

Relevant publications should mention "event sourcing," "domain-driven design," "object relational mapping," or something similar in the title. Closely related to ES is the field of "process mining," but our perspective focuses on ES as a storage mechanism in this paper, so we do not include that search term. We narrowed down the results list by applying the same query regarding relevant topics, including COMPUTING and excluding others, as follows[1]:

```
AND  (LIMIT-TO (SUBJAREA , "COMP" ) )
AND
(EXCLUDE (SUBJAREA , "MEDI") OR EXCLUDE (SUBJAREA , "MATH") OR
 EXCLUDE (SUBJAREA , "ENGI") OR EXCLUDE (SUBJAREA , "SOCI") OR
```

[1] Subquery to exclude topics, is correct with OR in contrast to our intuition, due to Scopus query builder.

```
EXCLUDE (SUBJAREA , "DECI") OR EXCLUDE (SUBJAREA , "ARTS") OR
EXCLUDE (SUBJAREA , "BUSI") OR EXCLUDE (SUBJAREA , "PHYS") OR
EXCLUDE (SUBJAREA , "MATE") OR EXCLUDE (SUBJAREA , "BIOC") OR
EXCLUDE (SUBJAREA , "ENER") OR EXCLUDE (SUBJAREA , "AGRI") OR
EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA , "CHEM") OR
EXCLUDE (SUBJAREA , "ENVI") OR EXCLUDE (SUBJAREA , "NEUR") OR
EXCLUDE (SUBJAREA , "PSYC") OR EXCLUDE (SUBJAREA , "EART") OR
EXCLUDE (SUBJAREA , "IMMU") OR EXCLUDE (SUBJAREA , "PHAR") )
```

Unbounded search returned 764 results, and the bounded one, only 169. Unfortunately, the keywords we use have different meanings in different disciplines. We excluded six unreachable articles and five books. One hundred fifty-eight articles were retrieved. Afterward, we excluded ninety-three articles because their content was out of our scope, due to their title or abstract review. Nine articles were out of scope, even though their title/abstract was relevant but not the actual content. Finally, we removed thirteen results dealing explicitly with DDD without mentioning ORM or ES because DDD is not our primary focus.

Our selection process is summarized in PRISMA flow diagram (Fig. 1) and resulted in 42 articles. This diagram was generated using the tool provided by Haddaway et al. [22].

The research questions, summarized in Table 2, are the basis of our research.

Table 2. Research questions.

RQ 1	Is there a shift from traditional ORM to ES over the years? *Rational:* Spot current research with ORM and ES, the shift or existing obstacles with the transition
RQ 2	Are modern cloud application development needs covered by academia? *Rational:* Evaluate which aspects of ORM and ES, are covered, in the academia

3 Results

Naturally, our search results fall into two groups: Publications mentioning ES and ORM. Therefore, we consolidate results mentioning DDD into one of these groups. Moreover, early publications for ES refer more to EIP than ES. Figure 2 shows the appearance per group and year. We got 13 articles on ES and 29 articles on ORM. In the following subsections, we summarize publications under review in chronological order.

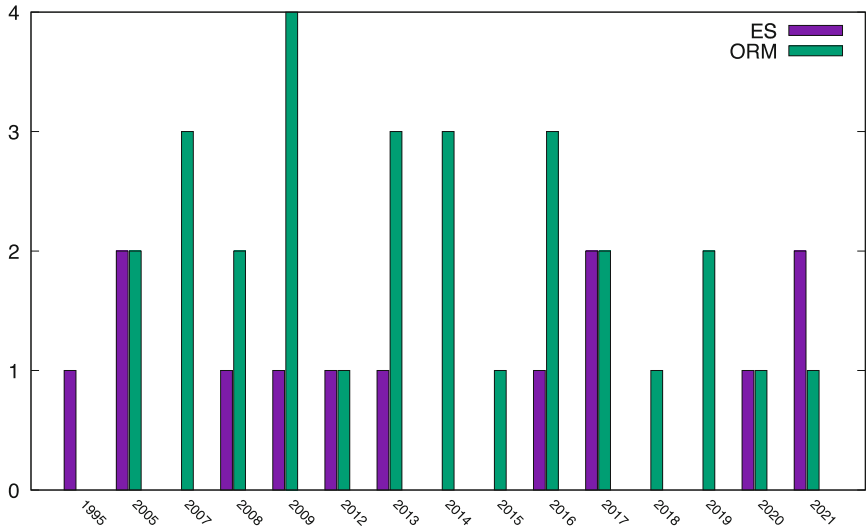

Fig. 2. Paper distribution over the years.

3.1 Event Sourcing

Publications in this category follow a pattern based on the year of publication. Works before 2010 mention "event" on the aspect of EIP rather than ES. Technologies such as Java and CORBA, Publish/Subscribe Pattern and Middleware are also mentioned. Common concepts include Event Domains, Event Types, and Event-Based/Event-Driven Architecture.

Domain and Port views of a system are described by Poo et al. [44], in 1995, as differentiation of "contemporary" object-oriented development methods. It stresses the dynamic behavior of objects and the importance of "events" during system design. "Object-Relationship diagrams at the event level to allow the analysis to focus on a smaller subset of objects that exist in application domain", which reminds us of bounded context in DDD, later to appear in the literature.

Aleksy et al. [3] discuss the management of Event Domains, utilizing CORBA [2] to implement the Publisher/Subscriber design pattern [23]. Pesonen and Bacon [41] present secure event types and public-key cryptography in an attempt to secure publish/subscribe systems. Baldoni et al. [6] discuss challenges in Event-Based data dissemination in event-driven architectures. Bacon et al. [4], mention privacy, security, and policies in the context of Event-Based systems noting that such systems require fundamentally different approaches to security due to their heterogeneous and loosely coupled nature and propose a role-based access control (RBAC) solution.

Concepts like CQRS and Event sourcing emerged after 2010, a year that coincides with Cloud's third period of existence [1]. In a case study, Kabbedijk [25] propose seven subpatterns to CQRS while discussing the variability of online business software. Rajkovic et al. [45] propose CQRS to improve query response

times in a medical information system. Erb et al. [14] propose ES, in the context of the Actor Model, to track causality in distributed applications. Maddodi and Jansen [33] provide a case study for a CQRS-based enterprise application. Overeem et al. [39] discuss schema evolution challenges in event sources system, also discussing CQRS and concepts of continuous deployment. Maddodi et al. [34] discuss performance measurements in event sourcing applications. Braun et al. [8] discuss eventual consistency challenges in distributed data-intensive systems. Finally, Lima et al. [30] propose the addition of tracing to Event Sourcing Systems as a way to improve observability further.

3.2 Object Relational Mapping

Regarding ORM, there is no distinct pattern among publications across the years. Most research investigates the Java framework and Hibernate ORM, while references to Microsoft Entity Model Framework are less frequent. Most work leans towards proposing frameworks to ease the modeling process and the maintainability of mappings, including fuzzy logic and genetic algorithms. There are extended discussions of ORM performance and ORM problems due to ORIM.

Philippi [42] discusses ORIM and the model-driven generation of object relational mappings. Snyder and O'Connor [48] propose an ORM implementation in Java. Landre et al. [29] provides a case study of a system reimplementation utilizing object-oriented techniques, DDD, and a "proper" ORM tool. Lodhi and Ghazali [31] propose another ORM technique. Wellhausen [57] propose a pattern to improve the query performance of ORM. O'Neil [37] review Hibernate over Microsoft's Entity Data Model in the context of web application development. Pohajalainen and Taina [43] propose enhancements of eager and lazy fetching of associations, thus improving query performance. Kowark et al. [27] implement a reflective object relational mapper that obsoletes manual maintenance and synchronization. Torres et al. [51] propose a UML profile, according to JPA persistence standards, for ORM modeling, verification, and software implementation. Van Zyl et al. [53] discusses performance optimizations in ORM. Xia et al. [58] discuss Hibernate as an ORM implementation and its performance implications while assuming Object-Oriented Databases will outperform the need for ORMs.

Wallek et al. [56] propose a fuzzy expert system to assist ORM during application design. Bernstein et al. [7] discuss the need for ORM to validate user-defined mappings, proposing an efficient algorithm for the case. Gawarkiewicz et al. [19] benchmarks recursive tree-like query performance in ORM. Murakami et al. [35] propose an ORM implementation based on a conceptual model and compare it with other ORM implementations. Chen et al. [10,11] evaluate performance and discusses anti-patterns in ORM, proposing a framework for automated anti-pattern detection. Ghandeharizadeh and Mutha [20] evaluate Hibernate's performance over its JDBC counterpart. Torres et al. [50] propose a graphical notation to help visualize ORM. Goetz and Kuhn [21] propose an approach to reduce configuration efforts of ORM and support schema evolution. Lorentz et al. [32]

review ORM strategies, guidelines, and best practices. Singh et al. [46] provide a genetic algorithm to optimize the performance of ORM configurations. Bagheri et al. [5] implement a formally precise technique for automated synthesis and dynamic analysis of trade off spaces for ORM. Torres et al. [52] provide an extended ORM review among different ORM implementations. Ismailova and Kosikov [24] propose a metamodel to support searching objects and their sources to support ORM. Nazario et al. [36] propose a framework to detect twelve ORM-related problems. Vial [55] discusses practical experience in using ORM engines, outlining complexity due to ORIM. Colley et al. [12] reviews the effects of ORIM, in terms of ORM efficiency, among other data persistence alternatives. Sotiropoulos et al. [49] implement a testing framework for cross-checking different ORM implementations among different DBMS.

4 Discussion

During our research, we can not observe a decline in ORM publications or a change of interest in research topics, despite the emergence of competing technologies, such as NoSQL databases and Event Stores. Unfortunately, we did not observe the "extinction" of ORM due to the emergence of Object-Oriented Databases. In order to make results more obvious we categorize papers in our review according to publication type and context of interest to support our two research questions.

To answer RQ1, we expect research for any technical subject to pass three main stages during its evolution. At first, there are frameworks and methodologies to apply the subject, then, as it emerges, we see more and more case studies and evaluations, and finally, maturity calls for a review.

Table 3. Aspects/Categories of interest.

Publication type	
Review	Is some sort of review, systematic or not
Tool/methodology	Presents a tool or a methodology regarding the specific technology
Case study/benchmark	Include a case study or benchmark
Contextual	
Performance	Deals with performance issues
Security	Deals with security aspects or mentions security
Privacy	Deals with privacy aspects or mentions security
Policy/legal	Deals or mentions policy or legal issues
Testing/validation	Mentions testing or a validation methodology

Contextually we seek, for RQ2, research publications regarding performance, security, privacy, policy/legal, and testing/validation. Although these are not

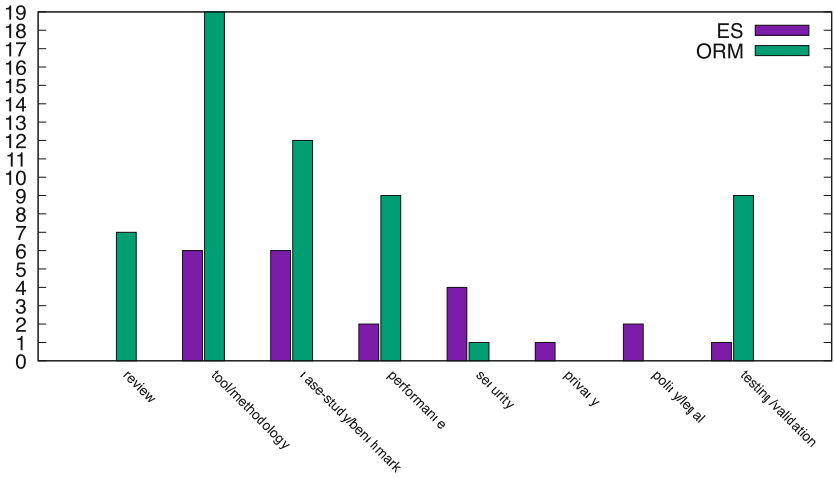

Fig. 3. Paper summary.

functional requirements of many software systems, their importance is critical and must be taken into account early in the system design phases for a system to be functional. Poor performance in the Cloud era costs money. Security, privacy, and policy have mostly legal implications for business owners. Finally, testing/validation is a critical aspect of software quality. Table 3 summarizes our review directions.

Figure 3 shows paper distribution graphically, while Table 4 provides a detailed grid view of our findings. As we can see, most ORM-related publications propose a tool, methodology, or guidance. In the case of ES, there is no explicit review paper. Our review found no specific implementations, although many well-known providers, like EventStore, exist. Therefore, we can infer that ES and CQRS implementations are primarily ad-hoc. Observing the summary table, we can see a smooth distribution for ORMs across all years. The same applies to ES, except for the case study/benchmark research with a six-year gap between 2013 and 2020. Case studies for the latter imply that ES-related technologies are actively applied again. So the short answer to RQ1 is that we do not observe any shift from ORM to ES technologies.

On the other hand, while performance and testing are a hot topic in ORM, ES research is more aware of implementing security, privacy, and legal issues. In contrast, the amount of publications in that area is not encouraging. We do not imply that there is no research on privacy and security, however, such research should materialize more in implementing tools and frameworks around these technologies. So the answer to RQ2 is negative. We need more intensive research in ES for data persistence. While many techniques and patterns apply in practice, methods and tools require more documentation under our academic lens.

Table 4. Summary of review findings by category (ES/ORM).

Year	Review	Tool/ methodology	Case-study/ benchmark	Performance	Security	Privacy	Policy/ legal	Testing/ validation
1995		1/0						
2005	0/1	2/2	1/1		1/0			0/1
2007		0/1	0/2	0/1				0/1
2008	0/1	0/1	0/1					
2009	0/1	0/2	0/2	0/2	1/0	1/0	1/0	0/1
2012		0/1	1/0		1/0			0/1
2013	0/1	0/3	1/1	1/1				0/1
2014		0/2	0/1	0/2				0/1
2015		0/1						
2016	0/1	1/2	0/2	0/2				0/1
2017	0/1	1/1	0/1		1/0			
2018		0/1						
2019	0/1	0/1						0/1
2020			1/0	1/1	0/1			
2021		1/1	2/1				1/0	1/1
Total	0/7	6/19	6/12	2/9	4/1	1/0	2/0	1/9

Table 5. ORM challenges.

Challenges	Description
Configuration	ORM are usually suboptimally configured
Mapping strategies (class inheritance)	Correct mapping relies to programmer experience
Performance	ORIM affects performance during read/write cycles
Quering/n+1 query problem	Generated queries are not always optimized
Maintenance/schema synchronization	Object model and db schema fall easily out of sync
Roundtrip testing	Validate correct operation of the ORM framework

Finally, in Table 5, we summarize the most important ORM challenges we observed during our review. ES and CQRS propose technics that utilize optimized read models to mitigate most of these challenges in a cloud-friendly way, especially performance issues. In addition, Domain-Driven Design also provides modeling guidance to overcome most ORIM related problems.

5 Conclusion

We performed a systematic literature review on object relational mapping and event sourcing. Our research aimed to evaluate how these two alternatives regarding data persistence have evolved over the years, in modern application development. ORM dominates data persistence technologies despite its failure to implement persistence for object-oriented systems accurately. On the other hand, event sourcing is considered complicated. To our perception, event sourcing has such a reputation because it alters how we view a software system. We are used

to modeling a static view of reality using class and entity-relationship diagrams. Event sourcing requires a more dynamic view of software systems. On the contrary, we are not used to modeling such a view quickly. In this direction, our future research will focus on implementing some "object to event sourcing" mapping. That is to utilize event sourcing as the data persistence layer and apply well-known object-oriented modeling methods during the system design, with e-Government as the primary domain.

References

1. Cloud computing. https://en.wikipedia.org/wiki/Cloud_computing#2010s. Accessed 20 Dec 2021
2. Common Object Request Broker Architecture (CORBA). https://www.omg.org/corba/. Accessed 23 Dec 2021
3. Aleksy, M., Köblitz, L., Schader, M.: Management of event domains-the mediator-approach. In: Proceedings of the International Conference on Parallel and Distributed Systems-ICPADS. vol. 2, pp. 438–442 (2005). https://doi.org/10.1109/ICPADS.2005.182
4. Bacon, J., Eyers, D., Singh, J., Shand, B., Migliavacca, M., Pietzuch, P.: Security in multi-domain event-based systems [sicherheit in ereignis-basierten mehrdománensystemen]. IT-Inf. Technol. **51**(5), 277–284 (2009). https://doi.org/10.1524/itit.2009.0552
5. Bagheri, H., Tang, C., Sullivan, K.: Automated synthesis and dynamic analysis of tradeoff spaces for object-relational mapping. IEEE Trans. Softw. Eng. **43**(2), 145–163 (2017). https://doi.org/10.1109/TSE.2016.2587646
6. Baldoni, R., Querzoni, L., Scipioni, S.: Event-based data dissemination on inter-administrative domains: is it viable? In: Proceedings of the IEEE Computer Society Workshop on Future Trends of Distributed Computing Systems, pp. 44–50 (2008). https://doi.org/10.1109/FTDCS.2008.14
7. Bernstein, P., Jacob, M., Pérez, J., Rull, G., Terwilliger, J.: Incremental mapping compilation in an object-to-relational mapping system. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 1269–1280 (2013). https://doi.org/10.1145/2463676.2465294
8. Braun, S., Bieniusa, A., Elberzhager, F.: Advanced domain-driven design for consistency in distributed data-intensive systems. In: Proceedings of the 8th Workshop on Principles and Practice of Consistency for Distributed Data, PaPoC 2021 (2021). https://doi.org/10.1145/3447865.3457969
9. Castellano, M., Pastore, N., Arcieri, F., Summo, V., de Grecis, G.: An e-government cooperative framework for government agencies. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences. IEEE (2005). https://doi.org/10.1109/hicss.2005.82
10. Chen, T.H., Shang, W., Jiang, Z., Hassan, A., Nasser, M., Flora, P.: Detecting performance anti-patterns for applications developed using object-relational mapping. In: Proceedings-International Conference on Software Engineering, pp. 1001–1012 (2014). https://doi.org/10.1145/2568225.2568259
11. Chen, T.H., Shang, W., Jiang, Z., Hassan, A., Nasser, M., Flora, P.: Finding and evaluating the performance impact of redundant data access for applications that are developed using object-relational mapping frameworks. IEEE Trans. Softw. Eng. **42**(12), 1148–1161 (2016). https://doi.org/10.1109/TSE.2016.2553039

12. Colley, D., Stanier, C., Asaduzzaman, M.: Investigating the effects of object-relational impedance mismatch on the efficiency of object-relational mapping frameworks. J. Database Manag. **31**(4), 1–23 (2020). https://doi.org/10.4018/JDM.2020100101
13. Dahan, U.: Clarified CQRS (2009). https://udidahan.com/2009/12/0/
14. Erb, B., Habiger, G., Hauck, F.: On the potential of event sourcing for retroactive actor-based programming. In: ACM International Conference Proceeding Series, vol. 17-July-2016 (2016). https://doi.org/10.1145/2957319.2957378
15. Evans, E.: Domain-Driven Design. Addison-Wesley (2004)
16. Fowler, M.: Patterns of Enterprise Application Architecture. Addison-Wesley (2003)
17. Fowler, M.: Event Sourcing (2005). https://martinfowler.com/eaaDev/EventSourcing.html
18. Fowler, M.: CQRS (2011). https://www.martinfowler.com/bliki/CQRS.html
19. Gawarkiewicz, M., Wisńiewski, P., Stencel, K.: Enhanced segment trees in object-relational mapping. In: ACM International Conference Proceeding Series, pp. 122–128 (2013). https://doi.org/10.1145/2490257.2490291
20. Ghandeharizadeh, S., Mutha, A.: An evaluation of the hibernate object-relational mapping for processing interactive social networking actions. In: ACM International Conference Proceeding Series, vol. 04–06-December-2014, pp. 64–70 (2014). https://doi.org/10.1145/2684200.2684285
21. Götz, S., Kühn, T.: Models@run.time for object-relational mapping supporting schema evolution. In: CEUR Workshop Proceedings, vol. 1474, pp. 41–50 (2015)
22. Haddaway, N.R., McGuinness, L.A., Pritchard, C.C.: PRISMA2020: R package and ShinyApp for producing PRISMA 2020 compliant flow diagrams (2021). https://doi.org/10.5281/ZENODO.5082518
23. Hohpe, G., Woolf, B.: Enterprise Integration Patterns. Addison-Wesley (2004)
24. Ismailova, L., Kosikov, S.: Metamodel of transformations of concepts to support the object-relational mapping. In: Procedia Computer Science, vol. 145, pp. 260–265 (2018). https://doi.org/10.1016/j.procs.2018.11.055
25. Kabbedijk, J., Jansen, S., Brinkkemper, S.: A case study of the variability consequences of the CQRS pattern in online business software. In: ACM International Conference Proceeding Series (2012). https://doi.org/10.1145/2602928.2603078
26. Kalloniatis, C., Mouratidis, H., Vassilis, M., Islam, S., Gritzalis, S., Kavakli, E.: Towards the design of secure and privacy-oriented information systems in the cloud: identifying the major concepts. Comput. Stan. Interfaces **36**(4), 759–775 (2014). https://doi.org/10.1016/j.csi.2013.12.010
27. Kowark, T., Hirschfeld, R., Haupt, M.: Object-relational mapping with SqueakSave. In: Proceedings of the International Workshop on Smalltalk Technologies 2009, IWST 2009-ESUG 2009 Smalltalk Joint Event, pp. 87–100 (2009). https://doi.org/10.1145/1735935.1735951
28. Kumar, S., Rishi, R.: A relative analysis of modern temporal data models. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 2851–2855 (2016)
29. Landre, E., Wesenberg, H., Ølmheim, J.: Agile enterprise software development using domain-driven design and test first. In: Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications, OOPSLA, pp. 983–993 (2007). https://doi.org/10.1145/1297846.1297967
30. Lima, S., Correia, J., Araujo, F., Cardoso, J.: Improving observability in event sourcing systems. J. Syst. Softw. **181**, 111015 (2021). https://doi.org/10.1016/j.jss.2021.111015

31. Lodhi, F., Ghazali, M.: Design of a simple and effective object-to-relational mapping technique. In: Proceedings of the ACM Symposium on Applied Computing, pp. 1445–1449 (2007). https://doi.org/10.1145/1244002.1244312
32. Lorenz, M., Hesse, G., Rudolph, J.P.: Object-relational mapping revised-a guideline review and consolidation. In: ICSOFT 2016-Proceedings of the 11th International Joint Conference on Software Technologies, vol. 1, pp. 157–168 (2016). https://doi.org/10.5220/0005974201570168
33. Maddodi, G., Jansen, S.: Responsive software architecture patterns for workload variations: a case-study in a CQRS-based enterprise application. In: CEUR Workshop Proceedings, vol. 2047, p. 30 (2017)
34. Maddodi, G., Jansen, S., Overeem, M.: Aggregate architecture simulation in event-sourcing applications using layered queuing networks. In: ICPE 2020-Proceedings of the ACM/SPEC International Conference on Performance Engineering, pp. 238–245 (2020). https://doi.org/10.1145/3358960.3375797
35. Murakami, T., Amagasa, T., Kitagawa, H.: Dbpowder: a flexible object-relational mapping framework based on a conceptual model. In: Proceedings-International Computer Software and Applications Conference, pp. 589–598 (2013). https://doi.org/10.1109/COMPSAC.2013.93
36. Nazário, M., Guerra, E., Bonifácio, R., Pinto, G.: Detecting and reporting object-relational mapping problems: an industrial report. In: International Symposium on Empirical Software Engineering and Measurement, vol. 2019-Septemer (2019). https://doi.org/10.1109/ESEM.2019.8870163
37. O'Neil, E.: Object/relational mapping 2008: Hibernate and the entity data model (edm). In: Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 1351–1356 (2008). https://doi.org/10.1145/1376616.1376773
38. Oumkaltoum, B., Idrissi, M.E., Benany, M.M.E., Omar, E.B.: Business intelligence and EDA based architecture for interoperability of e-government data services. In: 2019 IEEE International Smart Cities Conference (ISC2). IEEE (2019). https://doi.org/10.1109/isc246665.2019.9071769
39. Overeem, M., Spoor, M., Jansen, S.: The dark side of event sourcing: managing data conversion. In: SANER 2017–24th IEEE International Conference on Software Analysis, Evolution, and Reengineering, pp. 193–204 (2017). https://doi.org/10.1109/SANER.2017.7884621
40. Page, M.J., et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ **372**, n71 (2021). https://doi.org/10.1136/bmj.n71
41. Pesonen, L., Bacon, J.: Secure event types in content-based, multi-domain publish/subscribe systems. In: SEM 2005-Proceedings of the 5th International Workshop on Software Engineering and Middleware, pp. 98–105 (2005). https://doi.org/10.1145/1108473.1108495
42. Philippi, S.: Model driven generation and testing of object-relational mappings. J. Syst. Softw. **77**(2), 193–207 (2005). https://doi.org/10.1016/j.jss.2004.07.252
43. Pohjalainen, P., Taina, J.: Self-configuring object-to-relational mapping queries. In: Principles and Practice of Programming in Java-Proceedings of the 6th International Conference, PPPJ 2008, pp. 53–59 (2008). https://doi.org/10.1145/1411732.1411740
44. Poo, D., Lee, S.Y.: Domain object identification through events and functions. Inf. Softw. Technol. **37**(11), 609–621 (1995). https://doi.org/10.1016/0950-5849(95)98298-T
45. Rajković, P., Janković, D., Milenković, A.: Using CQRS pattern for improving performances in medical information systems. In: CEUR Workshop Proceedings, vol. 1036, pp. 86–91 (2013)

46. Singh, R., Bezemer, C.P., Shang, W., Hassan, A.: Optimizing the performance-related configurations of object-relational mapping frameworks using a multi-objective genetic algorithm. In: ICPE 2016-Proceedings of the 7th ACM/SPEC International Conference on Performance Engineering, pp. 309–320 (2016). https://doi.org/10.1145/2851553.2851576
47. Smith, K.E., Zdonik, S.B.: Intermedia: a case study of the differences between relational and object-oriented database systems. ACM SIGPLAN Not. **22**(12), 452–465 (1987). https://doi.org/10.1145/38807.38849
48. Snyder, M., O'Connor, T.: Object-relational mapping in java with SimpleORM. Dr. Dobb's J. **30**(12), 34–36 (2005)
49. Sotiropoulos, T., Chaliasos, S., Atlidakis, V., Mitropoulos, D., Spinellis, D.: Data-oriented differential testing of object-relational mapping systems. In: Proceedings-International Conference on Software Engineering, pp. 1535–1547 (2021). https://doi.org/10.1109/ICSE43902.2021.00137
50. Torres, A., Galante, R., Pimenta, M.: ENORM: an essential notation for object-relational mapping. SIGMOD Rec. **43**(2), 23–28 (2014). https://doi.org/10.1145/2694413.2694418
51. Torres, A., Galante, R., Pimenta, M.: Towards a uml profile for model-driven object-relational mapping. In: SBES 2009–23rd Brazilian Symposium on Software Engineering, pp. 94–103 (2009). https://doi.org/10.1109/SBES.2009.22
52. Torres, A., Galante, R., Pimenta, M., Martins, A.: Twenty years of object-relational mapping: a survey on patterns, solutions, and their implications on application design. Inf. Softw. Technol. **82**, 1–18 (2017). https://doi.org/10.1016/j.infsof.2016.09.009
53. Van Zyl, P., Kourie, D., Coetzee, L., Boake, A.: The influence of optimisations on the performance of an object relational mapping tool. In: ACM International Conference Proceeding Series, pp. 150–159 (2009). https://doi.org/10.1145/1632149.1632169
54. Vernon, V.: Implementing Domain-Driven Design. Addison-Wesley (2013)
55. Vial, G.: Lessons in persisting object data using object-relational mapping. IEEE Softw. **36**(6), 43–52 (2019). https://doi.org/10.1109/MS.2018.227105428
56. Walek, B., Bartoš J., Klimeš C.: Fuzzy tool for reasoning in object relational mapping of information system. In: 2012 35th International Conference on Telecommunications and Signal Processing, TSP 2012-Proceedings, pp. 81–85 (2012). https://doi.org/10.1109/TSP.2012.6256257
57. Wellhausen, T.: Object prefetch filter, a pattern for improving the performance of object retrieval of object-relational mapping tools. In: EuroPLoP 2007–12th European Conference on Pattern Languages of Programs, pp. 527–536 (2007)
58. Xia, C., Yu, G., Tang, M.: Efficient implement of orm (object/relational mapping) use in j2ee framework: hibernate. In: Proceedings-2009 International Conference on Computational Intelligence and Software Engineering, CiSE 2009 (2009). https://doi.org/10.1109/CISE.2009.5365905
59. Young, G.: CQRS documents (2010). https://cqrs.files.wordpress.com/2010/11/cqrs_documents.pdf

Foundation Models in Healthcare: Opportunities, Biases and Regulatory Prospects in Europe

Malwina Anna Wójcik(✉)

Science and Technology, University of Bologna, Bologna, Italy
malwinaanna.wojcik2@unibo.it

Abstract. This article concerns the rise of a new paradigm in AI - "foundation models," which are pre-trained on broad data at scale and subsequently adapted to particular downstream tasks. In particular, it explores the issue from the perspective of healthcare and biomedicine, focusing on the benefits of foundation models, as well as their propensity to encode bias, which threatens to exacerbate discriminatory practices already experienced by patients in Europe. Section 1 offers a brief introduction concerning the use of AI in healthcare and biomedicine and the problem of divergencies in access to and quality of healthcare across Europe. Section 2 familiarises the reader with the technical qualities of foundation models and recent developments in the field. Section 3 explains how the new health data strategy proposed by the EU could foster the development of foundation models in healthcare. Section 4 elaborates on their benefits in healthcare and biomedicine, while Sect. 5 explores the risk of bias exhibited by foundation models. Section 6 comments on the uncertain status of foundation models under the proposed Artificial Intelligence Act and offers brief recommendations concerning future regulation. Section 7 concludes.

Keywords: Foundation models · General purpose models · Healthcare · Artificial intelligence act · The European strategy for data · Algorithmic bias

1 Introduction

The growing deployment of Artificial Intelligence (AI) in medical diagnosis, prognosis and benefit allocation offers a promise of reducing healthcare disparities by breaking the healthcare iron triangle, simultaneously optimizing cost, access, and quality of care through faster and more accurate decision making. As underlined by the Broadband Commission for Sustainable Development report, integration of AI-enabled tools into healthcare delivery can help to address national and global health challenges, such as shortage of health workers, global health threats, dual burden of disease, growing inequalities and misinformation [11]. Similarly, AI carries enormous potential in biomedical contexts. For example, AI-based technologies can be used to identify and develop clinical best practices and fuel research in the field of genomics which requires analysis of large and complex datasets [53]. Deployment of AI in medicine development is expected

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 32–46, 2022.
https://doi.org/10.1007/978-3-031-12673-4_3

to simplify and accelerate target validation and identification of biomarkers, analysis of clinical data, pharmacovigilance and clinical use optimization [36].

Researchers argue that progress of AI in healthcare and biomedicine can be augmented by the use of powerful models, such as OpenAI's GPT-3, that are pre-trained on broad data at scale and subsequently adapted to particular downstream tasks [37]. An interdisciplinary study conducted at the Stanford Institute for Human-Centered Artificial Intelligence describes these models as "foundation" models in order to emphasize the paradigm shift in building task-specific AI systems which are increasingly based on general-purpose systems [10]. While foundation models exhibit impressive capabilities that could transform healthcare delivery, biomedical research and drug development, there is increasing evidence of their adverse societal impact, including entrenchment of bias [7].

Exacerbation of social inequalities constitutes one of the crucial challenges for the future deployment of foundation models in healthcare and biomedicine in Europe, which continues to struggle with significant divergencies in access to and quality of healthcare [23]. In 2020 the European Commission identified healthcare inequalities as one of the four main challenges facing public health in the EU [28]. Researchers highlight socially determined health disparities within Europe including, inter alia, differences in life expectancy, mortality rates for cancer and cardiovascular diseases, patterns of unhealthy lifestyle, unequal access to healthcare, and unmet medical needs [51]. The report of Fundamental Rights Agency (FRA) revealed that vulnerable groups of patients are often subject to discriminatory practices [26]. A more recent study conducted by Equinet, the network of European equality bodies, indicated that the existing patterns of discrimination in healthcare have been exacerbated due to the COVID-19 pandemic [31].

While the new European strategy on health data is likely to have a positive effect on innovation, enabling the development of foundation models in healthcare and biomedicine, the proposed Artificial Intelligence Act (AIA) [43] does not offer sufficient protection against the harms caused by foundation models. This paper argues that the absence of specific obligations concerning developers of foundation models and the proposed exclusion of these models from the scope of the Act is highly problematic from the point of view of algorithmic bias in healthcare. It underlines that fostering safe innovation for patients in Europe requires establishing a clear regulatory framework for foundation models in healthcare and beyond. The rest of the paper is structured as follows: Sect. 2 explains the key features of foundation models; Sect. 3 explores the new European strategy on data, pointing out how it could facilitate the development of foundation models in healthcare; Sect. 4 introduces possible opportunities offered by foundation models in healthcare and biomedicine; Sect. 5 explores the risk of bias that they exhibit; Sect. 6 comments on the uncertain status of foundation models under the AIA and offers brief recommendations concerning future regulation; Sect. 7 concludes.

2 What Are Foundation Models?

So far, the development of foundation models has been particularly advanced in the field of Natural Language Processing (NLP). In the recent years we have witnessed a consistent growth of large language models that are characterized by a huge size of training data

and millions or billions of parameters. In machine learning, the term "parameter" refers to values that the model itself chooses and optimizes during training [3]. A prominent example of a large language model is Google's BERT that was developed with a purpose of understanding the context behind search engine queries. Encompassing 345 million parameters, BERT can be adapted to perform a wide range of language understanding tasks, such as text classification, reading comprehension or question answering. GPT-3, OpenAI's latest large language prediction model, is largely trained on data scrapped from the Internet and has as many as 175 billion parameters. This allows the model to thrive at generating text, including creative and fluent human-like tales. Building on GPT-3 language understanding capabilities, OpenAI developed DALL-E, a model which creates images based on natural language description. Its latest version, DALL-E 2, exhibits enhanced performance in generating creative, photorealistic and accurate images in high resolution [47].

What makes foundation models extremely powerful is their larger-than-ever scale, which is made possible by the availability of big data, considerable advancements in computer hardware and the introduction of transformer networks, which enable capturing and representing real-world images and information. However, although foundation models are often labelled as "the new AI", they are in fact based on two well-established machine learning techniques: deep neural networks and self-supervised learning.

Deep neural networks mimic the functioning of human brain through a multilayer artificial neuronal network, whereby "each hidden layer combines the values in its preceding layer and learns more complicated functions of the input" [3]. Therefore, deep neural networks thrive at tasks requiring growing levels of abstraction, such as image recognition. An important feature of deep learning is the possibility to pre-train the model on a surrogate task and subsequently fine-tune it to specific applications. This idea, often referred to as "transfer learning", is the key component of foundation models, which are characterized by adaptability to downstream tasks.

Self-supervised learning is crucial for training of foundation models, because the gargantuan amount of data on which they are based would be extremely cumbersome and costly to label. The process of self-supervised learning takes place on unannotated data, allowing the algorithm to analyze and label information without human input. Large NLP models, such as GPT-3 or BERT, use self-supervised learning to "fill-in the gaps", for example by predicting the missing word on the basis of surrounding words.

The Stanford study illustrates the development of foundation models by reference to two trends in AI: emergence and homogenization [10].

Emergence has to do with the nature of machine learning. While traditional, rule-based AI systems generate outputs based on pre-defined rules programmed by humans, in machine learning the relevant rules "emerge" as the system induces them from data. In deep learning, multi-layered model fed with large amount of data permits the progressive emergence of higher-level features. Finally, in foundation models we witness the emergence of functionalities, such as in-context learning, which allows a large language model to easily adapt to a downstream task for which it was not specifically trained. For example, GPT-3 can be adapted simply by giving the machine a "prompt" which takes form of instructions in natural language [12].

Homogenization is a trend which leads to building machine learning technologies which can be utilized for a wide range of tasks. The rise of homogenization was accelerated by the developments in deep learning, as the same deep neural networks architecture can be used for various applications. Foundation models go a step further, homogenizing the model itself. For example, the majority of modern cutting-edge NLP models are adapted from foundation models like BERT [10]. There is a lot of evidence suggesting that homogenization is also progressing across different machine learning disciplines – the same tools are used to work on text, images, speech, tabular data or protein sequences [10].

Homogenization of foundation models increases cost-effectiveness of AI systems and provides opportunities for domains in which insufficient data is an obstacle for training task-specific algorithms. At the same time, homogenization of models can lead to a single point of failure, as flaws in foundation model cause flaws in all of its downstream applications. This potential problem is exacerbated by the phenomenon of emergence of new functionalities, as capabilities and risks of foundation models are not fully explored.

3 Europe's New Strategy for Big Data in Health – A Fertile Ground for Foundation Models?

Data is the lifeblood of all algorithms, foundation models in particular. According to the European Commission, "big data in health encompasses high volume, high diversity biological, clinical, environmental, and lifestyle information collected from single individuals to large cohorts, in relation to their health and wellness status, at one or several time points" [5]. These data can come from a variety of sources: healthcare providers, insurers, universities, NGOs, governments, pharmaceutical R&D, Internet of Medical Things and scientific publications. Moreover, it is not uncommon for tech developers to build predictive models in healthcare using non-traditional sources of health data, including consumer patterns, social media or fitness apps [32].

Diverse sources of health data make them naturally multi-modal. This means that they encompass various data types, including: structured and unstructured text, such as doctor's notes or electronic health records (EHR); images, such as X-ray scans; videos, such as ultrasounds; time series, such as ECG; and sequences, such as genetic data. While the best outcomes for patients can be achieved by holistically analysing the entirety of available data, ML models in healthcare are predominantly developed separately for each data modality, focusing either or image or non-image data. Recent studies concerning multimodal models in medical prognosis and diagnosis have shown that they generally perform better than their unimodal counterparts [20]. However, their success is dependent on the availability of patient data, which is not always guaranteed in the routine medical practice.

Both in case of unimodal and multimodal models, particular difficulties can arise when task-specific data is limited, leading to inaccurate results. For example, a ML model designed to predict the risk of hospital patients developing pneumonia classified some asthmatic patients as low-risk; this is an obvious mistake, because asthma is a well-known risk factor for pneumonia [13]. The error occurred because the medical records of asthmatic patients were missing from the training dataset, as such patients were often

admitted directly to the intensive care unit. Building models based on incomplete data can lead to marginalization and exclusion of vulnerable groups of patients. For instance, as reported by the Council of Europe Human Rights Commissioner, "incomplete data in the e-health platform, coupled with a lack of in-person interviews, resulted in loss of social benefits for certain persons with disabilities and older persons with disabilities" in Estonia [35]. In spite of efforts to the contrary, certain marginalized communities, such as Roma and Travelers, continue to face obstacles and discrimination in access to healthcare [18, 26, 27]. This causes their underrepresentation in the main sources of European health data, such as EHR, leading to selection bias in algorithms. Moreover, lessons from the COVID-19 pandemic suggest that the European health data lack reliability and interoperability [24]. This halts meaningful cooperation between Member States in the domain of public health, preventing effective response to problems such as cross-border health threats and the ongoing migration crisis.

The abovementioned challenges are addressed by the new European strategy for data [17], that constitutes an important step towards increasing availability of high quality, interoperable health data and fostering data-driven innovation in healthcare. The strategy foresees introduction of new legal instruments governing health data. First, the recently adopted Data Governance Act (DGA) [42] complements the Open Data Directive [25], enabling the reuse of public sector data which cannot be made available as open data because they are subject to the rights of others, including protection of personal data, intellectual property and commercial confidentiality. Health data falls under the scope of the DGA, which establishes mechanisms for sharing of data held by actors in the public health system, facilitating its secondary use, for example for the purpose of medical research and development of healthcare AI. Moreover, the DGA contains provisions on data altruism, enabling citizens and businesses to voluntarily share their data for the benefit of the society. This solution offers a potential to harvest more data about rare diseases and advance personal medicine through AI. Second, the DGA is accompanied by the initiative to develop the European Health Data Space (EHDS), which will promote exchange and access to health data for improving healthcare delivery, research and policy making. The recently published proposal for the Regulation on the EHDS, operationalizes sharing of electronic health data, introducing a system of permits for data reuse and designating responsible bodies in each Member State [45]. Quality concerns are at the heart of the EHDS, which foresees introduction of common requirements for EHR systems and "FAIR-ification" of various modalities of health data, making them "findable, accessible, interoperable and reusable" [16]. Thirdly, the proposal for the Data Act (DA) [44] sets rules allowing consumers and businesses to access and use data generated by the Internet of Things devices, including Internet of Medical Things. The proposal also grants public sector bodies and EU institutions the right to access data held by private enterprises when it is necessary for the public interest, including health emergencies.

Summing up, the EU's bold strategy for health data offers a chance to address the problems of data availability, quality and interoperability, which currently prevent Europe from exploiting the full potential of health data in building safe, accurate and fair AI models for health and biomedicine. In particular, the new European data landscape could offer a fertile ground for development of foundation models, which could be trained

on high quality data coming from the EHDS. The added value of foundation models is that they offer an opportunity to capture the interaction between different modalities of data and, in this way, create a powerful bio-medical knowledge base. Thanks to their high adaptation capabilities foundation models can be fine-tuned to various individual downstream applications in healthcare and biomedicine and adjusted to specific patient demographics. Furthermore, the DA rules facilitating user access to data generated by the use of products and services create a friendly environment for development of interactive properties of foundation models. The Stanford study underlines that foundation models are likely to change the paradigm of user-AI interaction by "allowing us to rapidly prototype and build highly dynamic and generative AI-infused applications" [10]. It is possible that in the future, the borderline between developers and users will get blurred, as the latter would be able to contribute to adjust the model to their own needs and values, for example by making healthcare AI culturally appropriate.

4 Foundation Models in Healthcare and Biomedicine – Opportunities

Various tasks in healthcare could benefit from the use of foundation models, increasing efficacy and decreasing costs. Researchers claim that the most realistic and feasible use of existing foundation models in the health sector involves NLP applications. Foundation models could be a great tool to extract information from free-text data coming from a variety of sources in the medical domain. Thus, they could serve as a useful interface for healthcare providers. For example, foundation models could help to navigate the complexity of EHRs and automate routine tasks in healthcare delivery, such as retrieving relevant information and medical literature, preparing patient documentation, suggesting tests, treatments and discharges or triaging non-critical patients [37, 41, 48]. Moreover, the translation functions of large language models, such as GPT-3, could reduce the language barrier between patients and healthcare professionals [37], which remains a significant obstacle for migrants and refugees for accessing healthcare in Europe [26]. Last but not least, foundation models' ability to generate human-like conversation could facilitate communication with patients by increasing the role of chat bots in preventive care [21]. For instance, PaLM, Google's latest 540 billion parameter language model, shows promising exploratory capabilities, as it is able to interpret and explain complex information [15]. Thus, developments in foundation models could allow to create chatbots that are better suited to answer patient questions and explain relevant medical information in an accessible form [14]. The potential for improving patient interaction is especially big in geriatric medicine. Researchers argue that socially assistive robots, which could benefit from the NLP advancements offered by foundation models, could provide support to the growing population of elderly patients, especially in circumstances such as the COVID-19 pandemic [1, 38].

Biomedicine is another potent field of application for foundation models. The Stanford study points out two qualities that make foundation models particularly helpful in biomedical discovery: strong generative capability and potential to integrate diverse modalities of data [10]. For example, drug discovery typically involves years of experiments which aim to find molecules that bind the pre-identified target, such as proteins or

genes. The generative qualities of foundation models can accelerate the process, resulting in less experiments and better target-molecule combinations. The ability to integrate various scales and sources of data and to transfer knowledge across various data modalities pushes the boundaries of efficacy even further. For instance, researchers have shown how a transformer model trained on natural language can be fine-tuned to predict protein folding [39, 52]. Discovering how the amino acid sequence of a particular protein affects its three-dimensional structure has been a key question in biochemistry, but the progress in "decoding" proteins has been slow, hindering the advancements in drug development. Once a candidate drug is developed, foundation models could also have a role in supporting clinical trials. For example, they could help to identify eligible patients based on their data profiles or design trial protocols [34]. Finally, foundation models could play a role in development of personalized medicine. Again, their ability to analyze multimodal patient data could lead to more accurate treatment recommendations [10].

5 Foundational Models and Bias Risks in Healthcare

The enthusiasm about the innovations offered by foundation models should be counterbalanced by the awareness of their potential dangers. Since foundation models are still at the early stage of development, many of their pitfalls remain unexplored, making them unsuitable for immediate deployment in high-stakes areas such as healthcare and biotechnology. To start with, the output generated by the state of art large language models is largely unpredictable. For example, in 2020, shortly after the introduction of GPT-3, a short experiment conducted by doctors and ML engineers has shown that the model's performance is not satisfactory even on relatively simple administrative tasks in healthcare [50]. While GPT-3 was able to help patient book an appointment or quickly identify relevant information in an insurance document, it could not perform simple operations on available information, such as adding fees for two medical examinations to calculate the total co-payment sum. The performance was also unsatisfactory when it came to medical diagnosis or treatment recommendation, as the outputs were often erroneous and oversimplified. Moreover, GPT-3 profoundly failed as mental health counsellor, suggesting recycling as a way to increase happiness or advising a suicidal patient to kill himself.

Apart from the central issue of accuracy, one of the biggest dangers of foundation models in healthcare is their propensity to reinforce bias. Although the sources and patterns of algorithmic discrimination in foundation models are not profoundly different from those already studied in the context of machine learning, certain properties of foundation models significantly exacerbate the problem of bias and create new challenges for its detection and mitigation.

Firstly, according to the latest AI Index Report issued by the Stanford University, as the size and capabilities of large language models keep growing, so does the likelihood and severity of bias [55]. For instance, a study of Gopher, a 280 billion parameter language model, indicated that it exhibits higher bias and toxicity levels than its smaller counterparts [46]. In their thought-provoking article Bender and Gebru question the "more data means better data" paradigm, warning against risks associated with the rise of large language models which are trained on the Internet corpora [7]. They point out that

in spite of the presumed inclusiveness of the Internet, text scrapped from the web does not offer an adequate representation of humanity, increasing power imbalances, entrenching dominant viewpoints and reinforcing inequalities. Bender and Gebru underline that the main contributors to the Internet are young males from developed countries. For instance, a survey found that the content of Wikipedia is dominated by Western history and politics while the history of minorities and non-Western cultures is poorly covered [6]. Girls and women constitute only 8,8–15% of contributors worldwide. At the same time, only 18,6% of biographies in English Wikipedia concern women [6]. In addition to underrepresentation, vulnerable communities, such as migrants or sexual minorities, are likely to fall victim of online hate speech and harassment, that are recently becoming a "pan-European" phenomenon [29]. On the other hand, the use of hate speech detecting algorithms to filter the training data can have a side effect of suppressing discourse of marginalized populations, such as LGBTQ people [22]. This problem is acknowledged by the developers of PaLM who argue that removing problematic and stereotypical associations from the model might result in "disproportionately excluding content about or authored by marginalized subgroups in the training data" [15].

The above mentioned problems of underrepresentation, overrepresentation and misrepresentation can cause foundation models trained on Internet corpora to encode various types of bias, such as stereotypes or negative associations for certain groups. For instance, it has been shown that word embeddings, a method used to represent textual data as vectors, can entrench stereotypes about gender roles [9]. Models such as BERT use contextual embeddings, which means that the representation of a word changes depending on its surrounding context. When pre-trained on a large hospital dataset containing EHR of patients admitted to intensive care unit, BERT's embeddings reinforced problematic relationships with regards to different genders, language speakers, ethnicities and insurance groups [54]. For instance, the model suggested to send White and Caucasian patients exhibiting belligerent and violent behavior to hospital, while African, African American and Black patients with the same symptoms to prison. In another study, Abid, Farooqi and Zou show how GPT-3 can capture societal bias against Muslim community, associating it with violence [2]. Typically for a large language model, the bias in sentence completion manifested in creative ways, making it more difficult to detect and mitigate. For instance, when prompted with a description of a girl wearing a headscarf, GPT-3 came up with an incoherent and violent story of the headscarf being drenched in blood. The case of Muslim women is just one example of models such as BERT and GPT-3 entrenching intersectional biases [33]. The risk of targeting intersectional groups is particularly serious in case of healthcare, as studies show that patients in Europe often experience discrimination based on a unique combination of different protected grounds [26, 31].

The second reason for foundation models' susceptibility to bias is associated with multimodality. Researchers point out that fusion of different modalities does not always lead to increased performance, because additional information can introduce inaccurate and unfair results [20]. For instance, the study of CLIP, which is a language-vision model, demonstrated significant reinforcement of historic race and gender bias. Fed with the image of a NASA astronaut Eileen Collins, the model chose "housewife in an orange jumpsuit" to be more appropriate caption than "an astronaut" [8]. CLIP was

also found to misclassify images of Black people as "non-human" [55]. One of the crucial problems with multimodal bias is the lack of well-established metrics for its measurements. Without appropriate bias detection techniques the use of models which fuse image and non-image data remains risky, especially in high-stake areas such as healthcare.

Lastly, as the Stanford study argues, addressing the problem of bias in foundation models requires careful analysis of the relationship between intrinsic biases and extrinsic harms [10]. The former pertain to properties of the foundation model, such as aforementioned underrepresentation, overrepresentation or misrepresentation of certain groups in the training dataset. Thus, once a foundation model exhibits intrinsic bias, its downstream applications are likely to become "fruits of the poisonous tree". On the other hand, extrinsic harm occurs during the process of adaptation itself, for example in case of performance disparities. In the future, developing successful approaches to debiasing foundation models will require detailed understanding of where in the model's ecosystem did the harm occur. Thus, allocation of responsibility between the providers of the foundation model and the developers of its concrete applications is likely be constitute a challenge. This creates additional burdens for claimants pursuing an algorithmic discrimination complaint, contributing to the phenomenon of underreporting, which is prevalent among patients who experience unfair treatment by healthcare professionals [31].

6 Foundation Models and the Artificial Intelligence Act

In light of the potential benefits and risks of foundation models, the EU must ensure that the forthcoming regulatory framework achieves the right balance between fostering innovation and protecting fundamental rights. This is particularly important in the area of healthcare and biotechnology; as the EU continues to struggle with health inequalities, the law must guarantee that the cutting edge technologies benefit rather than harm the health of marginalized communities. Unfortunately, the Artificial Intelligence Act Proposal does not offer adequate protection against the possible harms of foundation models.

First of all, it remains unclear whether foundation models fall under the scope of the AIA. The risk-based approach proposed by the Act focuses on the intended purpose of the system. Thus, as foundation models are intermediary assets with no specific purpose, it appears that only their downstream applications could be covered by the AIA. However, with no specific mention of foundation models, the Act leaves room for uncertainty. Even if we assumed, in the absence of an explicit declaration to the contrary, that the AIA places duties on developers of foundation models that are the base of systems classified as high-risk, defining these obligations is challenging. The generative capabilities of foundation models do not fit easily into the provider-user paradigm. For instance, downstream developers who build and release a specialized model for clinical text analysis based on a foundation model such as BERT [4], are likely to satisfy the definition of "providers" under the AIA (Art. 3(2)), incurring the associated obligations, including the duty to meet the conformity assessment procedure. In this case, it is unclear whether the developers of the foundation model itself would incur parallel obligations of providers and, if yes, what

would be their scope. In an alternative scenario, when a foundation model is used for a health related task with minimum adaptation, for instance by simply providing prompts in natural language, the entity responsible for the downstream application is likely to be considered a mere "user" under the AIA (Art. 3(4)), incurring only limited obligations. In this case, the developers of a foundation model are likely to be considered providers, which means that they would have to meet the provider's obligations for this specific intended purpose. These examples illustrate how the absence of foundation models in the AIA can lead to legal uncertainty for their developers and arbitrary protection for the affected people.

The amendments proposal submitted by the Slovenian Presidency [19] and the recent draft opinion by the European Parliament [30] attempt to address the aforementioned uncertainties. They introduce foundation models into the AIA, under the name of "general purpose AI system", which is defined as "AI system that is able to perform generally applicable functions such as image or speech recognition, audio or video generation, pattern detection, question answering, and translation, and is able to have multiple intended purposes." Both amendment proposals openly exclude general purpose models from the scope of AIA, placing compliance obligations on the developers of their downstream applications. According to the draft opinion by the European Parliament, the only obligation incurred by the provider of foundation model should be to provide the developer who adapts the model with "all essential, relevant and reasonably expected information that is necessary to comply with the obligations set out in this Regulation." The Slovenian presidency proposal is even more limited in this regard, mentioning the duties of foundation model provider only in the recital – the provider of general purpose model "should cooperate, as appropriate, with providers and users to enable their compliance with the relevant obligations under this Regulation and with the competent authorities established under this Regulation."

Of course, many of the downstream applications of foundation models in healthcare would qualify as high-risk under the AIA. This includes AI systems intended to be used as safety component of products that are subject to third party ex-ante conformity assessment under the Medical Devices Regulation 745/2017 (MDR) [49], or are themselves products subject to third party ex-ante conformity assessment under the MDR (Art. 6(1) and Annex II AIA). Thus, in order to be considered as high risk under the AIA, systems should be classified as moderate to high risk under the MDR. Medical devices that are classified as low risk under the MDR are not subject to third-party ex-ante assessment, and thus do not fall into the scope of high-risk systems under the AIA. According to Rule 11 of Annex VIII to the MDR, software is generally classified as low risk, unless it is used for medical diagnosis and therapy or to monitor physiological processes. Thus, although most of the medical software is likely to be classified as moderate to high risk, some possible downstream application of foundation models in healthcare, such as health apps which focus on prevention of diseases, could be assessed as low risk under the MDR, and thus excluded from the scope of high risk systems in the AIA. The same is true for systems which are unlikely to be classified as medical devices at all, for example a model matching patients to clinical trials. At the same time, it is worth emphasizing that the list of stand-alone systems classified as high risk according to Annex III of the AI Act, does not explicitly mention healthcare or biomedicine. It includes systems

providing "access and enjoyment of essential private and public services and benefits," that is systems used to evaluate the eligibility for public assistance benefits and services and systems used to dispatch emergency medical aid (Art. 5(a) and 5(c) of Annex III). In spite of their possible effect on the exercise of the right to health, certain systems, such as the aforementioned model matching patients to clinical trials or a model used to schedule medical appointments, are likely to escape the high-risk category.

Moreover, even if one argues that the substantial part of downstream applications of foundation models in healthcare and biomedicine are classified as high-risk, the potential exclusion of foundation models from the scope AIA remains deeply problematic from the point of view of algorithmic discrimination. It means that general purpose models do not need to comply with any criteria for high-risk systems, including the quality criteria for training, validation and testing of data sets (Art. 10 AIA) and appropriate bias mitigation measures implemented throughout their life cycle (Art. 15(3) AIA). Regulating only on the downstream applications of a foundation model allows to address the extrinsic harms, that arise in the process of adaptation, but not the intrinsic bias, that stems from the properties of a foundation model itself. This is very alarming given the overwhelming evidence of bias in general purpose models. The discussed AIA amendment proposals limit the obligations of foundation models providers to a hopelessly vague duty of cooperation, putting undue burden on the developers of the downstream applications. Placing the duty of compliance with conformity assessment procedure exclusively on the developers of downstream applications of foundation models can have a chilling effect on innovation, because micro, small and medium enterprises might lack sufficient tools and resources to meet the legal requirements [40]. In particular, it might be technically impossible for the downstream developers to effectively detect and remedy bias in general purpose models, preventing their safe and non-discriminatory deployment in healthcare.

I argue that in order to foster safe innovation in healthcare and beyond, the scope of the AIA should cover foundation models. Given that general purpose models are an emergent technology with largely unexplored harms, they should be directly classified as high-risk, if at least one of their potential intended purposes meets the definition of a high-risk system. Thus, downstream applications of a model, that are classified as high-risk, should trigger responsibilities for both the providers of foundation model and the adapting developers. To this end, there is a need to introduce sui generis rules concerning foundation models, because the existing proposal does not foresee multiple providers and establishes provider's obligations focusing on the intended purpose of a system. Therefore, the AIA should establish a clear framework of shared responsibilities between the developers of foundation models and developers of their downstream applications. Within that, appropriate bias detection and mitigation obligations must be specified.

7 Conclusion

The rise of foundation models, which are trained on a gargantuan amount of data and adapted to specific downstream applications, opens new opportunities for healthcare and biomedicine. Their generative capability and ability to integrate multimodal data make foundation models suitable candidates to advance medical prognosis, diagnosis and treatment recommendations and alleviate administrative burdens in administration

of healthcare. In biomedicine, general purpose models can facilitate clinical trials, drug discovery and the development of personalized medicine.

However, evidence suggests that foundation models can also exhibit discriminatory properties, perpetuating and exacerbating bias. The use of Internet corpora as the training set, multimodality and the complex interaction between intrinsic properties and extrinsic harms make bias in foundation models particularly difficult to detect and remedy. This is a huge challenge for the deployment of these models in healthcare. In light of the persisting inequalities in access to and quality of healthcare within the European Union, the regulation of healthcare AI should foster applications which serve marginalized patient communities and protect them from bias and discrimination.

While the new European strategy on data offers a fertile ground for development of foundation models in healthcare by increasing the availability of high quality, interoperable healthcare data, the AIA fails to protect patients against discrimination in foundation models. The current uncertainty about the inclusion of general purpose models under the scope of the AIA, and their explicit exclusion proposed by recent amendments, endanger the fundamental rights that the AIA pledges to protect. Thus, this paper has argued for the introduction of a regulatory framework for foundation models within the AIA, establishing a clear division of responsibilities between the developers of foundation models and developers of their downstream applications.

References

1. Abdi, J., et al.: Scoping review on the use of socially assistive robot technology in elderly care. BMJ Open **8**(2), e018815 (2018). https://doi.org/10.1136/bmjopen-2017-018815
2. Abid, A., Farooqi, M., Zou, J.: Persistent anti-Muslim bias in large language models. arXiv: 2101.05783. arXiv (2021).https://doi.org/10.48550/arXiv.2101.05783
3. Alpaydin, E.: Machine Learning, Revised and Updated Edition. The MIT Press (The MIT Press Essential Knowledge Series), Cambridge, Massachusetts (2021)
4. Alsentzer, E., et al.: Publicly available clinical BERT embeddings. arXiv:1904.03323. arXiv (2019). https://doi.org/10.48550/arXiv.1904.03323
5. Auffray, C., et al.: Making sense of big data in health research: towards an EU action plan. Genome Med. **8**(1), 71 (2016). https://doi.org/10.1186/s13073-016-0323-y
6. Barera, M.: Mind the gap: addressing structural equity and inclusion on Wikipedia. http://hdl.handle.net/10106/29572 (2020). Accessed 20 Mar 2022
7. Bender, E.M., et al.: On the dangers of stochastic parrots: can language models be too big?. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 610–623. Association for Computing Machinery (FAccT '21), New York, NY, USA (2021). doi:https://doi.org/10.1145/3442188.3445922
8. Birhane, A., Prabhu, V.U., Kahembwe, E.: Multimodal datasets: misogyny, pornography, and malignant stereotypes. arXiv:2110.01963. arXiv (2021). https://doi.org/10.48550/arXiv.2110.01963
9. Bolukbasi, T., et al.: 'Man is to computer programmer as woman is to homemaker? debiasing word embeddings In: Lee, D., et al. (eds.) Advances in Neural Information Processing Systems', Vol. 29. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf (2016). Accessed 20 Mar 2022
10. Bommasani, R., et al.: On the opportunities and risks of foundation models. arXiv:2108.07258. arXiv (2021). https://doi.org/10.48550/arXiv.2108.07258

11. Broadband Commission for Sustainable Development: Working group on digital and AI in health: reimagining global health through artificial intelligence: the roadmap to AI maturity (Sep 2020)
12. Brown, T.B., et al.: Language models are few-shot learners. arXiv:2005.14165. arXiv (2020). https://doi.org/10.48550/arXiv.2005.14165
13. Caruana, R., et al.: Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1721–1730. Association for Computing Machinery (KDD '15), New York, NY, USA (2015). https://doi.org/10.1145/2783258.2788613
14. Chaix, B., et al.: When chatbots meet patients: one-year prospective study of conversations between patients with breast cancer and a chatbot. JMIR Cancer **5**(1), e12856 (2019). https://doi.org/10.2196/12856
15. Chowdhery, A., et al.: PaLM: scaling language modeling with pathways. arXiv. http://arxiv.org/abs/2204.02311 (2022)
16. Communication from the Commission to the European Parliament and the Council: A European health data space: harnessing the power of health data for people, patients and innovation COM, 196 final (2022)
17. 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, 66 final (2020)
18. Council of Europe: Recommendation Rec: 10 of the committee of ministers to member states on better access to health care for Roma and Travellers in Europe (Adopted by the Committee of Ministers on 12 July 2006 at the 971st meeting of the Ministers' Deputies) (2006)
19. Council of the European Union: Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Slovenian Presidency compromise text. Brussels, 2021/0106(COD) (29 Nov 2021)
20. Cui, C., et al.: Deep multi-modal fusion of image and non-image data in disease diagnosis and prognosis: a review. arXiv:2203.15588. arXiv (2022). https://doi.org/10.48550/arXiv.2203.15588
21. Demner-Fushman, D., Mrabet, Y., Ben Abacha, A.: Consumer health information and question answering: helping consumers find answers to their health-related information needs. J. Am. Med. Inform. Assoc.: JAMIA **27**(2), 194–201 (2020). https://doi.org/10.1093/jamia/ocz152
22. Dias Oliva, T., Antonialli, D.M., Gomes, A.: Fighting hate speech, silencing drag queens? artificial intelligence in content moderation and risks to LGBTQ voices online. Sex. Cult. **25**(2), 700–732 (2020). https://doi.org/10.1007/s12119-020-09790-w
23. Federico, G.: Access to healthcare in the European Union: are EU patients (Effectively) protected against discriminatory practices? In: Rossi, L.S., Casolari, F. (eds.) The Principle of Equality in EU Law, pp. 229–253. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66137-7_8
24. Di Federico, G.: 'Stuck in the middle with you…wondering what it is I should do. Some Considerations on EU's Response to COVID-19' 7 EUROJUS, pp. 60–85 (2020)
25. Directive (EU): 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-use of Public Sector Information OJ L 172, pp. 56–83 (26 Jun 2019)
26. European Union Agency for Fundamental Rights: Inequalities and Multiple Discrimination in Access to and Quality of Healthcare. Publications Office, LU. (2013). https://doi.org/10.2811/17523. Accessed 12 May 2022
27. European Union Agency for Fundamental Rights: Coronavirus Pandemic in the EU: Impact on Roma and Travellers, Bulletin #5, 1 March - 30 June 2020. Publications Office, LU (2020). https://doi.org/10.2811/978921. (Accessed 13 May 2022

28. European Commission: Horizon 2020 Work Programme 2018–2020: Health, Demographic Change and Wellbeing. European Commission Decision C(2020)4029 of (17 Jun 2020)
29. European Commission: Study to Support the Preparation of the European's Commission Initiative to Extend the List of EU Crimes in Article 83 of the Treaty on the Functioning of the EU to Hate Speech and Hate crime. Publications Office of the European Union, Luxembourg (2021)
30. European Parliament: Draft Opinion of the Committee on Legal Affairs for the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs on the proposal for a regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Brussels, 2021/0106(COD) (2 Mar 2022)
31. Equinet: Equality, Diversity and Non-Discrimination in Healthcare: Learning from the Work of Equality Bodies, Brussels (2021)
32. Funahashi, K.: Big data in health care - predicting your future health' 94 southern California law review, pp. 355–390 (2021)
33. Guo, W., Caliskan, A.: Detecting emergent intersectional biases: contextualized word embeddings contain a distribution of human-like biases. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pp. 122–133 (2021). https://doi.org/10.1145/3461702.3462536
34. Harrer, S., et al.: Artificial intelligence for clinical trial design. Trends Pharmacol. Sci. **40**(8), 577–591 (2019). https://doi.org/10.1016/j.tips.2019.05.005
35. High-Level Conference: Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law. (Helsinki 2019) CommDH/Speech (2019)
36. International Coalition of Medicines Regulatory Authorities (ICMRA): Informal Innovation Network Horizon Scanning Assessment Report – Artificial Intelligence (Aug 2021)
37. Korngiebel, D.M., Mooney, S.D.: 'Considering the possibilities and pitfalls of generative pre-trained transformer 3 (GPT-3) in healthcare delivery. NPJ Digit. Med. **4**(1), 1–3 (2021). https://doi.org/10.1038/s41746-021-00464-x
38. Kulpa, E., Rahman, A.T., Vahia, I.V.: Approaches to assessing the impact of robotics in geriatric mental health care: a scoping review. Int. Rev. Psychiatry (Abingdon, England) **33**(4), 424–434 (2021). https://doi.org/10.1080/09540261.2020.1839391
39. Lu, K., et al.: Pretrained transformers as universal computation engines. arXiv:2103.05247. arXiv (2021). https://doi.org/10.48550/arXiv.2103.05247
40. Muller, C., et al.: AIA in-depth #1: objective, scope, definition.https://allai.nl/wp-content/uploads/2022/03/AIA-in-depth-Objective-Scope-and-Definition.pdf (ALLAI 2022). Accessed 20 May 2022
41. Percha, B.: Modern clinical text mining: a guide and review. Ann. Rev. Biomed. Data Sci. **4**, 165–187 (2021). https://doi.org/10.1146/annurev-biodatasci-030421-030931
42. Proposal for a Regulation of the European Parliament and of the Council on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) COM, 767 final (2020)
43. Proposal for the Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence and amending certain Union legislative acts COM, 206 final (2021)
44. Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) COM, 68 final (2022)
45. Proposal for a Regulation of the European Parliament and the Council on the European Health Data Space COM, 197 final (2022)
46. Rae, J.W., et al.: Scaling language models: methods, analysis & insights from training gopher. arXiv:2112.11446. arXiv (2022). https://doi.org/10.48550/arXiv.2112.11446

47. Ramesh, A. et al. (2022) Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125. arXiv. doi:https://doi.org/10.48550/arXiv.2204.06125
48. Rasmy, L., et al.: Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digit. Med. **4**(1), 86 (2021). https://doi.org/10.1038/s41746-021-00455-y
49. Regulation 2017/745 of the European Parliament and the Council of 5 April 2017 on medical devices amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC OJ L 117, pp. 1–175 (5 May 2017)
50. Rousseau, A., Baudelaire, C., Riera, K.: Doctor GPT-3: hype or reality?. Nabla. https://www.nabla.com/blog/gpt-3/ (27 Oct 2020). Accessed 20 Mar 2022
51. Scholz, N., European parliament, and directorate-general for parliamentary research services: Addressing health inequalities in the European Union: concepts, action, state of play: in-depth analysis. https://op.europa.eu/publication/manifestation_identifier/PUB_QA0120125ENN (2020). Accessed 12 May 2022
52. Tsaban, T., et al.: Harnessing protein folding neural networks for peptide–protein docking. Nat. Commun. **13**(1), 176 (2022). https://doi.org/10.1038/s41467-021-27838-9
53. World Health Organization: Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization, Geneva. https://apps.who.int/iris/handle/10665/341996 (2021). Accessed 12 May 2022
54. Zhang, H., et al.: Hurtful words: quantifying biases in clinical contextual word embeddings. arXiv:2003.11515. arXiv (2020).https://doi.org/10.48550/arXiv.2003.11515
55. Zhang D., et al.: The AI Index 2022 Annual Report' AI Index Steering Committee. Stanford Institute for Human-Centered AI (2022)

Semantic Technologies and Legal Issues

Towards a Method for the Formulation of an EGOV Strategy Considering International Rankings

Wagner S. Araujo[1,2(✉)], Delfina Soares[1], and João Carvalho[2]

[1] United Nations University (UNU-EGOV), Guimaraes, Portugal
wagner.s.araujo@unu.edu
[2] Algoritmi Centre, University of Minho, Guimaraes, Portugal

Abstract. Today, 151 of 193 Member States of the United Nations have a digital strategy, including those dedicated to EGOV. Formulating these strategies is complex and generally involves comparisons with other states to establish the country's positioning within the international situation, identify trends, or set up reasonable goals for the strategic plan. In this context, international rankings are a valuable source of information because they provide a multinational landscape on a wide range of relevant aspects and their evolution over time. This article presents the current results of a research project whose main objective is to develop a method to formulate EGOV strategies considering international rankings. This project started with an exploratory study that substantiated its relevance and allowed to define the method features: it should be flexible, easy to use, instructive, co-creative, comprehensive, and effective. Subsequent steps involved applying it to actual cases of EGOV strategy formulation in two iterative processes so far. This article describes the second iteration depicting the current version while applying and evaluating it by formulating the Cabo Verde EGOV strategy. Results also produced outputs that will be used in the next iteration of the research project. It is expected to benefit policymakers and the scientific community acting in the field.

Keywords: E-Government strategy · E-Governance strategy · EGOV strategy formulation method

1 Introduction

Today, 151 of 193 Member States of the United Nations have a digital strategy [1], including those dedicated to the public sector, or EGOV strategies. EGOV is a common concept used by academics and practitioners, consisting of the public sector's use of Information Technology (IT) to improve information and service delivery, encourage citizen participation in the decision-making process, and make government more accountable, transparent, and effective [2]. The association of IT with government or governance processes is not a simple task due to the high complexity of promoting accountable, effective, inclusive, transparent, and trustworthy public services that deliver people-centric outcomes [3]. It is, therefore, a multi and interdisciplinary domain that

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 49–62, 2022.
https://doi.org/10.1007/978-3-031-12673-4_4

requires knowledge from a wide range of areas, including Information Systems, Public Administration, and Political Science.

A key point is how the actors involved in the governance processes will act in the physical and digital worlds [4], which seems to demand a strategic approach [5]. EGOV strategies typically use benchmarking studies [6] in the formulation process stages, such as country diagnosis, context analysis, and set goals [7]. The list of rankings includes the United Nations E-Government Development Index (UN/EGDI) [1], the World Economic Forum Global Competitiveness Report [8], European Commission Digital Economy and Society Index [9], the recently launched World Bank Govtech Maturity Index [10], and others. Unfortunately, a method that systematises the process of EGOV strategy formulation considering International Rankings is still absent. It would benefit public officials in supporting their duties as EGOV strategists and scholars acting in the research field.

This research project aims to develop a method for the formulation of EGOV strategies considering international rankings. This article intends to show the results so far. The project started with an exploratory study that substantiated the relevance of such a method and allowed to define its features. The subsequent step involved the design, application, and evaluation of the method in actual cases. This article describes the results of the second iteration, including the depiction of the current version, its application in the Cabo Verde EGOV strategy formulation, and its evaluation. Previously, another iteration produced the first version, also applied in a real case.

The text is structured as follows: Sect. 2 presents the background and theory involved, contextualising the subject; Sect. 3 describes the research design; Sect. 4 presents the research results; Sect. 5 the discussion and Sect. 6 concludes.

2 Background and Theory

The association of Information Technology (IT) with government and governance is not a simple task, and a strategic approach seems necessary. Strategies are a top-level document that addresses directions, goals, components, principles, and implementation guidelines [11], having as usual parts the strategic vision, policies, programs, projects, and evaluation [12]. It is expected to answer questions like "Where is the country now?", "Where does the country want to be?" and "How does the country get there?" [13]. They are a reality in both developed and developing countries since they embarked on government modernisation activities [12], a movement supported since the first Summit on the Information Society promoted by the United Nations (UN) [14].

The formulation of these strategies should be compatible with EGOV purposes, such as making the government more efficient; improving the public service delivery; enhancing the accountability; and improving the relationship between citizens and businesses within the public sector [2, 12, 15, 16]. This list presents many perspectives and brings a wide range of objectives to pursue, and the formulation of an EGOV strategy is an alternative to handle it [11, 12]. These documents support the management of investments while turns possible an evaluation process through a set of indicators [13]. Regarding this, the assessment of EGOV has proven to be important [17] but complex due to the various perspectives involved, including the difficulty of quantifying qualitative objectives and the respective contexts.

In this context, the resort to the international rankings and their indicators is not rare [18]. International institutions regularly undertake significant studies to produce rankings of countries on a wide range of features, including information technology [19]. However, their use must be accompanied by a systematic study and reflection on this practice's implications, possibilities, and pitfalls. Many of them are built using a mix of indicators, with substantial decision power available to the compiler in choosing what specific indicators to include, selecting weightings, and smoothing over data unavailability [20]. Even ranking producers recognise limitations and alert that each country should decide the level and extent of their ranking use, balancing this practice with national development priorities [3].

This research project started with an exploratory study [7] that confirmed the importance of international rankings among EGOV strategists and public officials. This study was carried out in 2019 and involved semi-structured interviews with Brazilian public officials, including National Secretaries, Directors, a Government CIO, and Senior Advisors working directly with high-level executives. The respondents worked in at least one of the three Brazilian E-GOV strategies formulated since 2015. Results indicated that despite certain constraints, international rankings are relevant in the strategy formulation scenario [7]. It confirmed a previous study that international rankings are considered an important tool for defining national e-governance strategy and policies and program prioritisation, using them to review past efforts or establish new standards [18]. Finally, the study unveiled the expected features of such a method: the method should be flexible, easy to use, instructive, co-creative, comprehensive, and effective.

However, existing EGOV strategy formulation methods and frameworks do not explicitly consider international rankings. Chen [21], for example, summarises the differences between developed and developing countries, identifies critical factors for a successful e-government implementation, and proposes an implementation framework. Rabaiah and Vandijck proposed a generic strategic framework, describing a strategy's essential elements and components [11]. Mkude and Wimmer, by their side, proposed guidelines for the development of e-government systems, comparing many of the existing ones, resulting in a comprehensive strategic framework for the successful design of e-government systems in developing countries [12]. And Janowski presented a four-stage Digital Government Evolution Model comprising Digitisation, Transformation, Engagement, and Contextualisation stages [4].

To substantiate the design of the method was essential to resort to the existing literature in administration and management. The concept of a strategy used in this article comes from this literature: an explicit plan developed consciously and purposefully, made in advance of the specific decisions to which it applies [22]. This plan is also characterised by analytical, formal, and logical processes through which organisations scan the internal and external environment and develop policy options that differ from the status quo [23]. In developing the method, it was also important to differentiate the strategy process construct from the strategy content, with different but complementary definitions. While the strategy process reflects how alternatives and actions are selected [24], strategy content is the outcome of this process [25]. It is particularly relevant for the

project because international rankings embedded information will be used as the strategy content according to the strategy formulation process preconised by the proposed method.

This section briefly presented the theory used to develop the new method, highlighting the results from the previous exploratory study. The following section will describe the research design.

3 Research Design

This section presents the research design. As stated, the project started with an exploratory study that substantiated the relevance of such a method and allowed for defining its features. According to the results, it should be flexible, i.e., adjustable to the country context; easy to use, i.e., simplify the use of international rankings in EGOV strategy formulation; instructive, i.e., support the learning process and the association of rankings characteristics to EGOV purposes; co-creative, i.e., enable the participation of multiple stakeholders; comprehensive, i.e., broadly cover the EGOV purposes; and effective, i.e., deliver an EGOV strategy after a complete formulation process.

Subsequent steps involved applying the method to actual cases of EGOV strategy formulation in an iterative process structured according to the design science research approach. The choice of the approach occurs because it seeks to extend the boundaries of human and organisational capabilities by creating new and innovative artefacts [26]. As the method could be helpful for many countries, it is also convenient because of its generalisable characteristic [27]. The validation followed action research principles, as there was an opportunity to use the method's current version to formulate a real EGOV strategy involving researchers and policymakers. The choice occurred given that this methodology synergistically and holistically associates research and practice [28] and can produce highly relevant research results because it is grounded in practical action to solve an immediate problem situation while carefully informing theory [29]. In practice, the Action Research phases of Diagnosis, Action Planning, Action Taking, Evaluating, and Specifying Learning superpose two Design Science activities: Demonstration and Evaluation. Figure 1 illustrates the final research design.

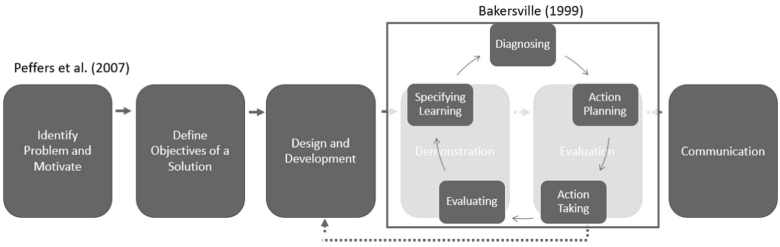

Fig. 1. Research design, adapted from Peffers et al. (2007) and Baskerville (1999

This section briefly presented the Research Design involving a combination of Design Science and Action Research approaches. It also described the exploratory research that

supplied the basis of the study. The following section will present the achieved results according to this methodology.

4 Results

The results will be presented following the combination of Design Science and Action Research approaches, according to the diagram shown in Fig. 1. The problem has been identified as the absence of a method for formulating EGOV strategies considering international rankings. The exploratory study confirmed the subject's relevance and defined its requirements or objectives to pursue. The method should be flexible, easy to use, instructive, co-creative, comprehensive, and effective.

4.1 Design and Developing Through the Design Science Approach

After defining objectives to pursue, the design and development phase included knowledge of theory brought to bear on the solution [30]. Hence, the proposed method has been based on the literature review [11, 12, 31] and on the content analysis of convenient selected official national digital EGOV strategies such as published by Argentina [32], Austria [33], Brazil [34], Chile [35], Mexico [36], Netherlands [37], Portugal [38], Thailand [39] and Turkey [40]. The evaluation of applying a previous version of the method in another actual case was also used to improve its process, composed of five stages: Diagnosis and Context Analysis; Definition of the Vision; Definition of Intervention Areas; Definition of Objectives; and Definition of Structuring Pillars.

The current version is based on two strategy constructs: strategy process [41] and strategy content [23]. While the strategy process reflects how strategic alternatives/actions are selected, strategy content is the outcome of this process or the selected alternatives/actions. The strategy formulation method and the associated content are presented in Fig. 2.

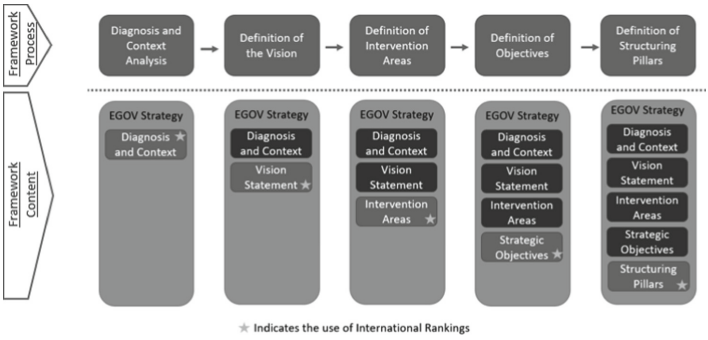

Fig. 2. Method strategy process and strategy content.

The strategy formulation process was constructed based on the previous academic work [11, 12]. Both authors based their work on official published national strategies,

looking for some commonalities. While Rabaiah and Vandijck studied EGOV strategies from 21 countries and the European Union, Mkude and Wimmer analysed nine guidelines for EGOV design. Figures 3 and 4 show how the method steam from the analysed literature.

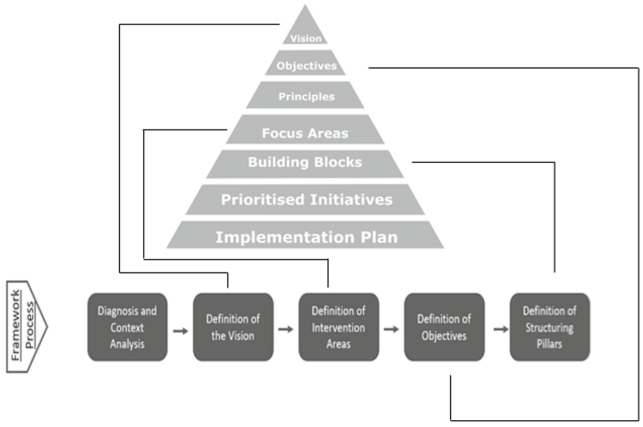

Fig. 3. The method process. Adapted from Rabaiah and Vandijck [11].

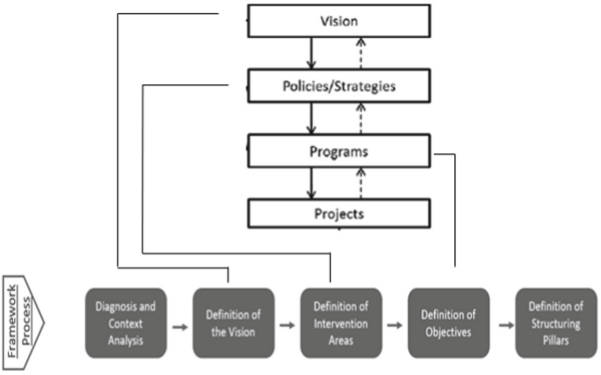

Fig. 4. Stages of the method process derived from Mkude and Wimmer [12].

The first stage, **Diagnosis and Context Analysis**, collects and analyses information about the country's national and international context. It is based on Heeks [13] question, "Where are we now? (p.44;50)". For the national context, inputs were the country's history in EGOV efforts, the existing regulatory framework, former and current government strategies, former and current international partnerships, a SWOT workshop with multiple stakeholders, interviews to collect the opinion of important key actors, inventory of government IT infrastructure, and an inventory on online and offline public services. For the international context, the inputs include a range of EGOV related international rankings and the analysis of countries of reference for Cabo Verde. This stage

produces the strategy content related to the national and the international context: the national context section includes the results of the SWOT analysis; the set of connections between current government strategies and EGOV purposes; and the expectations of future online public services based on the study of the public service inventory. A gap report about the government IT infrastructure; a 10-year analysis of country scores in selected international rankings; and a benchmarking report with countries-of-reference data/information. An important output of this stage is a list of relevant international rankings compatible with the country's context according to the opinion of key actors. The subsequent stages will receive as inputs the outputs produce here, assuming the Diagnosis and Context Analysis stage the role of information source for the next ones.

The **Definition of the Vision** is a stage to establish a single and strong statement that will frame the EGOV strategy content. All content produced in the subsequent stages should collaborate to reach the established vision. The strategic vision definition is also important because it resumes the policy in a single phrase to which the e-government strategists are held accountable [11]. Dimensions and components of international rankings can support the strategic vision construction. For example, if UN/EGDI is one of the selected international rankings, its subindexes, such as online services, telecommunication infrastructure, or human capital, can be used to construct a strategic vision like "A digital country, with state-of-the-art online public services, robust telecommunication infrastructure, and a full and interactive citizenship". The same rationale can be used for other selected international rankings.

The **Definition of Intervention Areas** is a stage that starts to organise the country's ambitions on EGOV, establishing the areas in which interventions are demanded to reach the vision. In the next stage, these "intervention areas" will group in clusters all strategic objectives. These areas stem from sources like the SWOT analysis results, the intersection of government general strategies and EGOV purposes, analysis of the public service inventory, government infrastructure gaps, and good practices from countries of reference. Dimensions and components of international rankings can also be used here. For example, if the country considers the WB/Ease of Doing Business Report a relevant index, one intervention area can be Business Generation. If cybersecurity is an issue, the International Telecommunication Union (ITU) Global Cybersecurity Index [42] can justify the creation of an Information Security Intervention Area. If e-participation is a priority, an Intervention Area destined for the Citizens' Involvement in Decision Making can be considered, as illustrated in Fig. 5.

The **Definition of Objectives** is a stage destined to declare strategic objectives. They can be political, economic, social, cultural, technological, or legal objectives or aspects necessary for implementation, such as change management, public-private partnership, and organisational and business architectures [12]. The analysis of the previous results on rankings can unveil opportunities for enhancement and a list of ranking measurements that fit in the country's context. One example of an objective that can be set in this stage is the construction of an E-Consultation Platform, an item measured by the United Nations E-Participation Index [1]. Another is an online service intended to minimise procedures, time, and costs associated with starting a new business, an item measured by the World Bank Ease of Doing Business Report [43]. Another one can be those

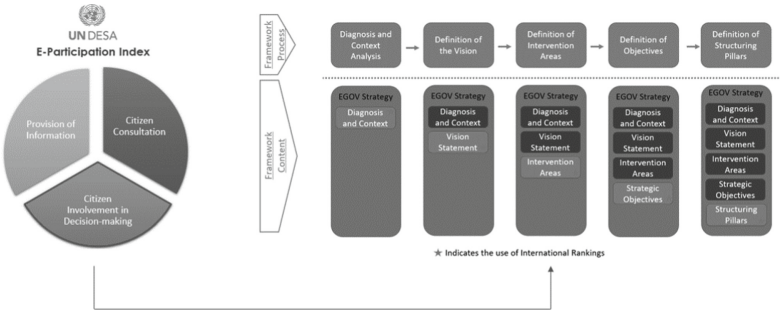

Fig. 5. An intervention area derived from the UN E-participation index dimension.

derived from the list of measurements present in the International Telecommunication Union Global Cybersecurity Index, as demonstrated in Fig. 6.

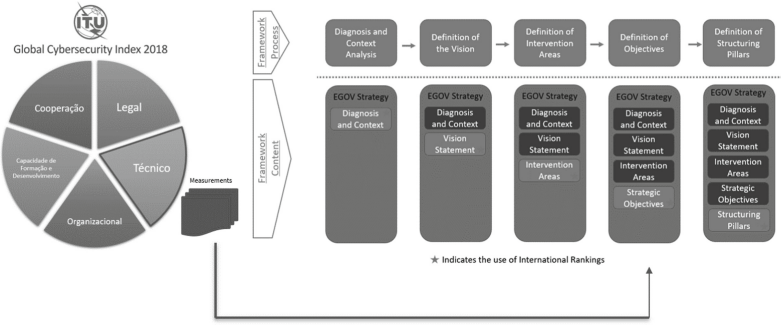

Fig. 6. Objectives derived from ITU global cybersecurity index measurements.

The **Definition of Structuring Pillars** is the stage that defines the necessary building blocks to implement the strategic objectives. Building blocks are components that improve the design of Digital Government systems [31]. They support the EGOV objectives, mostly technological and administrative/organisational and regulatory, defined according to established goals. An example is Digital Identity, an item measured in the United Nations E-Government Development Index [1]. Other examples are a Cybersecurity Government Agency and Cybersecurity Regulation, measured in the International Telecommunication Union Global Cybersecurity Index [44].

This subsection presented the Design and Development activity of the Design Science Approach. According to the Research Design, the Action Research activities used to demonstrate and evaluate the method will be depicted in the next subsection.

4.2 Demonstration and Evaluation Through the Action Research Approach

Action Research activities involved the development of the Cabo Verde EGOV strategy. The country is composed of ten islands in the middle of the Atlantic Ocean on the

western coast of Africa and has been investing heavily in EGOV development since 1998. According to the UN E-Government Survey, the country attained 2020 the High-Level Group in the Electronic Government Development Index.

The process was carried out in a partnership between the Cabo Verde Government and the Academy, which resulted in a team formed by five people, two senior country public officials and three EGOV researchers with a solid academic and practical background. The endeavour has been formalised through a collaboration protocol between the country and the academic institution, which stated a research and practice bridge environment. The government has been represented by the Ministry of Finance and the Ministry of Foreign Affairs. The process involved several countries' public officials and policymakers in 14 ministries and seven public administration entities. It also involved academics and businesspeople, including ten civil society and private sector actors and one from the academia/public university. It run between July and December 2020, ending with the submission to the Council of Ministers after the National Committee for Digital Strategy's approval in February 2021, valid for 2021 to 2023.

The Action Research Diagnosis Phase identified that a National Digital Strategy existed, but only a general reference to the public sector. Its focus was on the digital economy, including industry and business, civil society, and academic and scientific sectors. A National Committee for Digital Strategy has been set with many responsibilities, including approving a plan for the public sector. The Ministry of Finance, especially the "Casa do Cidadão" and the State Secretariat of Administrative Modernisation, identified the necessity to develop a specific digital strategy for the public sector and initiated an inventory of public services and general diagnosis tools. The ongoing methodology was unclear, with many paths and incohesive inputs and outputs through the stages. According to other regulations and instruments, the strategy has different names, signalling inconsistencies. Despite this scenario, it was identified that the country has an impressive background in EGOV and a motivated government team, aware of its potential and full of international references in the field. A group of focal points has already been stated, formed by representatives of leading agencies. The Information Society National Agency - NOSI was a reference inside and outside the country, mainly in Africa. International partnerships were vital to a country due to its dependency on donors' resources. EGOV development was heterogeneous through government agencies, for example, strong in the Finance Ministry but weak in the Health Ministry. As the country has a large population living abroad, a team working at the Embassy in Portugal developed a digital platform dedicated to the diaspora. This solo project should be included in the future strategy.

The Action Research Planning Phase defined the project target as the formal approval of the National EGOV strategy, defining intermediary steps to reach this objective. The base theory used for the method construction was made available to the action research team. The plan establishes the target for change and the approach to change [29], i.e., the formulation of the Cabo Verde EGOV strategy using the method for the formulation of EGOV strategies considering international rankings. The work period coincided with the COVID-19 pandemic, and most activities occurred online using the Microsoft Teams collaboration tool. Weekly meetings were set, but the collaboration tool allowed an efficient virtual office, where contact between team members occurred

by convenience using chat, file transfer, audio, and videoconferences. Activities on the ground like recorded interviews and the SWOT workshop have been planned to be conducted with local support by the country's public officials, who have been trained previously to achieve them.

The Action Taking phase was recorded in research notebooks as it progressed, and many meetings have been recorded using the collaboration tool. It occurred as planned, with a bit of distortion in the usual schedule due to restrictions inherent to the elections period in 2021 beginning. The former running process has been modified to method one, already depicted in Fig. 2. The **Diagnosis and Context Analysis** phase collected a comprehensive source of information, including official documents, the regulatory framework, country-of-reference benchmarks (Luxemburg, Mauritius, and Seychelles), IT infrastructure and public service inventories, questionnaires, recorded interviews, a SWOT workshop, and a set of international rankings selected according to the country context. This set included the United Nations E-Government Development Index (UN/EGDI), the World Bank Ease of Doing Business Report (WB/EDB), the International Telecommunication Union Global Cybersecurity Index (ITU/GCI), and the World Economic Forum Global Competitiveness Report.

After analysing information produced in the previous stage, the **Strategic Vision** has been defined as "A digital Cabo Verde, an agile, capable, resilient, reliable and transparent State, closer to citizens and business in all life events.". Nine **Intervention Areas** have been defined to reach the strategic vision, based on the Diagnosis and Context Analysis outputs: Digital Public Services; Digitalization and Simplification of Administrative Procedures; Access and Availability of Information; Integration and Interoperability; Infrastructure and Security; Technology; Legislation; Human Resources Literacy; and Governance. Some of these areas steamed from the rankings like ITU/GCI (Infrastructure and Security), UN/EGDI (Digital Public Services), and WB/EDB (Simplification of Administrative Procedures). From the previous phases, 80 **Objectives** have been defined to reach the established vision, each associated with the nine intervention areas. Finally, 56 **Structuring Pillars** have been designated as building blocks to offer structures to the objectives, classified into three categories: Administrative, Legislative, and Technological. International rankings measurements inspired many objectives and a range of structuring pillars. The National Committee of Digital Strategy delivered and approved the Cabo Verde EGOV Strategy for 2021 to 2023.

The Evaluating phase allows the conclusion that the objective was reached as agreed. Semi-structured interviews allowed us to conclude that the intervention produced the expected results for the country, a strategy formally approved.

There was important feedback for the method evaluation. The interviews revealed that the method is flexible because the Diagnosis and Context Analysis stage results in a context-oriented output. The remaining stages consistently use them as inputs. It is also easy to use because it supports rankings features use, facilitating their correlation with intervention areas, objectives, and structuring pillars. It can be considered instructive because it guides the use of international rankings content along the EGOV strategy formulation process. According to one of the interviewees, although having some knowledge about rankings, now he can critically understand their components, dimensions, and measurements. It is comprehensive because they are embedded during the process

execution from the first stage. It is co-creative because it occurs mainly in the Diagnosis and Context Analysis stage. Finally, it is effective because the team delivered the Cabo Verde EGOV strategy after a complete formulation process. Interviews also unveiled that the method brought cohesiveness and consistency to the process, offering future research and development directions. Critics were related to the absence of public consultation, which was unpracticable due to election period restrictions. A missing Action Plan was pointed out despite not existing in the previous project scope. Designing an Action Plan as a separate process is adequate because it is challenging to define resource allocation and mature deadlines during the strategy formulation.

This section presented the study results. The following areas will show the discussion and the conclusion of the article.

5 Discussion

Rankings, despite some restrictions, are part of the context involving public officials and policymakers during the EGOV strategies formulation. Unfortunately, this process is usually conducted *ad hoc*, without a method to support it. International rankings aren't constructed to formulate EGOV strategies but to another end, although it doesn't refuse the fact that many countries use them in this context. Ranking producers are aware of this use and alert that each country should decide the level and extent of their use. This scenario aligns with the previous exploratory study that is part of this research project. Besides confirming the utility of a method to support the EGOV strategy formulation considering international rankings, it highlighted the different perspectives of the involved stakeholders. Rankings are important for policymakers and politicians, but the importance differs because political and technical objectives differ. Both are important within the natural governance processes, but this reinforces that a systematic method is welcome.

It was the case in Cabo Verde. As a developing state is dependent on international funding, it was important to be capable of interpreting rankings information, the embedded academic and practical references on them, and their continuing evolution and upgrades over the years. Countries are naturally dependent on international reputation, and rankings are a natural source of comparable information regarding economic size, human development, ease of doing business, and so on. EGOV is not different. This context was important to Cabo Verde redesigning their EGOV Strategy formulation process to obtain the new method's advantage. Its flexibility was essential because the field is multidisciplinary, and the country's attributes are vast and diverse. The new method uses many information sources and consistently allocates them along with the formulation of strategy content. Its pre-defined process enables the work of a myriad of trackable data and information as used in each phase through the control of inputs and outputs. It turns virtually possible to use many sources, not only those from international rankings, supporting the correct allocation of data and information along the process, effectively producing EGOV strategy content. Ranking features also allow policymakers to learn about their characteristics, including components, dimensions, and measurements, due to the method's ease of use and instructive feature. As EGOV encompasses many purposes, the method helped identify intersections between international rankings and the

country's general strategies and government programs. The multiple stakeholders' participation was also supported, assuring significant representation of society in the strategy formulation. Finally, the method's effectiveness was assured through the delivery and final approval of the strategy document. This approval occurred with the participation and support of key-actors responsible for the technical and political endorsement of the document. Without its support, this information would probably be unsystematically used and poorly used in the strategy content, such as the choice of intervention areas, the proposition of objectives, and the definition of structuring pillars.

6 Conclusion

Despite many countries' uses of international rankings in the formulation of EGOV strategies, this process generally occurs without method. This article presented the current results of a research project whose main objective is to develop this method, designed to be flexible, easy to use, instructive, co-creative, comprehensive, and effective. Subsequent steps involved applying it to actual cases of EGOV strategy formulation in two iterative processes so far. The article depicted the current version of the method and its application and evaluation by formulating the Cabo Verde EGOV Strategy for 2021 to 2023. The results include outputs to be used in a new iteration of the same research project. The development occurred under a design science approach. As there was an opportunity to use and evaluate the method in a real case, an action research approach complemented the research design. The research project already presents results that benefit policymakers and scholars acting in the field. It occurs mainly because the method has been applied in two real cases of EGOV strategy formulation. Nonetheless, the research project still presents limitations. There isn't evidence of a stopping point for the design and development yet, which will demand further iterations, maybe evolving the formulation process and the associated strategy content.

References

1. Department of Economic and Social Affairs: e-Government survey 2020 - digital government in the decade of action for sustainable development: with addendum on COVID-19 response (2020)
2. UNESCO: UNESCO thesaurus. Unesco Thesaurus. [Online]. Available: http://vocabularies.unesco.org/browser/thesaurus/en/page/?uri=http://vocabularies.unesco.org/thesaurus/concept17004 (2019). Accessed 19 Oct 2019
3. UNITED NATIONS: e-Government survey 2018: gearing e-Government to support transformation towards sustainable and resilient societies (2018)
4. Janowski, T.: Digital government evolution: From transformation to contextualization. Gov. Inf. Q. **32**(3), 221–236 (2015)
5. Henderson, J.C., Venkatraman, H.: Strategic alignment: leveraging information technology for transforming organizations. IBM Syst. J. **38**(2.3), 472–484 (1999)
6. Heeks, R.: Understanding and measuring e-Government: international benchmarking studies. UNDESA Work. E-Participation e-Government Underst. Present Creat. Futur. Budapest, no. July, pp. 27–28 (2006)

7. Araujo, W., Soares, D., Carvalho, J.: International rankings in E-GOV strategy formulation: appraisal of their use and relevance. In: Proceedings of Ongoing Research, Practitioners, Posters, Workshops, and Projects of the International Conference EGOV-CeDEM-ePart 2021, vol. 3049, pp. 163–175 (2021)
8. WEF, W.E.F.: The global competitiveness report: how countries are performing on the road to recovery (2020)
9. European Commision: Digital economy and society index (DESI) 2020. Hum. Cap. (2020)
10. Dener, C., Nii-Aponsah, H., Ghunney, L.E., Johns, K.D.: GovTech Maturity Index (2021)
11. Rabaiah, A., Vandijck, E.: A strategic framework of e-Government: generic and best Practice. Electron. J. e-Government **7**(3), 241–258 (2009)
12. Mkude, C.G., Wimmer, M.A.: Strategic framework for designing e-Government in developing countries. In: Wimmer, M.A., Janssen, M., Scholl, H.J. (eds.) EGOV 2013. LNCS, vol. 8074, pp. 148–162. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40358-3_13
13. Heeks, R.: Implementing and Managing e-Government, vol. 23, no. 1. SAGE Publications Ltd, London (2006)
14. Almazan, R.S.: Building Digital Government Strategies: Principles and Practices. Springer Science+Business Media, New York, NY (2017)
15. Alarabiat, A., Soares, D., Ferreira, L., De Sá-Soares, F.: Analyzing e-Governance assessment initiatives: an exploratory study. In: Proc. 19th Int. Conf. Digit. Gov. Soc. Delft, Netherlands, no. ii, pp. 10 (2018)
16. Ojo, A., Janowski, T.: Integrating electronic government and public administration reform strategies: method and experience. In" Global Strategy and Practice of e-Governance: Examples from Around the World (2011)
17. Ogutu, J.O., Irungu, J.K.: Citizen-centric evaluation framework for e-Government systems in Kenya. The case of public service commission of Kenya Online Recruitment amp; Selection system. In: IST-Africa Conf. Exhib. (IST-Africa), 2013, pp. 1–8 (2013)
18. Soares, D., Araujo, W., Carvalho, J.: Electronic governance policy measurement, assessment and monitoring: a discussion. In: ICT Electron. Gov. 2017 Surv. use Inf. Commun. Technol. Brazilian public Sect., pp. 173–182 (2018)
19. Rorissa, A., Demissie, D., Pardo, T.: Benchmarking e-Government: a comparison of frameworks for computing e-Government index and ranking. Gov. Inf. Q. **28**(3), 354–362 (2011)
20. Davis, K.E., Kingsbury, B., Merry, S.E.: Indicators as a technology of global governance. Law Soc. Rev. **46**(1), 71–104 (2012)
21. Chen, Y.N., Chen, H.M., Huang, W., Ching, R.K.H.: e-Government strategies in developed and developing countries: an implementation framework and case study. J. Glob. Inf. Manag. **14**(1), 23–46 (2006)
22. Mintzberg, H.: Patterns in strategy formation. Manage. Sci. **24**(9), 934–948 (1978)
23. Andrews, R., Boyne, G.A., Law, J., Walker, R.M.: Strategy formulation, strategy content and performance: an empirical analysis. Public Manag. Rev. **11**(1), 1–22 (2009)
24. Hart, S L.: An integrative framework for strategy-making processes author(s): Stuart L . Hart Source : The Academy of Management Review, Vol. 17, No. 2, pp. 327–351 (Apr 1992). Published by : Academy of Management Stable URL : https://www.jstor.org/stable. Acad. Manag. Rev., vol. 17, no. 2, pp. 327–351 (1992)
25. Johanson, J.-E.: Strategy Formation and Policy Making in Government. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-03439-9
26. Bichler, M.: Design science in information systems research. Wirtschaftsinformatik **48**(2), 133–135 (2006). https://doi.org/10.1007/s11576-006-0028-8
27. Venable, J., Baskerville, R.: Eating our own cooking: toward a more rigorous design science of research methods. Electron. J. Bus. Res. Methods **10**(2), 141–153 (2012)

28. Avison, D.E., Davison, R.M., Malaurent, J.: Information systems action research: debunking myths and overcoming barriers. Inf. Manag. **55**(2), 177–187 (2018)
29. Baskerville, R.L.: Investigating information systems with action research. Commun. Assoc. Inf. Syst., **2** (1999)
30. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007)
31. Baheer, B.A., Lamas, D., Sousa, S.: A systematic literature review on existing digital government architectures: state-of-the-art, challenges, and prospects. Adm. Sci. **10**(2), 25 (2020)
32. Presidencia de la Nación Argentina: Agenda Digital Argentina. Boletín Oficial de la República Argentina. [Online]. Available: https://www.boletinoficial.gob.ar/detalleAviso/primera/195154/20181105 (2018). Accessed 16 Mar 2021
33. Federal Chancellery and Federal Ministry and R. and E. of Science: Digital Roadmap Austria. [Online]. Available: https://www.digitalroadmap.gv.at/fileadmin/downloads/digital_road_map_broschuere_eng.pdf (2016). Accessed 16 Mar 2021
34. Brasil: Estratégia de Governo Digital 2020–2022 — Português (Brasil). [Online]. Available: https://www.gov.br/governodigital/pt-br/EGD2020 (2020). Accessed 16 Mar 2021
35. Chile: Estrategia de transformación digital. Estrategia de Transformación Digital. [Online]. Available: https://innovadorespublicos.cl/documentation/publication/40/# (2018). Accessed 16 Mar 2021
36. Mexico, G.de L.R. : Estrategia digital nacional de mexico. [Online]. Available: https://www.gob.mx/cms/uploads/attachment/file/17083/Estrategia_Digital_Nacional.pdf (2013). Accessed 16 Mar 2021
37. Netherlands: Dutch Digitalisation Strategy. Report, Government.nl (2018)
38. Governo de portugal: plano de ação para a transição digital de portugal. República Portuguesa: Economia e Transição Digital. [Online]. Available: https://www.portugal.gov.pt/gc22/portugal-digital/plano-de-acao-para-a-transicao-digital-pdf.aspx 2020. Accessed 16 Mar 2021
39. EGA: Thailand digital government development plan 2017–2021 (Draft) (2017)
40. Turkey: National e-Government strategy and action plan. [Online]. Available: https://bilgem.tubitak.gov.tr/en/urunler/2016-2019-national-e-government-strategy-and-action-plan 2016. Accessed 16 Mar 2021
41. Mintzberg, H., Quinn, J.B.: The Strategy Process: Concepts, Contexts, Cases. Prentice-Hall International, London (1996)
42. International Telecommunications Union: Measuring the information society report 2018 (2018)
43. The World Bank Group: Ease of doing business score and ease of doing business ranking, pp. 77–86 (2020)
44. ITU: Global cybersecurity index 2020 - measuring commitment to cybersecurity (2020)

Semantic Annotation of Parliamentary Debates and Legislative Intelligence Enhancing Citizen Experience

Stéphane Gagnon[1(✉)] and Sabrina Azzi[2]

[1] Université du Québec en Outaouais, Gatineau, Canada
stephane.gagnon@uqo.ca
[2] House of Commons of Canada, Ottawa, Canada
sabrina.azzi@parl.gc.ca

Abstract. The concept of "Legislative Intelligence" (LegisIntel) refers to Artificial Intelligence (AI) and semantic analytics tools implemented in parliaments to enhance citizen experience in monitoring complex interrelations among various contents of parliamentary proceedings. The integration of a suite of digital solutions can build upon the core functionality of Semantic Annotation of Parliamentary Debates. Using well-established Natural Language Processing (NLP) technologies, linked to ontologies and Knowledge Graphs (KG), it can help identify the concepts and entities throughout texts, and index sentences and summaries as per a citizen's customized knowledge base. These annotations can then be leveraged to recommend relevant text excerpts end-users could build upon, within teams if they chose to do so, and possibly compose and customize legislative critiques and recommendations thoroughly tested for coherence, accuracy, and evidence. The present study proposes an international open-source initiative among parliaments to ensure the launch and viability of a suite of LegisIntel solutions. It reports on the completed initial phase of this initiative, aiming to prepare discussions in launching an international consultation among peers. The goals of this phase are to document the core functionality of LegisIntel solutions and formulate a proposed architecture that may serve to generate ideas from various developer communities. The Action Design Research (ADR) methodology is used in this process, with results focused on system artefacts such as an interface mockup, a functional design, and a model of infrastructure components. The conclusion addresses risks and outlines the next steps of this initiative.

Keywords: Parliamentary debate · Semantic annotation · Legislative intelligence

1 Introduction

The implementation of digital democracy requires solutions that go beyond traditional information architectures. The end-user experience of citizens and a variety of parliamentary stakeholders involves more than traditional "download and reuse" dissemination

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 63–76, 2022.
https://doi.org/10.1007/978-3-031-12673-4_5

policies that have underlined most of the Linked Open Data (LOD) initiatives. Among key factors causing complexity, digital solutions must take in account the dynamic nature of parliaments, the importance of semantics in interpreting contents, and the ever-diversifying knowledge domains addressed in legislative proceedings.

As an extension to these concepts, the concept "Legislative Intelligence" (LegisIntel) refers to Artificial Intelligence (AI) and semantic analytics tools implemented in parliaments to enhance citizen experience. These technologies can improve information search and use, automate certain tasks in analyzing complex conceptual interrelations, and ensure the relevance and consistency between continuously updated debates, laws, regulations, reports, and testimonies.

The integration of a suite of LegisIntel solutions can build upon the core functionality of Semantic Annotation of Parliamentary Debates. Its purpose is to identify the various concepts and entities throughout texts and index the key sentences and summaries through a citizen's customized knowledge base. These annotations can then be used to recommend relevant text excerpts end-users could build upon, within teams if they chose to do so, and possibly compose and customize legislative critiques and recommendations thoroughly tested for coherence, accuracy, and evidence.

Well-established technologies provide building blocks for this approach. Natural Language Processing (NLP) can help integrate heterogenous sources for legislative analysis. Standard components include Named Entity Recognition (NER), Part of Speech (POS) tagging, Recognizing Textual Entailment (RTE), and argument mining. These can be integrated with ontologies and Knowledge Graph (KG) platforms for further processing.

This environment is yet to be developed and represents a groundbreaking research endeavor. Therefore, following a partnership between Université du Québec and the House of Commons of Canada, an international open-source initiative among parliaments is proposed to ensure the launch and viability of a suite of LegisIntel solutions. Its development will be hosted on GitLab, while opening the opportunity for groups of parliaments to form end-user collaboration teams to develop specific functionality (https://gitlab.com/BTMprof/LegisIntel).

The present study reports on the initial phase of this initiative, focused on preparing the discussion to launch an international consultation among peers in parliaments. The goals of this phase are to document the core functionality of LegisIntel solutions and formulate a proposed architecture that may serve to generate ideas from various developer communities. The Action Design Research (ADR) methodology is used in this process, with results focused on system artefacts such as an interface mockup, a functional design, infrastructure components models, and a logical design for software implementation.

The next section outlines a literature review of how LegisIntel solutions have been implemented in parliaments. From this review, a set of objectives and requirements are formulated to guide this initial phase of the initiative. The ADR methodology is then summarized, describing the process followed in analysis and design. Results are presented with several design artefacts that can help to identify gaps and opportunities for further research and development. The conclusion helps to identify risk areas of this initiative and set an agenda for further discussion in the community.

2 Literature

Several parliaments around the world have recently experimented with innovative Legis-Intel solutions. An important dimension for classifying studies is the potential value they offer to end-users, especially to enhance citizen experience in political participation. Some recent studies of these parliamentary semantic applications are classified in Table 1. Among key technologies, there is increasing focus on NLP, semantic matching using ontologies, semantic rules, and KG datasets stored in NoSQL and Big Data databases.

Table 1. Applications and parliamentary end-user concerns.

#	End-user concern	Technologies	References
1	Accessibility and integration	LOD curation process, schemas, ontologies, and graph query	[1–8]
2	Readability and visualization	End-user interfaces, information synthesis, navigation	[9–15]
3	Usability and customization	NLP annotation, graph search, knowledge extraction	[16–24]

The generic end-user concerns of accessibility, readability, and usability must be understood beyond their traditional application in User Experience (UX) design. As such, the present study focuses specifically on UX for information search and use behavior that can be fulfilled by further automated NLP and semantic technologies.

Several notable research efforts have made advances to enhance citizen experience through semantic annotation. In Finland, a team has developed an extensive LOD curation process to annotate debates [25]. As part of the CLARIN project [26], this has led to a systematic integration of internal and external graphs, allowing to query parliamentary speeches through various semantic features.

In Chile, a focus on legal text annotation, using the UN Akoma-Ntoso XSD, have allowed for complex LOD integration. Among other advantages, this has enabled monitoring public finance voting behavior [27–29].

In Spain, teams have focused on annotating and integrating diverse multimedia resources and ensuring their coherent retrieval [17, 19, 30]. Annotation was linked to efforts to systematize parliamentary information architecture and ensure its seamless integration. Their recommendation functionality can also be extended beyond multimedia to cater to NLP and annotation solutions.

In the UK, semantic annotation of Hansard has allowed more systematic studies of linguistic change and parliamentary positioning [31–33].

In Estonia, annotation efforts have focused on identifying debate and argument polarity, a special case of sentiment analysis from mixed text and vote data [34, 35].

Finally, and most notably, several teams have worked with the European Parliament. Efforts have focused on LOD curation, in particular the reconciliation of multilingual transcriptions, with the aim to compare their impact on legislation and MEP behavior [12, 36–44].

While these studies rely on significantly different architectures some design patterns can be identified. First, most solutions tend to separate parliamentary LOD schemas from the fine-grained ontologies used to annotate text. This causes a difficulty in integrating ontologies across schemas and internal-external ontology reconciliation.

Second, NLP and ontology-driven annotation pipelines are neither integrated nor synchronized. Their full integration and simultaneous learning-alignment-enrichment cycles could rapidly increase the accuracy of semantic annotation and search.

Third, all rely on the typical single-end-user viewpoint, without looking at the potential of collaborative search and filtering.

Fourth, agent-driven and automated search recommendations are mostly overlooked, whereas the obvious effect of LOD on parliamentary corpora is the sheer scale, semantic complexity, and speed of events that no single citizen can effectively monitor.

These design gaps are the focus of the present study, attempting to propose a more modular approach. This should create the opportunity for key components of previous efforts to be integrated seamlessly. Interoperability would also be enhanced and in turn multilingual corpora reconciliation. Finally, text embeddings and models would also become easier to share across same-language corpora, enabling debate and legislative intelligence internationally, and broader monitoring as per a citizen's interests.

3 Objectives

Some recent prototypes have confirmed LegisIntel solutions are valuable for parliaments. However, the wide scope of functionality to be developed, and the challenges in integrating several technologies have limited progress, as found in a literature review.

To overcome key design gaps, an international open-source initiative among parliaments is proposed to ensure the launch and viability of a suite of LegisIntel solutions. Many institutions, whether experienced with semantic tools or beginning, can join forces and share the workload of developing a wide diversity of generic use cases. An open-source license will also allow a greater diversity of private and public partners to customize existing software packages to the needs of LegisIntel functionality. Finally, an open Big Data platform approach to NLP and semantic component reuse will ensure that existing libraries can be fully leveraged, while scaling to address a growing demand for improving the digital experience of citizens.

To initiate discussions, the present study offers a requirements analysis and system design that can help generate ideas among developer groups. It is best that formal and specific research objectives would be pinpointed after a community consultation has been launched. The objectives of this first phase are therefore exploratory and focus on research questions addressed typically in an innovative systems design lifecycle:

- RQ1: Who are the main end-user groups in parliamentary settings?
- RQ2: How do they differ in the functionality and requirements expected?
- RQ3: What functional design could serve to fulfill the requirements identified?
- RQ4: Which NLP and semantic analytics technologies are necessary?

4 Methodology

The research questions proposed are formulated within the Information Systems (IS) engineering and design research tradition. The outcome sought is also, at the present stage, a "discussion paper" to generate further ideas and motivate open-source contributions from numerous public-domain research teams. Meanwhile, it should leave open the methodologies to be followed by later stages, especially as they may focus on diverse branches of Computer Science (CS), such as ML, LOD, KG, and semantic technologies, all relying on different implementation methods and performance criteria.

Given its focus on architecture across a diverse community of democratic institutions, this study follows the Action Design Research (ADR) methodology [45]. Its purpose is to identify a change opportunity, devise actions to be carried out to favor this change, and design key artefacts for systems and solutions that will support change. In the context of the present study, a focus on phases 1. Problem Formulation and 2. Reflecting and Learning should be later substantiated by 3. Building, Intervention, and Evaluation (BIE) cycles, allowing more 4. Formalization of Learning through actual change outcomes. The ADR research process is linked directly artefacts and the phases typical of unified processes in engineering, involving diagnosis, design, implementation, and evolution or further improvement throughout the lifecycle of the system [46].

Within the context of an open-source initiative, it is expected that changing citizen and end-user experiences will be incremental. With multiple teams making contributions at various points in time, a focus on rigorous DevOps will ensure greater alignment between architecture and implementation artefacts contributed from various sources [47]. Because of the challenges in surveying and monitoring change across a wide diversity of end-users, and the difficulty of sharing work among teams with varying resources and capabilities, artefact delivery and their change impacts will need to be carefully monitored and realigned to ensure sustained relevance of the initiative [48], with continuous commits and updates on a public Git repo.

5 Results

5.1 Functionality and Requirements

In addressing RQ1 and RQ2, research begins from understanding the assets of a parliament, which are the numerous text, audio, and video contents linked to series of scheduled and loosely structured debates by members and witnesses invited in committees.

Parliamentary proceedings must fulfill the information needs of several end-user groups. In addition to providing some search solutions by grouping user categories as per functionality (e.g., information aggregated for a constituency, or voting records for legal professionals monitoring a bill), customizable profiles are necessary to ensure configurable search linked to individual interests. Therefore, a suite of LegisIntel solutions requires technologies facilitating an unlimited range of end-user experience, especially by optimizing Human-Computer Interactions (HCI), including collaboration among end-users and automated processes. It must also ensure ubiquitous and mobile solutions that ensure information is used in timely fashion.

Integrating heterogenous information sources and better identifying the various concepts and entities throughout texts are key functionality. Indexing the key sentences and summaries through their own custom knowledge base is also essential to enable automated search. Recommending relevant text excerpts which they can build upon as a team, and possibly compose and customize legislations that have been thoroughly tested for coherence and accuracy, can serve to the reuse-refactor cycle of proceedings.

These complex search and semantic functions require LOD assets, published by parliaments under public-domain licenses, and taking many forms. The best known are related to parliamentarians and other participants: profiles and interventions of parliamentarians, witness, and visitors, transcribed and translated debates, voting history of parliamentarians. Another dataset relates to parliamentary operations: agendas and schedules of chambers and committees, directories of laws and regulations at various stages of development. The last and most complex are major documents submitted and processed by parliamentarians: reports and investigations by external entities, public accounts, and supreme audit agencies.

Overall, while functionality is mostly linked to information search and use, there is an increasing concern with automating numerous tasks and ensuring end-users can focus their attention on interpretating parliamentary information. Its dynamic nature implies a heavy workload in extracting meaning from debates for legislative analysis.

As presented in Table 2, results linked to RQ1 and RQ2 show solutions can have functionality more relevant to specific end-user groups. These groups are outlined in Table 3, with groups sharing certain similarities in information use behavior:

- G1 and G2 are more concerned with in-depth and legal substance of decisions.
- G3, G4 and G5 have more concerns across a wide range of policies, and events.
- G6 and G7 are the most numerous but very focused on customizable alerts.

Table 2. Functionality of legislative intelligence solutions.

#	Common to all end-user groups	#	Specific to some end-user groups
F1	Semantic annotations	F6	Topic-based monitoring
F2	Semantic search	F7	Jurisdiction-based monitoring
F3	Automatic summarization	F8	Public accounts analytics
F4	Automatic translation	F9	Interface design and data exploration
F5	Automatic recommendation	F10	Collaborative annotation and analysis

The assignment of functionality to each group is only preliminary and serves to initiate discussions with end-users to become co-developers of this LegisIntel solution. They can be interpreted as the most value-adding functions to the core process of each user group. It also attempts to form coalitions of end-users that may share specific functionality of importance, and therefore creates stronger momentum in prioritizing them.

Table 3. Parliament end-user groups and benefits of legislative intelligence.

#	Group	Enhancements to information search and use	Functionality
G1	Legislative professionals	Researching the complex interrelations, and test the relevance and consistency between, the variety of laws and regulations	F1–F10
G2	Public policy analysts	Government executives and their analysts concerned by parliament proceedings touching on their policies and programs	F1–F5, F8–F10
G3	Municipal officials	Analyzing financial and service implications across a wide range of policies and ensuring a coherent response to jurisdiction parliamentarians	F1–F5, F7–F9
G4	Lobbyist groups	Monitoring upcoming laws and regulations, pinpointing potential implications, aligning their messages with key parliamentarians, and lobbying	F1–F6, F8-F10
G5	Investigative journalists	Journalists primarily of print media seeking detailed analyses on key topics, especially raising significant evidence and testimony	F1–F6, F8–F10
G6	Daily news journalists	Journalists of all media channels in need to stay in touch and ensure accuracy of latest political news and their potential impact	F1–F7
G7	Informed citizens	Academics, students, activists, politicians, etc., receiving alerts of new events related to topics and jurisdictions of their concerns	F1–F7

To illustrate the necessary requirements, a mockup of a system interface, enabling the semantic annotation of parliamentary debates, is shown in Fig. 1. Notice that several external knowledge graphs are linked to semantic tags, which should ideally be filtered given the choice of concepts from the LegisIntel ontology. The taxonomy presented to end-users should also be linked to their preferences and ongoing search patterns. These features will be added to an existing LOD portal of a parliament, ensuring that existing assets are fully integrated within the KG and ontologies.

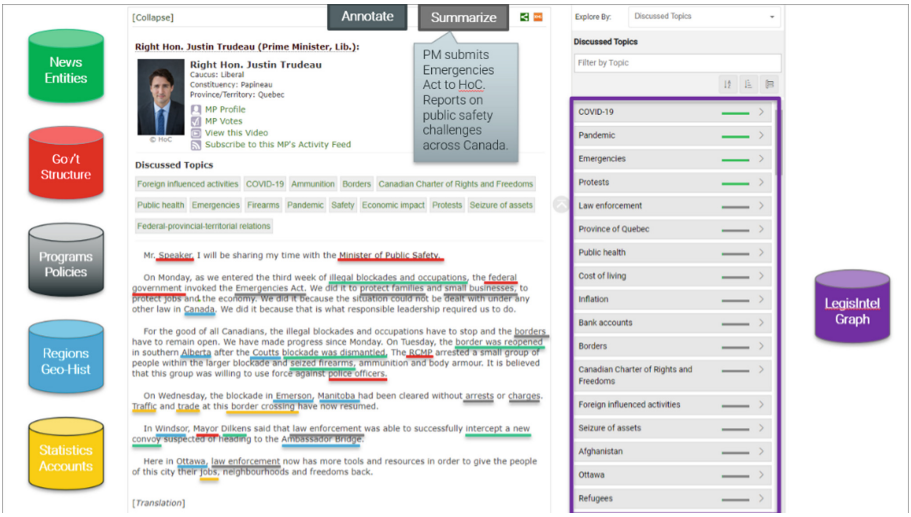

Fig. 1. Mockup of an interface for semantic annotation of parliamentary debates, with external knowledge graphs from various sources, and a LegisIntel ontology used as filtering taxonomy. The LOD portal without LegisIntel features is at https://www.ourcommons.ca/en/open-data.

5.2 Functional Design

In the present study, RQ3 raises a key issue of how to integrate a wide diversity of functionality to support all end-user groups. A functional design and overall process-oriented architecture is suggested, especially to overcome the gaps identified in the literature review. A focus on automation to overcome LOD overload is also favored.

The process and databases leveraging semantic annotations in parliamentary debates are shown in Fig. 2. All end-users can follow the same four steps: create a profile, choose topics from an ontology, start monitoring agents to identify relevant contents, and filter alerts as per evolving interests.

Most importantly, as shown in Fig. 1, the LegisIntel graph allows for seamless integration of debate semantics with both internal and external KG data. As such, it serves as a single index for concepts used to customize end-user experience. These concepts are also combined in filters, rules, and annotations, hence enabling widespread sharing and interoperability among end-users. Finally, they are leveraged by monitoring agents to match end-user trigger concepts with events, especially from external sources.

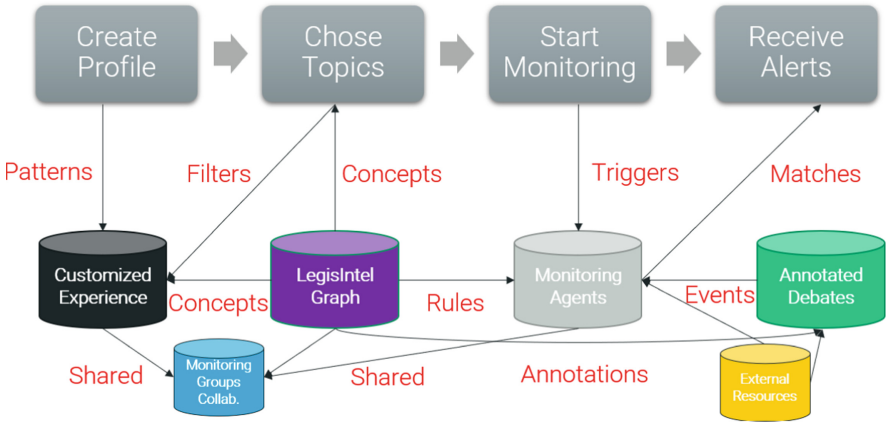

Fig. 2. Process and databases leveraging semantic annotations in parliamentary debates (rightmost database) for agent-driven monitoring of statements that match a citizen's interests.

5.3 Technological Infrastructure

Research on RQ4 led to discussions with parliamentary information architects, along with NLP and semantic technology researchers. It allowed to reach a consensus as to the degree of reuse and integration tasks. More importantly, it aimed to address design gaps identified in a literature review and to implement the functional design suggested.

As shown in Fig. 3, existing parliamentary assets (grey) should be easy to integrate and link to KG technologies. Components linked to end-user experience (green) are existing web apps that should be decoupled from the architecture. Numerous open-source and proprietary components exist and should be easy to integrate throughout the system (blue). The platform should emphasize vendor-neutral semantic standards.

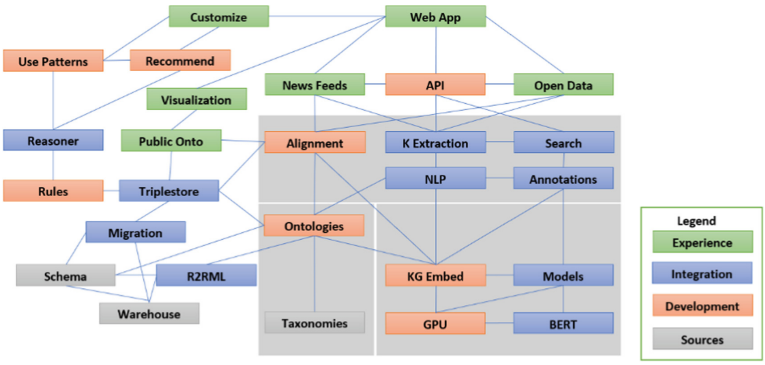

Fig. 3. Software components of a parliamentary debate semantic annotation and query system. (Color figure online)

The bulk of development efforts (orange) relate to five areas, of which three are a priority (large grey boxes). First, the ontology requires substantial collaborative and

iterative efforts. It is still unclear which engineering methods shall be favored, as the parliamentary community differs significantly from domains where ontologies have been mostly developed and reused, cf. healthcare informatics and e-business solutions.

Second, ontology alignment will represent significant challenges, especially given the multilingual nature of proceedings in several countries, and the need for reconciliation across corpora. Complex solution will be required to make alignment a continuous integration and delivery process and involve Big Data infrastructure to meet heavy corpora and large KG assets.

Third, significant investment must be made in Graphical Processor Unit (GPU) development. Platforms have improved rapidly, in particular BERT [49] and graph embeddings as an extension [50–53]. Model reuse depends heavily on pretraining but given the open nature of parliamentary proceedings, it should be easy to develop sharable and reliable models in the main languages. As well, as parliamentary LOD is time-sensitive, great attention will be required for developing a Time Ontology and Annotation type system that will enhance the optimization of GPU-enabled KG models.

Other development efforts should be emerging as these core technologies are incrementally implemented and reach a critical level of "alpha" testing. These other tasks include developing semantic search rules to leverage the ontology and KG models, to be linked to end-user experience and semantic reasoning for agent-driven functionality. As well, a community-built Application Programming Interface (API) shall emerge as this initiative reaches maturity and readiness for "beta" testing and deployments.

6 Conclusion

An international initiative was proposed to develop a new extensible architecture for Legislative Intelligence (LegisIntel). Following an Action Design Research (ADR) methodology, the first phase of this research was reported, focusing on four research questions attempting to identify requirements and architecture components. A literature review of NLP and semantic annotation technologies used in parliaments showed some key design gaps, related mostly to interoperability, integration, and automation.

The results show that parliamentary information behavior differs significantly among end-user groups, but that citizen experience can be enhanced by enabling collaborative monitoring. By automatically annotating debates and linking to internal and external KG assets, the proposed functional design can allow citizens to choose the most relevant concepts to monitor, while sharing search patterns and manage triggers and alerts.

Since most parliaments publish proceedings in LOD assets, the risks of this initiative are primarily at the level of sharing development efforts for numerous new components. NLP and semantic technologies benefit from mature open standards and platforms, but ontology engineering, alignment, and embedding will all require innovative efforts.

It is likely that this initiative may have significant practical implications for parliaments everywhere. The stimulation of democratic participation is an essential function of parliamentary institutions and leveraging LegisIntel solutions may help rapidly increase citizen involvement. As well, the use of Open-Source solutions will enable smaller institutions to share the same basic functionality as larger ones, while allowing groups of parliaments to form as needed to develop custom solutions. In the end, it may serve to strengthen their architecture and better integrate LOD and cloud strategies.

The next steps in this effort are to launch a wider consultation with development teams across IT teams within parliament and congress institutions. An immediate focus on multilingual and corpora reconciliation should be ensured, with the aim to make this initiative helpful in legislative and democratic monitoring. As well, given its open-source orientation, the initiative should emphasize early involvement of citizen groups.

References

1. Phan, T.T., Lam, H.-P., Hashmi, M., Choi, Y.: Towards construction of legal ontology for korean legislation. In: Aveiro, D., Dietz, J., Filipe, J. (eds.) IC3K - Proc. Int. Jt. Conf. Knowl. Discov., Knowl. Eng. Knowl. Manag., pp. 86–97. SciTePress (2020)
2. Mockus, M., Palmirani, M.: Legal ontology for open government data mashups. In: Edelmann, N., Parycek, P. (eds.) Proc. Int. Conf. E-Democr. Open Govern., CeDEM, pp. 113–124. Institute of Electrical and Electronics Engineers Inc. (2017). https://doi.org/10.1109/CeDEM.2017.25
3. Hasan, M.M., Anagnostopoulos, D., Kousiouris, G., Stamati, T., Loucopoulos, P., Nikolaidou, M.: An ontology based framework for e-government regulatory requirements compliance. Intl. J. E Serv. Mob. Appl. **11**, 22–42 (2019). https://doi.org/10.4018/IJESMA.2019040102
4. Lamharhar, H., Chiadmi, D., Benhlima, L.: Ontology-based knowledge representation for e-Government domain. In: Steinbauer M., Indrawan-Santiago M., Anderst-Kotsis G., Khalil I. (eds.) Int. Conf. Inf. Integr. Web-Based Appl. Serv., iiWAS - Proc. Association for Computing Machinery, Inc. (2015). https://doi.org/10.1145/2837185.2837203
5. Nečaský, M., Knap, T., Klímek, J., Holubová, I., Vidová-Hladká, B.: Linked open data for legislative domain - ontology and experimental data. In: Abramowicz, W. (ed.) BIS 2013. LNBIP, vol. 160, pp. 172–183. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41687-3_17
6. Fonou Dombeu, J.V., Huisman, M.: Semantic-driven e-government: application of uschold and king ontology building methodology for semantic ontology models development. Int. J. Web Semantic Technol. **2**, 111–120 (2011). https://doi.org/10.5121/ijwest.2011.2401
7. Loukis, E., Xenakis, A.L.: A methodology for ontology-based knowledge-level interoperability among parliaments. In: Amer. Conf. Inf. Sys., AMCIS, pp. 5112–5122. San Francisco, CA (2009)
8. Agnoloni, T., Bacci, L., Francesconi, E.: Ontology based legislative drafting: design and implementation of a multilingual knowledge resource. In: Gangemi, A., Euzenat, J. (eds.) EKAW 2008. LNCS (LNAI), vol. 5268, pp. 364–373. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87696-0_31
9. Rafail, P., Efthimios, T.: Knowledge graphs for public service description: τhe case of getting a passport in Greece. In: European, Mediterranean, and Middle Eastern Conference on Information Systems, pp. 270–286. Springer (2020)
10. Zittel, T., Nyhuis, D., Baumann, M.: Geographic representation in party-dominated legislatures: a quantitative text analysis of parliamentary questions in the german bundestag. Legis. Stud. Q. **44**, 681–711 (2019). https://doi.org/10.1111/lsq.12238
11. Eide, S.R.: The swedish poligraph: A semantic graph for argument mining of Swedish parliamentary data. In: ACL - Workshop Argument Min., ArgMining - Proc. Workshop, pp. 52–57. Association for Computational Linguistics (ACL) (2019)
12. Mendonça, I., Trouve, A., Fukuda, A.: Exploring the importance of negative links through the european parliament social graph. In: Proceedings of the 2017 International Conference on E-Society, E-Education and E-Technology, pp. 1–7. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3157737.3157744

13. Jakubíček, M., Kovář, V.: Czechparl: Corpus of stenographic protocols from Czech parliament. In: RASLAN - Recent Advances Slavonic Nat. Lang. Proc., 4th Workshop Recent Advances Slavonic Nat. Lang. Proc., Proc., pp. 41–46. Masaryk University, Karlova, Studanka (2010)
14. Xing, Z., Hillygus, S., Carin, L.: Evaluating U.S. Electoral representation with a joint statistical model of congressional roll-calls, legislative text, and voter registration data. In: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 1205–1214. Association for Computing Machinery (2017). https://doi.org/10.1145/3097983.3098151
15. Smith, S., Baek, J.Y.C., Kang, Z., Song, D., Ghaoui, L.E., Frank, M.: Predicting congressional votes based on campaign finance data. In: Proc. - Int. Conf. Mach. Learn. Appl., ICMLA, pp. 640–645. Boca Raton, FL (2012). https://doi.org/10.1109/ICMLA.2012.119
16. Maynard, S.: Automatic semantic annotation for the easification of action rule legislative sentences for specialist readers. In: Ashley, K.D., et al. (eds.) CEUR Workshop Proc., pp. 36–47. CEUR-WS (2021)
17. Quijano-Sánchez, L., Cantador, I.: Semantic annotation and retrieval of parliamentary content: A case study on the Spanish congress of deputies. In: Cantador, I., Chevalier, M., Melucci, M., Mothe, J. (eds.) CEUR Workshop Proc. CEUR-WS (2020)
18. Ruprechter, T., Khosmood, F., Kuboi, T., Dekhtyar, A., Gütl, C.: Gaining efficiency in human assisted transcription and speech annotation in legislative proceedings. In: Hinnant, C.C., Zuiderwijk, A. (eds.) ACM Int. Conf. Proc. Ser. Association for Computing Machinery (2018). https://doi.org/10.1145/3209281.3209410
19. De Campos, L.M., Fernández-Luna, J.M., Huete, J.F., Martín-Dancausa, C.J.: An integrated system for accessing the digital library of the parliament of Andalusia: Segmentation, annotation and retrieval of transcriptions and videos. In: Proc. Int. Workshop Pattern Recogn. Inf. Syst. - PRIS; In Conjunction ICEIS, pp. 38–47. Barcelona (2008)
20. Andrushchenko, M., et al.: Using parsed and annotated corpora to analyze parliamentarians' talk in Finland. J. Assoc. Soc. Inf. Sci. Technol. **73**, 288–302 (2022). https://doi.org/10.1002/asi.24500
21. Almeida, P., Marques-Pita, M., Gonçalves-Sá, J.: PTPARL-D: an annotated corpus of forty-four years of Portuguese parliamentary debates. Corpora. **16**, 337–348 (2021). https://doi.org/10.3366/cor.2021.0226
22. Koeva, S., Obreshkov, N., Yalamov, M.: Natural language processing pipeline to annotate bulgarian legislative data. In: Calzolari, N., et al. (eds.) LREC - Int. Conf. Lang. Resour. Eval., Conf. Proc., pp. 6988–6994. European Language Resources Association (ELRA) (2020)
23. Lapponi, E., Søyland, M.G., Velldal, E., Oepen, S.: The talk of Norway: a richly annotated corpus of the Norwegian parliament, 1998–2016. Lang. Resour. Eval. **52**(3), 873–893 (2018). https://doi.org/10.1007/s10579-018-9411-5
24. Höfler, S., Sugisaki, K.: Constructing and exploiting an automatically annotated resource of legislative texts. In: Calzolari, N., et al. (eds.) Int. Conf. Lang. Resourc. and Eval. – LREC, pp. 175–180. European Language Resources Association (ELRA) (2014)
25. Sinikallio, L., et al.: Plenary debates of the parliament of Finland as linked open data and in parla-CLARIN markup. In: Gromann, D., et al. (eds.) OpenAccess Ser. Informatics. Schloss Dagstuhl- Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing (2021). https://doi.org/10.4230/OASIcs.LDK.2021.8
26. Erjavec, T.: The ParlaMint corpora of parliamentary proceedings. Lang. Resour. Eval. (2022). https://doi.org/10.1007/s10579-021-09574-0
27. Cifuentes-Silva, F., Rivera-Polo, F., Labra-Gayo, J.E., Astudillo, H.: Describing the nature of legislation through roll call voting in the chilean national congress, a linked dataset description. Semantic web. (2021)

28. Cifuentes-Silva, F., Fernández-Álvarez, D., Labra-Gayo, J.E.: National budget as linked open data: new tools for supporting the sustainability of public finances. Sustainability **12**, 4551 (2020). https://doi.org/10.3390/su12114551
29. Cifuentes-Silva, F., Labra Gayo, J.E.: Legislative document content extraction based on semantic web technologies. In: Hitzler, P., et al. (eds.) ESWC 2019. LNCS, vol. 11503, pp. 558–573. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21348-0_36
30. Sánchez-Nielsen, E., Chávez-Gutiérrez, F., Lorenzo-Navarro, J.: A semantic parliamentary multimedia approach for retrieval of video clips with content understanding. Multimedia Syst. **25**(4), 337–354 (2019). https://doi.org/10.1007/s00530-019-00610-2
31. Coole, M., Rayson, P., Mariani, J.: Unfinished Business: Construction and Maintenance of a Semantically Tagged Historical Parliamentary Corpus, UK Hansard from 1803 to the present day. In: Proceedings of the Second ParlaCLARIN Workshop, pp. 23–27. European Language Resources Association, Marseille, France (2020)
32. Nanni, F., Menini, S., Tonelli, S., Ponzetto, S.P.: Semantifying the UK Hansard (1918–2018). In: Proceedings of the 18th Joint Conference on Digital Libraries, pp. 412–413. IEEE Press, Champaign, Illinois (2019). https://doi.org/10.1109/JCDL.2019.00095
33. Abercrombie, G., Batista-Navarro, R.: Semantic change in the language of UK parliamentary debates. In: Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, pp. 210–215. Association for Computational Linguistics, Florence, Italy (2019). https://doi.org/10.18653/v1/W19-4726
34. Koit, M.: How are the members of a parliament arguing? analysis of an argument corpus. In: Rocha, A.P., Steels, L., van den Herik, J. (eds.) ICAART - Proc. Int. Conf. Agents Artif. Intell., pp. 1046–1053. SciTePress (2021)
35. Koit, M.: Annotating arguments in a parliamentary corpus: an experience. In: Aveiro, D., Dietz, J., Filipe, J. (eds.) IC3K - Proc. Int. Jt. Conf. Knowl. Discov., Knowl. Eng. Knowl. Manag., pp. 213–218. SciTePress (2020)
36. Michon, S., Wiest, E.: A database about the members of European parliament: contributions and limitations of automated data collection in the study of European political elites. Bull. Sociol. Methodol. **152**, 125–141 (2021). https://doi.org/10.1177/07591063211040235
37. Däubler, T., Chiru, M., Hermansen, S.S.L.: Introducing COMEPELDA: comprehensive European parliament electoral data covering rules, parties and candidates. Eur. Union Polit. **23**, 351–371 (2021). https://doi.org/10.1177/14651165211053439
38. Pérez, M.C.: The group in the self a corpus-assisted discourse studies approach to personal and group communication at the European parliament. Pragmatics **29**, 357–383 (2019). https://doi.org/10.1353/gpr.2019.0000
39. Plevoets, K., Defrancq, B.: The cognitive load of interpreters in the European parliament A corpus-based study of predictors for the disfluency uh(m). Interpreting **20**, 1–28 (2018). https://doi.org/10.1075/intp.00001.ple
40. Defrancq, B.: The European parliament as a discourse community: its role in comparable analyses of data drawn from parallel interpreting corpora. Interpreters Newsl. **23**, 115–132 (2018). https://doi.org/10.13137/2421-714X/22401
41. Calzada-Pérez, M.: Researching the european parliament with corpus-assisted discourse studies from the micro- and macro-levels of text to the macro-context. Rev. Esp. Ling. Apli. **30**, 465–490 (2018). https://doi.org/10.1075/resla.00003.cal
42. Calzada Pérez, M.: What is kept and what is lost without translation? a corpus-assisted discourse study of the European Parliament's original and translated English. Perspect. Stud. Transl. **26**, 277–291 (2018). https://doi.org/10.1080/0907676X.2017.1359325
43. Veroz, M.A.: Translation in the European parliament: the study of the ideational function in technical texts (EN/FR/ES). Meta **62**, 19–44 (2017). https://doi.org/10.7202/1040465ar

44. Van Aggelen, A., Hollink, L., Kemman, M., Kleppe, M., Beunders, H.: The debates of the European parliament as linked open data. Semantic Web. **8**, 271–281 (2017). https://doi.org/10.3233/SW-160227
45. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action design research. MIS Q. **35**, 37–56 (2011)
46. Mullarkey, M.T., Hevner, A.R.: An elaborated action design research process model. Eur. J. Inf. Syst. **28**, 6–20 (2019). https://doi.org/10.1080/0960085X.2018.1451811
47. Gill, A.Q., Chew, E.: Configuration information system architecture: insights from applied action design research. Inform. Manag. **56**, 507–525 (2018). https://doi.org/10.1016/j.im.2018.09.011
48. Collatto, D.C., Dresch, A., Lacerda, D.P., Bentz, I.G.: Is action design research indeed necessary? analysis and synergies between action research and design science research. Syst. Pract. Action Res. **31**(3), 239–267 (2017). https://doi.org/10.1007/s11213-017-9424-9
49. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT - Conf. N. Am. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol. - Proc. Conf., pp. 4171–4186. Association for Computational Linguistics (ACL) (2019)
50. Zhao, Y., Zhang, A., Feng, H., Li, Q., Gallinari, P., Ren, F.: Knowledge graph entity typing via learning connecting embeddings. Knowl. Based Syst. **196**, 105808 (2020). https://doi.org/10.1016/j.knosys.2020.105808
51. Palumbo, E., Monti, D., Rizzo, G., Troncy, R., Baralis, E.: entity2rec: property-specific knowledge graph embeddings for item recommendation. Expert Sys. Appl. **151**, 113235 (2020). https://doi.org/10.1016/j.eswa.2020.113235
52. Leblay, J., Chekol, M.W., Liu, X.: Towards temporal knowledge graph embeddings with arbitrary time precision. In: Int. Conf. Inf. Knowledge Manage., pp. 685–694. Association for Computing Machinery (2020). https://doi.org/10.1145/3340531.3412028
53. Assi, A., Mcheick, H., Dhifli, W.: Data linking over RDF knowledge graphs: a survey. Concurrency Comput. **32**, e5746 (2020). https://doi.org/10.1002/cpe.5746

Can the GDPR Allay Privacy Concerns Towards Smart Products? The Effect of a Compliance Seal on Perceived Data Security, Trust, and Intention to Use

Oliver Michler(✉), Christian Stummer, and Reinhold Decker

Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany
{oliver.michler,christian.stummer,
reinhold.decker}@uni-bielefeld.de

Abstract. Smart consumer products are designed to provide their users with various benefits. To utilise smart products and enjoy their benefits, users usually have to provide some kind of information to the product and its manufacturer—often personal data. This can raise privacy and data security concerns and may hamper the use of smart products. The European Union's (EU) General Data Protection Regulation (GDPR) addresses these concerns and provides requirements for an appropriate handling of data. Our study assumes that a positive perception of the GDPR can encourage smart product usage. Therefore, we explore the effect of a GDPR compliance seal that signals data security and trustworthiness. By means of an online experiment with 142 participants from Germany, we investigate the seal's effect on perceived data security, perceived trust, and the intention to use a smart robot vacuum cleaner. The results indicate that a GDPR compliance seal indeed has a positive effect on perceived data security and, through perceived data security as a mediator, also on perceived trust. While the direct impact of a GDPR compliance seal on the intention to use a smart product lacks statistical significance, our model reveals an indirect effect via perceived data security and perceived trust as well as a positive total effect of the seal on intention to use.

Keywords: Smart products · GDPR seal · Trust · Privacy concerns · Data security

1 Introduction

Smart products continue to find their way into private lives and households [1]. Certain smart products process extensive data from different sources, and by linking this data, manufacturers can gain comprehensive insights into the everyday life and personal preferences of users [2]. For example, smart speakers react to certain keywords using voice recording [3]; digital assistants leverage personal preferences to provide personalised services [4]; and smartwatches, fitness trackers, and health apps monitor physical activities and body conditions [5].

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 77–91, 2022.
https://doi.org/10.1007/978-3-031-12673-4_6

The suspicion that smart products collect a substantial amount of data and that they may do so in personal situations as well can hamper consumers' willingness to use these products and occasionally even result in resistance towards them [4–6]. Therefore, in order to encourage the usage of smart (consumer) products, it is necessary to address potential causes for distrust, like the collection of data and personal information, and thereby actively build consumer confidence in smart products [7].

The General Data Protection Regulation (GDPR) implemented in the European Union (EU) deals with data collection, data processing, and privacy issues arising from modern technologies and services [8]. Therefore, it also applies to potential issues for distrust in smart products and their manufacturers. However, numerous smart product users are not aware of the applicability of the GDPR and the implied benefits [9, 10]. Hence, they are also not aware of the measures that manufacturers must implement to increase privacy and data security.

The study at hand addresses this issue and, therefore, investigates whether a GDPR compliance seal can impact perceived data security, perceived trust, and the intention to use a smart product. To this end, we conducted an online experiment in which we presented a smart robot vacuum cleaner to two groups of consumers from Germany. In the description of the robot vacuum cleaner presented to the first (experimental) group, we highlighted a GDPR compliance seal, but this was not done for the second (control) group.

Accordingly, the research contribution of our study lies in the investigation of the benefits of communicating GDPR compliance using a seal while promoting a smart product. Apart from the theoretical insights, our findings can also help manufacturers of smart products to better assess the use of a GDPR compliance seal and the alleged effect.

The remainder of this paper is organised as follows: Sect. 2 deals with a classification of smart products, addresses possible security concerns of smart product users, describes how the GDPR is supposed to increase perceived privacy and data security for EU citizens, and outlines the current state of research on the effects of the GDPR on data security and trust in smart products. Section 3 introduces our assumptions regarding the positive effects of a GDPR compliance seal. Section 4 describes the online experiment used to verify the assumptions and presents the results of the research. Finally, Sect. 5 provides theoretical and managerial implications as well as limitations and directions for further research.

2 Background

2.1 Smart Products

While there is no undisputed general definition of smart products, recently, Raff et al. [11] propose four archetypes of smart products—that is, digital products, connected products, responsive products, and intelligent products. The respective archetypes are characterised by certain criteria and build on one another. For the purpose of our study, we are particularly interested in the last two categories. Corresponding products can collect and process data, interpret data in a specific context, and react autonomously. Smart speakers are an illustrative example of a responsive product. If such responsive products

are also able to make their own independent decisions (e.g., by means of reasoning based on artificial intelligence (AI) software for learning, improving, and anticipating) and, thus, act proactively, they can be classified as intelligent products (e.g., driverless cars).

The smart product used in our study is a smart robot vacuum cleaner equipped with a microphone (which enables voice control) and a video camera (which is used for superior cleaning functionality). This robot vacuum cleaner qualifies as a responsive product, but certain (future) versions might also be intelligent following the categorisation by Raff et al. [11].

2.2 Security Concerns Among Smart Product Users

The use of smart products and the associated sharing of personal information entails the risks of online attacks and digital profiling. Recently, there have been more frequent hacks and attacks on smart devices as part of malicious activities [12]. The consequences of such attacks can be minor malfunctions in the form of temporarily non-functioning devices, unavailable services [13], or deliberate damage from, for example, blocking smart locks [14] or influencing smart thermostats [15]. Digital crimes are also not uncommon—for example, stealing sensitive personal information [16], publication of stolen data [13, 17, 18], or spying on users of smart products [19]. Moreover, attackers may occasionally hack end users' devices to carry out viral attacks, like distributed denial of service (DDoS) attacks on service providers [20].

In contrast to attacks, digital profiling is not a harmful activity performed by an outsider but causes fear due to the possibilities of combining collected information from various sources regarding a specific consumer to create a data-rich profile that defines their characteristics, expected behaviour, and preferences [2, 21]. Smart products collect and use a substantial amount of personal information regarding their users to run and optimise their services [18]. For example, in a smart home, cameras could end up recording private situations, and digital assistants might listen to personal conversations [13]. At home, in particular, where privacy is required, this creates uncertainty among users [7].

Furthermore, the use of smart products can also result in a feeling of loss of control [7]. Users might no longer feel that they can control all the actions that the product performs or that they can completely switch off the smart product at any time [22]. In terms of data security and privacy, a loss of control could also refer to data sovereignty over one's own information [16]. Users no longer consider themselves to be the exclusive owners of their data, because they do not know what data is collected at which points in time and for what purposes [17]. In the eventuality that manufacturers of smart products pass on the collected data to third parties, users also lose control over who has access to their data [16].

This raises concerns regarding potential monitoring, the security of personal data, and personal anonymity [22]. Such privacy and data security concerns influence consumers' decisions to use smart products [23, 24]. Therefore, in order to encourage the use of smart products, it is necessary for manufacturers to address these concerns accordingly.

2.3 Improved Privacy and Data Security with GDPR

One way to deal with data security and privacy concerns is to establish legal requirements. The GDPR addresses privacy and data security concerns with a number of regulations [25]. It considers the challenges associated with data security that arise from new technologies, like smart products [8]. In a nutshell, the aim of the GDPR is to protect the privacy and personal data of EU citizens and to give users more control over the use of their data by third party [26].

Only a few studies have examined the effect of the GDPR on privacy concerns and the willingness to share data with smart products. Their findings suggest that users of smart products have developed a better understanding of how the GDPR protects their privacy and contributes to data security [27, 28]. Even if there are still mixed feelings regarding control over one's own data [28], the GDPR has generally led to increased trust between users and the smart product or its manufacturer [2]. This is due to enhanced perceived security and reduced privacy concerns [29, 30] and also has a positive impact on the willingness to share data [31].

However, even though it can be expected that most EU citizens are familiar with the GDPR by now and that they have an increased knowledge of the GDPR regulations, numerous users of (smart) products are not aware of the applicability of the GDPR, the opportunities the GDPR provides them with as data subjects, and the specific measures that product manufacturers are compelled to take because of it [9, 10, 32]. Due to their lack of awareness and occasionally moderate interest in data privacy [28], potential users often lack the knowledge that would enable them to eliminate data security concerns based on legal requirements. Thus, measures to protect personal data are not taken for granted.

This uncertainty of applicability could be overcome with the help of a seal that indicates compliance with the GDPR. A seal represents a common means of communicating compliance with certain specifications at the product display level. The CE seal is a successful example of communicating compliance with EU-wide guidelines in a simple and common manner [33]. Articles 42 and 43 of the GDPR explicitly address the possibility of using a seal for the communication of GDPR compliance [34].

3 Hypothesis Development

The present study contributes to a research stream that is concerned with the (positive) effect of the GDPR on transparency, perceived security, and trust between users and smart products [2, 30]. While prior studies explicitly refer to the GDPR and dwell on the measures taken in this context, in our study, we assume that a simple GDPR compliance seal, which is displayed in a product's promotion/advertisement and does not describe any specific GDPR measures, already has a positive influence on the perception of a specific product. Hence, we investigated the effect of a GDPR compliance seal on perceived data security, perceived trust, and intention to use.

In doing so, our study relies on the signalling theory [35], which can be applied in settings in which the communication of certain information from one party to another party is used to resolve information asymmetry and associated uncertainty [36, 37]. Thus, the signalling theory is suitable in our research setting, in which manufacturers of smart

products as the signallers use a GDPR compliance seal to signal that they comply with the requirements of the GDPR. This message needs to be recognised and understood by potential users of the respective smart products as receivers.

With regard to privacy seals, we draw from previous research in the e-commerce industry which suggests that companies that want to use a reputable privacy seal must follow certain standards [33, 38], and that such privacy seals indeed succeed in signalling the trustworthiness and safety of websites [39–41]. This positive effect was also confirmed for generic security labels in the smart product context [42, 43].

Taken together—and given the positive impact of the GDPR on trust in smart products—we, therefore, assume that a GDPR compliance seal would show positive effects as well. While previous studies often focussed on purchase intention as the dependent (outcome) variable, in our case, we assume that the purchase is only the first necessary step, and usage of the smart product is what providers are actually aiming at. Therefore, in our study, we examined the consumers' intention to use the smart product.

H1: The use of a GDPR compliance seal increases the consumers' intention to use a smart product (direct effect).

The additional trust generated by a privacy seal in the e-commerce industry is usually driven by signalling security and communicating that the data and the privacy of users are protected [33, 34, 44, 45]. The latter has become a necessity, as consumers are becoming increasingly aware of the value of their data [46], while, simultaneously, media often report on the misuse of data that has been accessed by third parties, for example, through hacker attacks or unauthorised disclosure [13, 16].

With regard to data security, prior studies have already shown that this is an important issue for users of smart products [6, 7]. Accordingly, data security and privacy concerns represent a challenge for manufacturers, which may raise consumer resistance towards smart products if not addressed in a suitable manner [5, 47].

The GDPR sets minimum requirements for the protection of consumer data [25]. This protection is signalled through a GDPR compliance seal, which is expected to increase perceived data security.

H2: The use of a GDPR compliance seal increases the perceived data security of a smart product.

Smart products collect data to fulfil their functions and purposes. However, such data collection can deter consumers from using smart products [4–6]. In order to solve this dilemma, and to overcome the respective adoption barrier, trust in the protection of the collected data needs to be created, and the manufacturer of the smart product must ensure privacy [7].

A privacy seal can help in this respect, as has been shown for websites provided in the e-commerce industry [39–41]. The key element here is that consumers trust the certification by an independent institution and transfer the trust in this institution to the certified object [38].

The same mechanism can work for a GDPR compliance seal. Combining the increased perceived data security (see H2) with the trust-building effect of the seal, we can assume that a GDPR compliance seal also increases trust in a smart product.

H3: The effect of a GDPR compliance seal, mediated by perceived data security, increases perceived trust in a smart product.

Finally, we expect that by increasing perceived data security and building trust in a smart product, a GDPR compliance seal will also have a positive indirect effect on the intention to use a smart product. Perceived data security and perceived trust function as mediators of the effect of a GDPR compliance seal on the intention to use.

Therefore, for manufacturers of smart products, a GDPR compliance seal could be a straightforward means of communicating compliance with data protection requirements. Thus, in addition to the direct effect of a GDPR compliance seal on the intention to use (see H1), we assume a positive indirect effect of a GDPR compliance seal when it is used in product communication. Combining this indirect effect with the direct effect of the seal is likely to also result in a positive overall effect of a GDPR compliance seal on the intention to use a smart product.

H4: The effect of a GDPR compliance seal, mediated by perceived data security and perceived trust, increases the intention to use a smart product (indirect effect).

H5: The sum of the direct and indirect effects of a GDPR compliance seal has a positive impact on the intention to use a smart product (total effect).

4 Empirical Study

4.1 Data Collection and Measurement

To test our hypotheses, we conducted an online experiment with 142 participants from Germany who were referred to our questionnaire through the research panel *respondi* (www.respondi.com). The sample comprised 45% female and 55% male participants, with an average age of 45.6 years (SD = 10.7 years). The participants were familiar with the GDPR or at least knew about its relevance regarding data protection.

At the beginning of the experiment, the participants were asked to put themselves in the situation of searching for a smart robot vacuum cleaner to be purchased by them. This robot vacuum cleaner is portrayed as a (responsive) smart product that could raise concerns based on its access to voice and video recordings [48]. The participants were then randomly assigned to one of two groups; one group (experimental group) was provided with a product advertisement that included a GDPR compliance seal in the bottom right corner of the advertisement (see Fig. 1), while the other group (control group) is part of a setting without such a seal (see Fig. 2). Furthermore, the participants were asked to read a thorough description of the vacuum cleaner. The description was identical in both settings; the only exception was that in one setting, the participants were informed that the robot vacuum cleaner is GDPR-compliant, while in the other setting, the GDPR was not mentioned at all. Note that the advertisement as well as the description of the robot vacuum cleaner were provided in German.

The realism of the product display was confirmed by a pre-test ($N_1 = 30$). The same pre-test was used to evaluate several alternative manufacturer names, from which the name 'Jarvis' was rated as most suitable.

After viewing the product display, the participants were asked about their intention to use the smart robot vacuum cleaner, their perception of data security, and their trust in the smart product manufacturer. All three aspects were queried using 7-point Likert scales (1 = 'strongly disagree' to 7 = 'strongly agree'). The questions for assessing the intention to use were based on two items suggested by Lu et al. [49] and one new item. In addition, five items from Kim et al. [50] were used to identify perceived data security. To assess the perceived trust in the manufacturer of the smart product, we measured three frequently used characteristics—benevolence, integrity, and ability [51]. To this end, we used seven items from the scale provided by Bhattacherjee [52]. All the above-mentioned items are listed in the Appendix.

The questionnaire was then tested in a second pre-test with 51 participants ($N_2 = 51$). Both the second pre-test and the main experiment confirmed the success of the manipulation and the realism of the situation and product display.

Fig. 1. Product advertisement with a GDPR compliance seal

Fig. 2. Product advertisement without a GDPR compliance seal

4.2 Results

Overall, we identified several effects, all of which are displayed in Fig. 3. Most prominently, our analysis suggests a positive, significant total effect of the GDPR compliance seal on the intention to use (0.701, $p = 0.035$). While the direct effect of the seal on usage intention does not prove significance (0.371, $p = 0.207$), the indirect effect of the seal on the intention to use a smart product—through perceived data security and perceived trust—is significantly positive (0.119, 95% CI [0.011, 0.268]).

The indirect effect ultimately results from an increase in perceived trust that has a significant positive effect on intention to use (0.692, $p < 0.001$). As we did not find a significant effect of the GDPR compliance seal directly on perceived trust (0.166, $p = 0.401$), we resorted to a serial mediation that began with a significant effect of the compliance seal on perceived data security (0.626, $p = 0.027$), which, on its part, had a significant effect on perceived trust (0.275, $p < 0.001$). It is noteworthy that the mediation is only partial in all these instances, as the direct effect of perceived data security on usage intention is also not significant (0.154, $p = 0.101$).

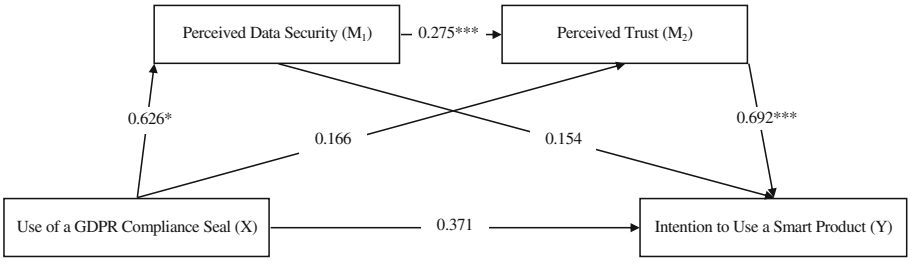

Fig. 3. Statistical diagram of the effects of a GDPR compliance seal at the following levels of significance: * = significant at $p < 0.05$; ** = significant at $p < 0.01$; *** = significant at $p < 0.001$.

4.3 Discussion

The objective of our study was to demonstrate that the communication of GDPR compliance by using a GDPR compliance seal has a positive effect on the intention to use a smart product. We assumed that the effect was mediated by perceived data security and perceived trust.

Our first hypothesis states that the GDPR compliance seal has a direct positive effect on the intention to use a smart product. The empirical data does not support this assumption; thus, we have to reject H1.

In our second hypothesis, we addressed the first part of the expected indirect effect on usage intention when assuming that the GDPR compliance seal has a positive effect on perceived data security. Indeed, the perceived data security is significantly higher when the promotion of a smart product includes a GDPR compliance seal ($M_{1(\text{without seal})} = 2.432$; $M_{1(\text{with seal})} = 3.058$), which indicates that the product is compliant with the requirements of the GDPR. Thus, we have evidence that leads us to confirm H2.

For the second part of the expected indirect effect, we tested whether the GDPR compliance seal, mediated by perceived data security, also increases perceived trust in a smart product. Indeed, the perceived trust is significantly higher for smart products that have a GDPR compliance seal ($M_{2(\text{without seal})} = 4.590$; $M_{2(\text{with seal})} = 4.928$). Thus, our results reveal that a significant mediation exists and, accordingly, the GDPR compliance seal has a positive effect on perceived trust, which confirms H3.

Furthermore, as result of a serial mediation through positive effects of the seal on perceived data security and perceived trust, we also expected a positive indirect effect of a GDPR compliance seal on the intention to use a smart product and, indeed, we found the corresponding increase in the intention to use the smart robot vacuum cleaner in our experimental study. Consequently, we were able to confirm H4.

As the bottom line, our model reveals a significantly positive total effect of the GDPR compliance seal on the intention to use the smart robot vacuum cleaner ($Y_{(\text{without seal})} = 3.464$; $Y_{(\text{with seal})} = 4.165$). This finding leads us to confirm hypothesis H5—that the sum of the direct and indirect effects of a GDPR compliance seal has a positive impact on the intention to use a smart product.

5 Conclusion

5.1 Theoretical Implications

Our study examined the effect of a GDPR compliance seal on perceived data security, perceived trust, and, indirectly, on the intention to use smart products. This study contributes to a stream of research with (i) studies in the e-commerce industry, which identified a positive effect of privacy seals on perceived data security [33, 34, 44, 45], (ii) studies on the impact of perceived data security and perceived trust on the usage of smart products [6, 7], and (iii) studies on the effect of the GDPR on perceived data security [30], in particular in the smart product context [29]. While our results are in line with these earlier studies, we contribute to this field by demonstrating the positive effect of communicating compliance with the GDPR by using a simple seal. To this

end, we applied the signalling theory to expand the findings regarding the use of certification seals by combining them with active communication of compliance with legal requirements.

5.2 Managerial Implications

Our study offers several new insights for the manufacturers of smart products. First and foremost, a seal can communicate the GDPR compliance of smart products in a simple and condensed format. The use of a GDPR compliance seal in the course of product promotion is found to allay privacy concerns among smart product users and, thus, increase perceived data security, perceived trust, and intention to use. Although it was not tested in our experimental study, we can safely assume that using a GDPR compliance seal is considerably more effective in these respects than traditional (lengthy) legal statements, which are often ignored by consumers [53].

Logically, the increased intention to use a smart product, when it is marketed by means of a GDPR compliance seal, should also increase consumers' purchase intention. Consequently, the use of the seal would likely improve sales figures—at least in markets in which data security and privacy concerns play an important role.

Furthermore, the seal increases perceived data security, which should make users more willing to share data with the manufacturer of a smart product, which is a mandatory requirement for the use of most smart products. Thus, both the increase in usage of a smart product and the increased willingness to share data provide manufacturers the opportunity to collect more information on the use and behaviour of a smart product in certain situations. In turn, this information can be used to further develop and improve smart products, thereby ensuring that ultimately the functionality of the product increases and, thus, the consumer benefits from the use of a GDPR compliance seal.

Since compliance with the GDPR is a requirement in EU markets, meeting GDPR regulations should not lead to any additional development efforts. Manufacturers may have to pass a certification process to be allowed to use a GDPR compliance seal, but the advantages of using a seal must clearly outweigh the associated effort of such a certification.

5.3 Limitations and Future Research

Our study makes a valuable contribution to the extant research on the use of smart products. Nevertheless, it has certain limitations, but these also provide promising avenues for further research.

First, we only collected data from consumers in Germany, an EU country known for its relatively strong focus on data protection issues [54]. Future studies should examine the effect of a GDPR compliance seal in different countries within and outside of the EU.

Second, we believe that the smart product that was used in our empirical study (i.e., the smart robot vacuum cleaner) is well suited for our study due to its video streaming functionality and listening capabilities to react to voice control, which implies that it potentially collects a substantial amount of data from users' homes—that is, from the

personal space of its users. However, a different type of smart product could confirm our results and possibly measure an even stronger effect of the GDPR compliance seal if it is used for a product that collects more personal, more critical, or simply a larger amount of data. Smart health products or smart speakers in combination with digital assistants could be suitable for such an exploration.

Third, we confirmed several hypotheses that suggest a positive effect of a GDPR compliance seal. However, further research is necessary to better understand the psychological effects associated with the usage of a GDPR compliance seal. It is possible that for certain consumers, the seal confirms their assumption that the smart product needs to be GDPR-compliant. For other consumers, the seal might reduce data protection concerns. Moreover, there might also be consumers who only become aware of data security threats because of the seal (and would not have given a thought to such concerns otherwise). Such calming, alarming, or confirmatory effects of a compliance seal on consumers could be investigated in future research.

Finally, future studies could account for additional factors. For example, a conjoint analysis could investigate the impact of the country of origin, the product's price, or its brand. Moreover, the effect of the GDPR compliance seal could be compared with the effect of a more general third-party data security or privacy seal that does not refer to the legal regulations set by the GDPR.

Appendix

Table 1.

Table 1. Questionnaire items for measuring intention to use, perceived data security, and trust.

Construct	Origin	(Adapted) item
Intention to use	Lu et al. 2005	I am willing to use the smart robot vacuum cleaner
		Using the smart robot vacuum cleaner is worthwhile
	New item	It is likely that I will use the smart robot vacuum cleaner
Perceived data security	Kim et al. 2008	I am concerned that this smart robot vacuum cleaner is collecting too much personal information from me
		I am concerned that the manufacturer of the smart robot vacuum cleaner will use my personal information for other purposes without my authorisation

(*continued*)

Table 1. (*continued*)

Construct	Origin	(Adapted) item
		I am concerned that the manufacturer of the smart robot vacuum cleaner will share my personal information with other entities without my authorisation
		I am concerned that unauthorised persons (i.e. hackers) have access to my personal information
		I am concerned that the manufacturer of the smart robot vacuum cleaner will sell my personal information to others without my permission
Perceived trust (ability)	Bhattacherjee 2002	The manufacturer has the skills and expertise to develop a smart robot vacuum cleaner which will perform in an expected manner
		The manufacturer has access to the information needed to let the robot vacuum cleaner perform appropriately
		The manufacturer has the ability to meet most customer needs
Perceived trust (integrity)	Bhattacherjee 2002	The manufacturer is fair in its use of private user data
Perceived trust (benevolence)	Bhattacherjee 2002	The manufacturer keeps its customers' best interest in mind
		The manufacturer makes good-faith efforts to address most customer concerns
Overall trust	Bhattacherjee 2002	Overall, the manufacturer is trustworthy

References

1. Strategy Analytics: Global smart home market roaring back in 2021. https://news.strategyanalytics.com/press-releases/press-release-details/2021/Strategy-Analytics-Global-Smart-Home-Market-Roaring-Back-in-2021/default.aspx (2021)
2. Wachter, S.: Normative challenges of identification in the internet of things: privacy, profiling, discrimination, and the GDPR. Comput. Law Secur. Rev. **34**, 436–449 (2018)
3. Kowalczuk, P.: Consumer acceptance of smart speakers: a mixed methods approach. J. Res. Interact. Mark. **12**, 418–431 (2018)
4. Kumar, V., Dixit, A., Javalgi, R.G., Dass, M.: Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing. J. Acad. Mark. Sci. **44**(1), 24–45 (2015)
5. Mani, Z., Chouk, I.: Drivers of consumers' resistance to smart products. J. Mark. Manag. **33**, 76–97 (2017)

6. Dominici, G., Roblek, V., Abbate, T., Tani, M.: Click and drive: consumer attitude to product development: towards future transformations of the driving experience. Bus. Process. Manag. J. **22**, 420–434 (2016)
7. Wilson, C., Hargreaves, T., Hauxwell-Baldwin, R.: Benefits and risks of smart home technologies. Energy Policy **103**, 72–83 (2017)
8. Badii, C., Bellini, P., Difino, A., Nesi, P.: Smart city IoT platform respecting GDPR privacy and security aspects. IEEE Access **8**, 23601–23623 (2020)
9. Ogury LTD: GDPR one year on: survey findings show consumer awareness with data use is concerningly low. https://www.prnewswire.com/news-releases/gdpr-one-year-on-survey-findings-show-consumer-awareness-with-data-use-is-concerningly-low-300855176.html (2019)
10. Proton Technologies AG: Do consumers know their GDPR data privacy rights? https://gdpr.eu/consumers-gdpr-data-privacy-rights/ (2019)
11. Raff, S., Wentzel, D., Obwegeser, N.: Smart products: conceptual review, synthesis, and research directions. J. Prod. Innov. Manag. **37**, 379–404 (2020)
12. Doffman, Z.: Cyberattacks on IoT devices surge 300% in 2019, 'measured in billions', report claims. https://www.forbes.com/sites/zakdoffman/2019/09/14/dangerous-cyberattacks-on-iot-devices-up-300-in-2019-now-rampant-report-claims/#8cf531a58926 (2019)
13. Bugeja, J., Jacobsson, A., Davidsson, P.: On privacy and security challenges in smart connected homes. In: Proceedings of the 2016 European Intelligence and Security Informatics Conference (EISIC), pp. 172–175. IEEE (2016)
14. Ho, G., Leung, D., Mishra, P., Hosseini, A., Song, D., Wagner, D.: Smart locks: lessons for securing commodity Internet of Things devices. In: Chen, X., Wang, X., Huang, X. (eds.) Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security, pp. 461–472. ACM, New York, NY (2016)
15. Bilton, N.: Nest thermostat glitch leaves users in the cold. https://www.nytimes.com/2016/01/14/fashion/nest-thermostat-glitch-battery-dies-software-freeze.html (2016)
16. Zheng, S., Apthorpe, N., Chetty, M., Feamster, N.: User perceptions of smart home IoT privacy. Proc. ACM Hum.-Comput. Interact. **2**, 1–20 (2018)
17. Kumar, N., Madhuri, J., ChanneGowda, M.: Review on security and privacy concerns in Internet of Things. In: Proceedings of the 2017 International Conference on IoT and Application (ICIOT). IEEE (2017)
18. Psychoula, I., Singh, D., Chen, L., Chen, F., Holzinger, A., Ning, H.: Users' privacy concerns in IoT based applications. In: Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation Conference. IEEE (2018)
19. Oluwafemi, T., Kohno, T., Gupta, S., Patel, S.: Experimental security analyses of non-networked compact fluorescent lamps: a case study of home automation security. In: Proceedings of the Learning from Authoritative Security Experiment Results (LASER), pp. 13–24 (2013)
20. Ronen, E., Shamir, A., Weingarten, A.-O., O'Flynn, C.: IoT goes nuclear: creating a ZigBee chain reaction. In: Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), pp. 195–212. IEEE (2017)
21. Kostkova, P., et al.: Who owns the data? open data for healthcare. Front. Public Health **4**, 7 (2016)
22. Wilkowska, W., Ziefle, M., Himmel, S.: Perceptions of personal privacy in smart home technologies: do user assessments vary depending on the research method? In: Tryfonas, T., Askoxylakis, I. (eds.) HAS 2015. LNCS, vol. 9190, pp. 592–603. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20376-8_53

23. Consumers International: Testing our trust: consumers and the Internet of Things 2017 review. https://www.consumersinternational.org/media/154746/iot2017review-2nded.pdf (2017)
24. Greverie, F., Buvat, J., Nambiar, R., Appell, D., Bisht, A.: Securing the Internet of Things opportunity: putting cybersecurity at the heart of the IoT. https://www.capgemini.com/wp-content/uploads/2017/07/securing_the_internet_of_things_opportunity_putting_cyber_security_at_the_heart_of_the_iot.pdf (2014)
25. de Hert, P., Papakonstantinou, V.: The new general data protection regulation: still a sound system for the protection of individuals? Comput. Law Secur. Rev. **32**, 179–194 (2016)
26. Furey, E., Blue, J.: Can I trust her? Intelligent personal assistants and GDPR. In: Proceedings of the 2019 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–6. IEEE (2019)
27. Fox, G., Tonge, C., Lynn, T., Mooney, J.: Communicating compliance: developing a GDPR privacy label. In: Proceedings of the Twenty-fourth Americas Conference on Information Systems, pp. 1–5 (2018)
28. Presthus, W., Sørum, H.: Consumer perspectives on information privacy following the implementation of the GDPR. Int. J. Inf. Syst. Proj. Manag. **7**, 19–34 (2019)
29. Paul, C., Scheibe, K., Nilakanta, S.: Privacy concerns regarding wearable IoT devices: how it is influenced by GDPR? In: Proceedings of the 53rd Hawaii International Conference on System Sciences, pp. 4388–4397 (2020)
30. Zhang, J., Hassandoust, F., Williams, J.E.: Online customer trust in the context of the general data protection regulation (GDPR). Pacific Asia J. Assoc. Informa. Syst. **12**, 86–122 (2020)
31. Karampela, M., Ouhbi, S., Isomursu, M.: Exploring users' willingness to share their health and personal data under the prism of the new GDPR: implications in healthcare. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2019, pp. 6509–6512 (2019)
32. Varkonyi, G.G., Kertesz, A., Varadi, S.: Privacy-awareness of users in our cloudy smart world. In: Proceedings of the Fourth International Conference on Fog and Mobile Edge Computing (FMEC), pp. 189–196. IEEE (2019)
33. Rodrigues, R., Wright, D., Wadhwa, K.: Developing a privacy seal scheme (that works). Int. Data Priv. Law **3**, 100–116 (2013)
34. Johansen, J., Fischer-Hübner, S.: Making GDPR usable: a model to support usability evaluations of privacy. In: Friedewald, M., Önen, M., Lievens, E., Krenn, S., Fricker, S. (eds.) Privacy and Identity 2019. IAICT, vol. 576, pp. 275–291. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-42504-3_18
35. Spence, M.: Job market signaling. Quart. J. Econ. **87**, 355–374 (1977)
36. Connelly, B.L., Certo, S.T., Ireland, R.D., Reutzel, C.R.: Signaling theory: a review and assessment. J. Manag. **37**, 39–67 (2011)
37. Spence, M.: Signaling in retrospect and the informational structure of markets. Am. Econ. Rev. **92**, 434–459 (2002)
38. Cook, L.: The role of third-party seals in building trust online. e-Serv. J. **2**, 71–84 (2003)
39. McKnight, D.H., Choudhury, V., Kacmar, C.: Developing and validating trust measures for e-commerce: an integrative typology. Inf. Syst. Res. **13**, 334–359 (2002)
40. McKnight, D.H., Kacmar, C.J., Choudhury, V.: Shifting factors and the ineffectiveness of third party assurance seals: a two-stage model of initial trust in a web business. Electron. Mark. **14**, 252–266 (2004)
41. Moores, T.: Do consumers understand the role of privacy seals in e-commerce? Commun. ACM **48**, 86–91 (2005)
42. Johnson, S.D., Blythe, J.M., Manning, M., Wong, G.T.W.: The impact of IoT security labelling on consumer product choice and willingness to pay. PLoS ONE **15**, e0227800 (2020)

43. Emami-Naeini, P., Dheenadhayalan, J., Agarwal, Y., Cranor, L.F.: Which privacy and security attributes most impact consumers' risk perception and willingness to purchase IoT devices? In: Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP), pp. 1937–1954. IEEE (2021)
44. Rifon, N.J., LaRose, R., Choi, S.M.: Your privacy is sealed: effects of web privacy seals on trust and personal disclosures. J. Consum. Aff. **39**, 339–362 (2005)
45. Shankar, V., Urban, G.L., Sultan, F.: Online trust: a stakeholder perspective, concepts, implications, and future directions. J. Strat. Inf. Syst. **11**, 325–344 (2002)
46. Walker, K.L.: Surrendering information through the looking glass: transparency, trust, and protection. J. Public Policy Mark. **35**, 144–158 (2016)
47. Coughlan, T., Brown, M., Mortier, R., Houghton, R.J., Goulden, M., Lawson, G.: Exploring acceptance and consequences of the Internet of Things in the home. In: Proceedings of the 2012 IEEE International Conference on Green Computing and Communications, pp. 148–155. IEEE (2012)
48. Princi, E., Krämer, N.: I spy with my little sensor eye: effect of data-tracking and convenience on the intention to use smart technology. In: Proceedings of the 53rd Hawaii International Conference on System Sciences (2020)
49. Lu, H.-P., Hsu, C.-L., Hsu, H.-Y.: An empirical study of the effect of perceived risk upon intention to use online applications. Inf. Manag. Comput. Secur. **13**, 106–120 (2005)
50. Kim, D.J., Ferrin, D.L., Rao, H.R.: A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. Decis. Support Syst. **44**, 544–564 (2008)
51. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Acad. Manag. Rev. **20**, 709–734 (1995)
52. Bhattacherjee, A.: Individual trust in online firms: scale development and initial test. J. Manag. Inf. Syst. **19**, 211–241 (2002)
53. Obar, J.A., Oeldorf-Hirsch, A.: The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services. Inf. Commun. Soc. **23**, 128–147 (2020)
54. heyData GmbH: Europe in the data protection ranking. https://en.heydata.eu/europa-im-datenschutz-ranking (2021)

Artificial Intelligence and Machine Learning in E-Government Context

Comparative Analysis of Classification Algorithms Applied to Circular Trading Prediction Scenarios

Douglas Silva(✉), Sergio T. Carvalho, and Nadia Silva

Informatics Institute, Federal University of Goias, Goiania, GO, Brazil
{douglasbernardes,sergio,nadia}@inf.ufg.br

Abstract. Goods and services trading taxation is the main source of revenue for Brazilian states, therefore its evasion directly affects public services. A particular case of evasion concerns the issuance and use of cold invoices – ones referring to transactions registered at Tax Administration, but which did not actually take place. Following the proposal by Mathews et al. [16], this work reports the application of classic supervised learning algorithms to identify circular trading behaviors involving taxpayers from Brazilian State of Goias, through the analysis of their goods and services trading operations. Experiments showed similar results to the original ones, but pointing to k-Nearest-Neighbours (and not Logistic Regression) as the most accurate technique for this purpose – given Brazilian context's characteristics.

Keywords: Machine learning · Classifying algorithms · Circular trading · Value-added tax · Statistical analysis

1 Introduction

Taxation on goods and services trading operations is the main source of revenue for states and provinces worldwide [14] – and in Brazil it accounts for almost 90% of state revenue [10]. It is considered an indirect tax, once it is embedded in the amount paid by consumer for product/service taken, but it is not passed on to Government directly by him, yet by the companies that were part of the consumption chain – e.g., the producer of the raw material, item's manufacturer, the company responsible for its distribution, and finally the retail company responsible for selling and delivering it.

It relies on an universal formula which indicates that tax due to a taxpayer is basically the output tax received (on selling/provisioning operations) minus the input tax paid (on purchasing/acquiring operations) [12].

It is a non-cumulative tax, due proportionally to each taxpayer that compose the consumption chain – known as a taxpayer. Therefore, each taxpayer is only responsible for passing along the tax related to what it "adds" to final value of each product or service. At the end of tax period, usually a calendar month, he

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 95–109, 2022.
https://doi.org/10.1007/978-3-031-12673-4_7

declares and pays all tax that fell to him in that period, based on a universal formula: due tax is basically the sum of tax that fell to him on sales/provisioning operations, minus the sum of tax that was due to it by the purchase/consumption operations it carried out in the same period [12] – which goes as a kind of "tax credit". This credit exists to ensure that tax is not charged twice (first one embedded in product's purchase, paid by the one who sold it, and second one in the sale/provision resulting from this acquisition, to be paid by the second taxpayer, now in the role of seller/provider, but who had been the buyer/taker in the first transaction).

Some taxes of this nature are *sales tax* in the United States [4], *ICMS* (that stands for *Tax on Circulation of Goods and Services*) in Brazil, *GST (Goods and Services Tax)* in countries like Australia, Canada, Singapore and recently India [14,17], and variations of *VAT (Value-Added Tax)* – that is used in most countries, like China [28] and European Union. Even though with some peculiarities, they share the same characteristics and non-cumulative character – and consequently their collection faces common difficulties.

Trading goods and services generates, at each transaction, one or more tax documents with a complete record of the items and parties involved in the transaction [28], including their tax classification and due tax rate. This data is usually registered on an electronic *invoice*, which – once verified and authorized by Tax Administration – has even legal validity.

As all sales of goods and provision of services are taxed, and each operation generates at least one tax document, volume of generated data grows daily and exponentially [22] – and it can be used for both good and bad purposes. If, on one hand, data analysis can favor proposition of public policies (e.g. on granting or not a tax benefit to a certain segment of economy), on the other hand, the massive volume of data – often incomplete and inaccurate – can be used by taxpayers to mask malicious attempts of tax evasion.

Evasion occurs when any action by the taxpayer – intentionally or not – leads to due tax non-collection by the Public Administration. Brazil, for example, has one of the highest tax burdens in the world, being considered disproportionate by many taxpayers (according to the Brazilian Institute of Planning and Taxation, in 2021 Brazilians worked for about five months just to pay taxes) [26]). In order to minimize impact on their income and wealth, taxpayers often look for illegal ways for not collecting all tax due to them [11].

As tax evasion generates damage to public coffers and directly impacts the provision of services to citizens, regardless of taxpayer's intent, it is essential to identify and mitigate it as soon as possible. Therefore, techniques that allow analysis of the huge volume of generated tax data can lead to identification of patterns that deviate from taxpayers' expected behavior and consequently indicate possible attempts of evasion. They can even provide not-so-intuitive analysis and findings, which business experts often did not even consider when looking for undue taxpayer behavior.

It turns out that fraud and tax evasion are only revealed when one of the companies involved in the sale process is audited or investigated [5]. An audit

procedure, however, usually only takes place at the end of a tax period, if there is any sign of misbehavior. This means that fraudulent action has already taken place, there has already been financial loss, and from then on it is only up to try recovering such amounts. In contrast, the appropriate approach would be to anticipate fraudster's action, foreseeing behaviors that indicate a possible fraud before there is, effectively, damage to the treasury.

A particular and interesting case found in literature is that of multiple circular operations (also known as *circular trading*) [12,13,15,16,18,19]. According to a systematic mapping study carried out in 2021 [23], it has been quite recurrent in research associated with tax evasion in the trade of goods and services. Furthermore, it also affects financial market [8], in a context similar to the one described here.

Circular trading occurs when taxpayers carry out false transactions in a circular manner (A sells to B, B sells to C, and C sells back to A, but always with *cold* invoices, without product/service deliver, just to simulate revenue and generate VAT credit).

Invoice is called *cold* when registered operation is not effectively carried out, i.e., invoice is issued, but no product is delivered or service is provided. It is a "simulated" transaction, with shady interests – for simulating revenue, simulating transaction flow between companies, or even providing undue tax credit to the taxpayer indicated as the buyer on the invoice, for example.

It happens that in Brazil, unlike what happens on VAT context, ICMS' administration is regionalized. This means that States have autonomy to, among other issues, specialize tax law, defining rules that are particularly convenient for them. It also turns local the management of main registration data and data from internally originated transactions. All these characteristics make it difficult to identify suspicious actions, such as circular trading, since such data are not usually shared between States and tax evaders take advantage of this, generally issuing cold invoices to companies located in a different State than theirs.

Methods traditionally used to solve aforementioned problems involve an extensive use of auditing, in which a business specialist manually observes reports or individual transactions in an attempt to discover unexpected or fraudulent behavior. Such a method is time-consuming, costly and imprecise, and in *big data* scenario it is impractical.

Machine learning algorithms and techniques then open up a huge range of possibilities. And the first step on it is to understand how algorithms used to identify suspicious behavior would behave with this regionalized data – and what kind of insights this analysis would be able to bring.

Following this line, our research reproduced the experiments conducted by Mathews et al. [16] to understand whether circular trading behaviors also manifests itself in the trade of goods and services in Brazil, and whether it is possible to predict them before they generate greater damage to public treasure.

Remainder of the article is organized as follows. Section 2 presents its theoretical references, showing classic supervised learning techniques used in the experiment. Section 3 presents research methodology, while Sect. 4 describes the

original experiment and its conclusions. Section 5 then presents our results when reproducing the experiment using data from transactions carried out in the State of Goias, in Brazil, and a comparative analysis of these results. Finally, Sect. 6 presents final considerations of this work.

2 Classical Supervised Learning Methods

Machine learning comprises computational methods and techniques that change aspects of its behavior as experience is acquired [20]. An analytical model is conceived, up to a certain point, with the definition of parameters that can be adjusted during its execution, while *learning* is considered to be the execution of an algorithm that optimizes these parameters based on training data and previous experience. Model can be predictive, when making predictions about future behaviors, descriptive, when acquiring knowledge from data, or both.

Machine learning uses statistical theory to build mathematical models, since its primary task is to make inferences from sampled data. This field of research works on two fronts: first, in training, to produce efficient algorithms that solve the optimization problem, as well as to store and process the huge amount of available data. Second, once a model is learned, it acts on its representation and algorithmic solution for inference, which also need to be efficient. In certain applications, learning algorithm efficiency – measured by its space and time complexity – can be as important as its predictive accuracy.

Machine learning techniques are generally classified into four categories or learning paradigms, depending on the nature of training data and feedback data available to the system.

In supervised learning, for every available training pattern (or set of characteristics) there is a desired known response – as known as label [1]. In other words, the model is trained with examples of inputs and their desired outputs, provided by a "teacher" or "supervisor", and the goal is to learn a general rule that maps inputs to outputs [25]. When training is performed, the model identifies which features are similar to each other – in patterns that have the same label – and associates these features with that label, assigning it to new patterns that show to be suchlike.

Unlike supervised learning, in unsupervised learning there is no desired output associated with each pattern, so the data is not pre-labeled. Therefore, there is no role of "teacher" or "supervisor", and only the input data is delivered to the model. The model is then expected to be able to capture, represent or express properties existing in the dataset – and propose the classes itself [1]. Its main interest is to identify how patterns (or data) can be organized into groups (or *clusters*) that express their similarity.

There is also semi-supervised learning, a middle ground between the two previous paradigms. Thus, for the construction of the model, both labeled and unlabeled data [7] are used. In other words, the "teacher" or "supervisor" provides a training set with some (often many) of the target outputs missing.

Finally, there is also reinforcement learning. In this paradigm, although it is not possible to indicate the correct output for each data (as in the supervised case), algorithm has access to information about the quality of the output produced by the model, in the form of a reward or punishment signal [25]. Reinforcement learning algorithms try to find a policy that maps the states of the environment to the actions that the agent must perform in those states, in order to maximize some notion of long-term reward.

The experiment reproduced in this research uses classical techniques of supervised learning aiming to classify transactions in the trade of goods and services into two classes: those that potentially form circular trading, and those that do not. The techniques evaluated in the experiments are described below.

2.1 KNN

k-Nearest-Neighbours (KNN) is a non-parametric classification method that analyzes a sample and classifies it based on the label of its k nearest neighbors. The majority label among the k data records closest to the sample (i.e. in its neighborhood) is generally used to decide its classification – whether or not it may weight the analysis based on the distance between them [6].

In a way, KNN's results ends up being skewed by k. There are several criteria for its choice, but conventionally the algorithm is executed several times, with different values of k, and then the one with the best performance is used.

2.2 SVM

A *Support-Vector Machine* (SVM) proposes to establish a decision limit between two classes, in order to enable prediction of labels from one or more feature vectors [9]. This decision boundary, known as hyperplane, is oriented in such a way that it is as far away as possible from the closest data points for each of the classes. These closest points are called support-vectors, and the distance between the hyperplane and the first point of each class is commonly called margin.

Initially, SVM separates the classes according to the hyperplane, and defines the points that belong to each one of them. It then maximizes the margin by setting its distances. When a new sample is given, it uses the feature vectors to classify it, analyzing the points of each class and disregarding the *outliers*.

2.3 Random Forest

A *Random Forest* is a classification method that consists of a collection of tree-structured classifiers, being these trees' growth given by a set of identically distributed and independent random variables. Each tree finally casts a unit vote for the class it considers most popular for the input sample, based on its own random vector, and vote count indicates to which class the sample will be assigned by the model [3].

2.4 Logistic Regression

Logistic Regression is a statistical classification method that aims to predict the class of an input sample from categorical, continuous or binary variables, derived from samples used during training. It proposes a regression model based on binomially distributed *response variables*, which models the probability of an event to occur as a function of the characteristics extracted from the training set – also known as prediction variables [24] .

3 Methodology

The research carried out through this experiment consisted of an exploratory and experimental study, with a quantitative approach and applied research, conducted in a laboratory environment [21].

We selected an article entitled *Link prediction techniques to handle tax evasion* [16], obtained in the systematic mapping study carried out in 2021 [23], to reproduce the experiment proposed in it. The objective was to analyze possible behaviors of circular trading in the Brazilian context – given that circular trading is recurrently associated with tax evasion in the European scenario, but quite unnoticed in the context of ICMS in Goias. More specifically, the selected article proposes the use of supervised learning algorithms to indicate, based on goods and services trading operations' characteristics and some statistical variables resulting from them, which transactions have the potential to form, in the future, a collusion of circular trading.

We chose to initially reproduce the experiment using a methodology called *Simplicity-first* [27], when analyzing practical datasets. The methodology's idea is to carry out implementation with the least possible number of changes to the original modeling, even if it initially seems that important data are being disregarded – and thus, allowing a more appropriate comparative analysis with the baseline.

In this way, we extracted data with the exactly same characteristics as the dataset used in the original experiment, but from transactions carried out in the State of Goias. We then followed the steps proposed by the article for generating statistical variables from the original data and using them on the experiment.

Finally, after reproducing the experiment following article's guidelines, we added the cross-validation method to the implementation, providing greater reliability to the results by avoiding *overfitting* [2].

4 Original Experiment

According to the aforementioned systematic mapping, one of the recurring behaviors involving tax evasion in the trade of goods and services, within what has been researched by Data Science, is *Circular Trading*.

As mentioned, circular trading occurs when taxpayers carry out false transactions in a circular manner (A sells to B, B sells to C, and C sells to A, but

always with cold invoices, with no product/service delivered), just to increase business turnover of involved companies or to mask some type of fraudulent behavior. It is illustrated in Fig. 1.

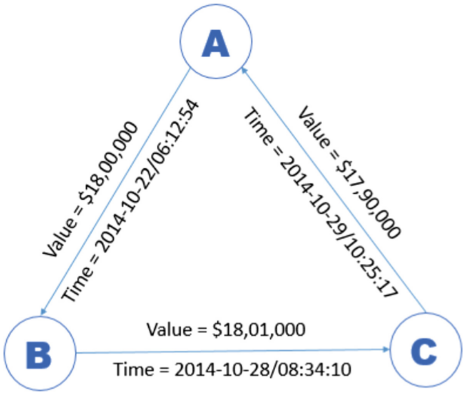

Fig. 1. Example of circular trading [obtained from [16]].

As can be seen in Fig. 1, relationships between taxpayers are mapped as a graph, where vertices represent the involved taxpayers and edges represent carried out transactions – being directed just like transaction itself, and weighted according to the tax amount highlighted in each transaction.

Circular trading has characteristics similar to what is observed in Goias when using ICMS credit from invoice-only companies. An *invoice-only company* is the one opened with the sole purpose of issuing cold invoices and generating false ICMS credit and which, when identified as a fraudster, is simply abandoned. As it is usually constituted in the name of an intermediary (called in Brazil a "orange" partner), there is nobody effectively responsible for the damage to public treasury, even if there is a drawn up notice of infraction.

In this scenario, the article proposes the extraction of the following data from tax documents that recorded commercial transactions carried out in 2015 in the state of Telangana, India:

No.	Seller's ID	Buyer's ID	Time	Value in $
1	a	b	2012/11/10/11:23:00	20000
2	c	d	2012/11/10/11:06:00	30000
3	d	b	2012/11/10/10:08:00	19000
4	m	n	2012/11/10/09:09:10	17000

Fig. 2. Sales database [obtained from [16]].

Although an invoice has a greater level of detail than that shown in Fig. 2, for the purpose of this experiment only these four data are used: unique identifiers for

seller and buyer, date and time when transaction was registered and total amount of tax due for that transaction – each one with its relevance on identifying a possible circular trading.

The sales records indicated in Fig. 2 are then mapped, allowing the analysis of transactions that result from each one – and, thus, identifying possible occurrences of circular trading.

Its called a *triad* each tuple formed by three vertices – the two that compose the original transaction analyzed, in terms of Fig. 2, and a third vertex that would be the receiver (or buyer) of a transaction arising from the original transaction – that is, a transaction that started at the receiver of the original transaction.

This concept is illustrated in Fig. 3. For each taxpayer in this condition (being receiver of a transaction arising from the original one), a tuple is generated. When, in the original dataset, there is a transaction that leaves this third vertex towards the original seller, from the first transaction (that is, if there is a transaction from u to v), it is deemed that the triad is formed and circular trading took place.

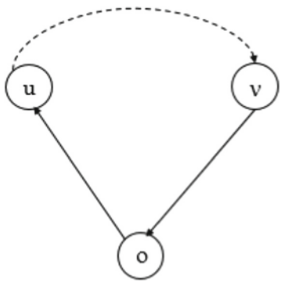

Fig. 3. Triad *uvo* [obtained from [16]].

Figure 4 illustrates the result of this new dataset, now composed by potential triads – derived from the original dataset.

No.	u	v	o	JC	FR	TC	IV	PPR	Link
1	a	b	c	0.1176	6.0000	4	0.7056	0.16789774	1
2	x	y	z	0.0714	2.0000	0	0.1428	0.01175115	0

Fig. 4. Database of potential triads obtained from transactions between taxpayers [16].

For each potential triad composed by taxpayers u, v and o, we calculate five statistical variables derived from the flow and volume of transactions carried out by them – within the triad or not. They are:

– *Jaccard coefficient (JC)*, based on the volume of neighbors between two vertices;

- *Flow Ratio (FR)*, a ratio between volume of transactions and volume of transacted tax credit;
- count of *Two Cycles (TC)*, based on the volume of purchases and sales transacted between a pair of vertices;
- *Interaction Variable (IV)*, obtained from the product of JC and FR;
- *Personalized PageRank (PPR)*, a customization of the PageRank algorithm that suggests that a taxpayer who carries out circular trading tends to relate to others who also do so.

Finally, if there is, in original dataset, a transaction that takes origin at the third vertex and is destined to the original seller, closing the triad – that is, if there is a transaction from u to v –, tuple is labeled with $Link = 1$. If this transaction does not exist, tuple is labeled with $Link = 0$.

Such statistical variables, along with the label (0 or 1) assigned to them, are then submitted to supervised classification models – and the results presented in the article are shown below.

4.1 Original Results

For validating and evaluating the proposed method, four classic classification models were implemented: KNN *(K-Nearest Neighbors)*, *Random Forest*, SVM *(Support Vector Machine)* and Logistic Regression. Among these, only Logistic Regression model's results were presented (in detail). According to the authors, it obtained the best accuracy among the four classifiers – as can be seen in Fig. 5.

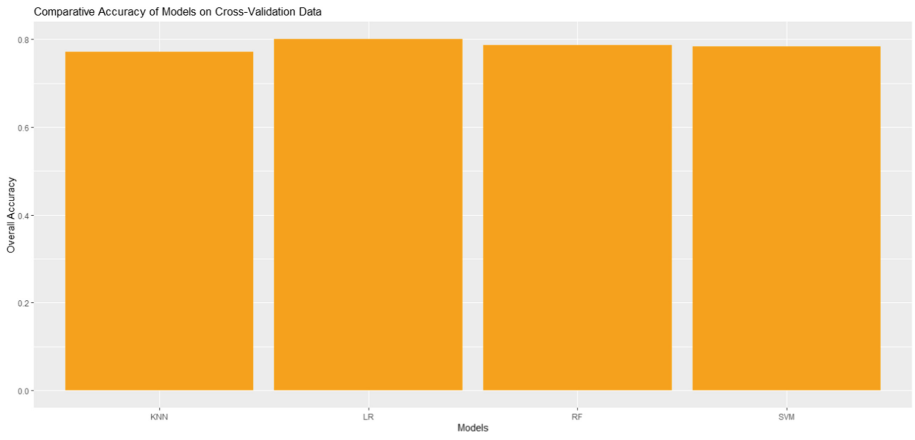

Fig. 5. Comparison chart - models [obtained from [16]].

In order to evaluate model's performance, its accuracy, precision, *recall* and F1-score were calculated. All these parameters were calculated for both training and test datasets, as shown in Table 1. As can be seen, the accuracy in the test dataset was 80%.

Table 1. Original experiment - logistic regression performance evaluation.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Train	80.9	85.41	74.54	79.60
Test	80.0	83.83	74.10	78.66

5 Experiment Reproduction

5.1 Data Source - ICMS Transactions in the State of Goias (Brazil)

The experiment was reproduced using data from electronic invoices issued in Goias between December 29 and 31, 2019. Regarding the data provided in Fig. 2, following ones were selected:

- as *Seller's ID*, the State Enrollment (local registration ID) of the invoice sender – which is the seller of the product or service provider;
- as *Buyer's ID*, the State Enrollment (local registration ID) of the invoice receiver – who is the product purchaser or service taker;
- as *Time*, invoice issuing date and time;
- as *Value in $*, the total ICMS amount highlighted on the invoice.

This query obtained 272,309 results, which were then used when conducting the experiment.

5.2 Conducting the Experiment

Some particularities were observed on data extracted from the brazilian invoices. Firstly, State Enrollments of companies domiciled outside the State (generally, senders of obtained invoices) are not usually informed, or are informed with a generic value *999999999*. As we are using this data as a taxpayer identifier, this would bias the model, once algorithm could consider any taxpayer with a State Enrollment equal to *999999999* as if they were the same company, even if they are domiciled in different federative units and have no affiliation. Thus, the first preprocessing performed was to ignore transactions (original or resulting) whose participant (sender or receiver) had a State Enrollment equal to *999999999*.

Second, Brazilian tax law allows, in some cases, companies to issue an invoice to themselves (including ones with ICMS credit highlighted). This is what happens, for example, when merchandise leaves a store for street sales, and returns to the establishment at the end of the day. Both operations give rise to the issuance of outgoing and incoming invoices, respectively. Thus, the second preprocessing was responsible for ignoring transactions (original or resulting) whose participants were the same taxpayer.

Thirdly, it was necessary that, following articles guidelines, the original, arising and resulting transactions (the one that returns to the original point, closing the triad) should be executed in chronological order. Thus, a tuple should be

disregarded as a potential triad if there was a flow of transactions between the three companies, but not if they had occurred in a chronological sequence.

Finally, as the analysis of transactions arising from the originals always analyzes all possibilities, triads containing the same taxpayers could be found. For example, a triad where $u = \text{A}$, $v = \text{B}$ and $o = \text{C}$, and another where $u = \text{B}$, $v = \text{C}$ and $o = \text{A}$, would be found at different circumstances – but they would still be the same triad (since they are the same taxpayers, with transactions in the same chronological sequence). Thus, the preprocessing algorithm should ignore equivalent tuples, in order to avoid biasing the training set.

Considering all these aspects, and the invoices issued in the period, the training set containing potential triads was generated and submitted to the four classic classification models mentioned above.

The preprocessing algorithm and the classification models were all implemented in Python programming language.

5.3 Experiment Results

When executing the preprocessing algorithm, considering all the exclusion criteria mentioned in Sect. 5.2, for every 10,000 invoices analyzed, only 1,690 were effectively analyzed, on average. These 1,690 transactions, however, generated 148,838 tuples for the training set, of which 289 triads were confirmed.

As performed in the article, we implemented, for validation and evaluation of the proposed method, the four classic classification models: KNN, *Random Forest*, SVM and Logistic Regression. Unlike the article, however, Logistic Regression presented the worst performance, with 51% accuracy over the test dataset, as show in Table 2. The best performance was that of KNN, with approximately 90% accuracy and an average score of 98.9% in the cross-validation with 10 groups.

Table 2. Predictive performance comparison of different Machine-Learning algorithms in scenarios of circular trading.

Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LR - baseline	80.0	83.83	74.10	78.66
KNN	**90.0**	**91.0**	**89.0**	**90.0**
SVM	71.0	97.0	61.0	67.0
Random Forest	47.0	49.0	50.0	50.0
Logistic Regression	53.0	59.0	50.0	51.0

We validated the model with best performance – in our case, the KNN – using the two metrics: the ROC Curve (Receiver Operating Characteristic Curve) and the Precision-Recall Curve.

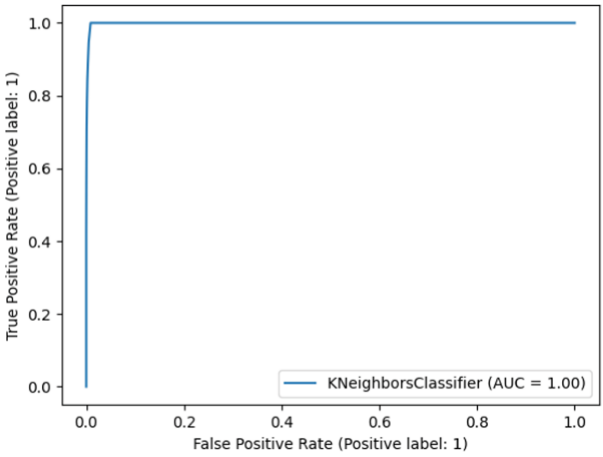

Fig. 6. ROC curve for train data.

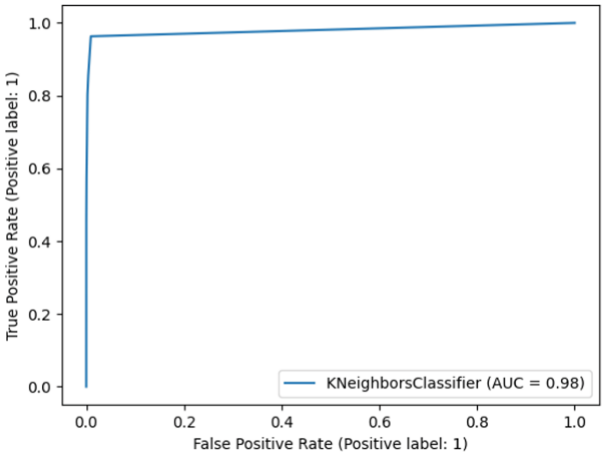

Fig. 7. ROC curve for test data.

As shown in Figs. 6 and 7, the area under the train dataset ROC curve is 1.0 and the test dataset ROC curve is 0.98. Since ROC curves are close, we can conclude that the model is not over fitting. Since the area under the train dataset ROC curve is more than 0.70, one can say that model is not under fitting.

Precision-Recall Curve, on the other hand, shown in Fig. 8, is a useful measure of prediction success when the classes are very imbalanced – which is our case in Goias. It shows the trade-off between precision and recall for different threshold. As a high area under the curve represents both high recall and high precision, we can assume that the classifier is returning accurate results, as well as returning a majority of all positive results – even though potential-edge dataset is clearly unbalanced.

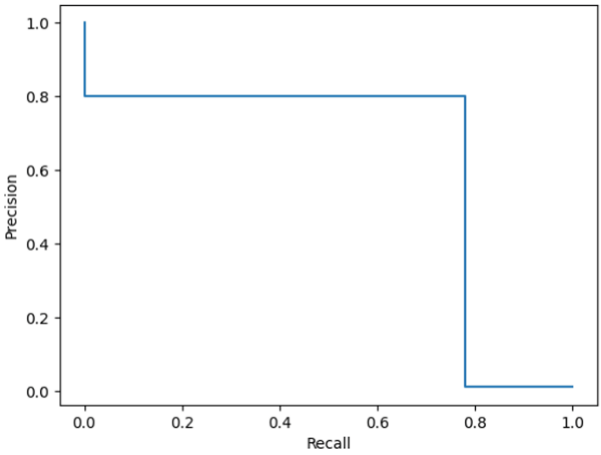

Fig. 8. Precision-recall curve.

6 Conclusions

This work sought to report an experiment carried out to verify how the ICMS data from Goias would behave in learning models proposed in the literature, and then compare obtained results with original ones presented it the article. The objective was to assess the feasibility of using data of taxpayers from Goias, Brazil – and their operations in the trade of goods and services – in identifying possible fraud/tax evasion behaviors.

The results were interesting, since one of the classifiers had an accuracy close to 90% with data from Goias. However, this was achieved with the KNN model, while the original experiment had its best performance with the Logistic Regression one – which in our case had the worst performance.

In principle, as the statistical variables analyzed to identify potential circular trading are inferences out of the flow and behavior of transactions between taxpayers, the performance of the evaluated techniques tends to be directly associated with their adjacency to the behavior of taxpayers – while goods trading – in each context. In other words, taxpayers relate differently in each case, so the problem ends up being better handled by the technique that best model the relationships between them – in each situation.

With regard to Brazilian public administration, however, more important than which technique showed best performance is the fact that a possible new indication of tax fraud was found that, apparently, had not yet been taken into account by the Brazilian Tax Administration – at least in the State of Goias, and that there are machine learning techniques capable of pointing out this behavior from the analysis of tax documents already available, with a very significant accuracy.

As the approach only uses data from the most essential tax documents (the invoices, which record transactions themselves), the analysis can be performed when invoice authorization is requested, and not just after the cold transaction had already taken place – and may even conditioning the authorization itself. Therefore, when a taxpayer requests Tax Administration to issue an invoice for the transaction that is being carried out, the system responsible for this request may first check the machine learning model, seeking to identify whether the characteristics of entities involved in the transaction – and the history of transactions already carried out – indicate a possible behavior of circulating trading and, from there, suspending (for evaluation) or even prohibiting the authorization of the invoice. Thus, it prevents a false transaction from taking effect – and its consequences to manifest – before it causes damage to public treasury and prejudices services provided to citizens.

References

1. Abbass, H.A., Sarker, R., Newton, C.S.: Data Mining: A Heuristic Approach: A Heuristic Approach. IGI Global (2001)
2. Berrar, D.: Cross-validation (2019)
3. Breiman, L.: Random forest. Mach. Learn. **45**(1), 5–32 (2001)
4. Buxton, E., Kriz, K., Cremeens, M., Jay, K.: An auto regressive deep learning model for sales tax forecasting from multiple short time series. In: 18th IEEE International Conference on Machine Learning and Applications, pp. 1359–1364 (2019)
5. Junqué de Fortuny, E., Stankova, M., Moeyersoms, J., Minnaert, B., Provost, F., Martens, D.: Corporate residence fraud detection. In: KDD 2014, pp. 1650–1659 (2014)
6. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K.: KNN model-based approach in classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds.) OTM 2003. LNCS, vol. 2888, pp. 986–996. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-39964-3_62
7. Han, J., Pei, J., Kamber, M.: Data Mining: Concepts and Techniques. Elsevier, Amsterdam (2011)
8. Hayashi, F.Y., et al.: Manipulação do mercado de capitais baseada em informações: análise do art. 27-c da lei nº 6.385/1976 sob a perspectiva da tutela constitucional da liberdade de expressão (2021)
9. Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., Xu, W.: Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics **15**(1), 41–51 (2018)
10. Jatobá, A.E., Moura, D.L., Martins, I.C., Ramos, H.S., Aquino, A.L.: CALT: uma ferramenta automática para cobrança do icms em operações interestaduais. In: IX Workshop de Computação Aplicada em Governo Eletrônico, pp. 143–154. SBC (2021)
11. Jordão, R.V.D., Silva, M.S., Brasil, H.G., Vasconcelos, M.C.R.L.: Um caleidoscópio de perspectivas sobre o sistema público de escrituração digital-sped. Rev. Contemp. Contabilidade **12**(26), 119–140 (2015)
12. Mathews, J., Mehta, P., Babu, C.S., Kasi Visweswara Rao, S.V.: An algorithmic approach to handle circular trading in commercial taxation system, pp. 67–75 (2018)

13. Mathews, J., Mehta, P., Sobhan Babu, C., Kasi Visweswara Rao, S.V.: Clustering collusive dealers in commercial taxation system. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) IntelliSys 2018. AISC, vol. 869, pp. 703–717. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-01057-7_54
14. Mathews, J., Mehta, P., Kuchibhotla, S., Bisht, D., Chintapalli, S.B., Rao, S.V.K.V.: Regression analysis towards estimating tax evasion in goods and services tax. In: 2018 IEEE/WIC/ACM WI, pp. 758–761 (2018)
15. Priya, Mathews, J., Kumar, K.S., Babu, C., Rao, S., et al.: A collusion set detection in value added tax using Benford's analysis. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) SAI 2018. Advances in Intelligent Systems and Computing, vol. 858, pp. 909–921. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01174-1_70
16. Mathews, J., Mehta, P., Suryamukhi, Babu, S.: Link prediction techniques to handle tax evasion. In: 8th ACM IKDD CODS and 26th COMAD, pp. 307–315 (2021)
17. Mehta, P., Mathews, J., Suryamukhi, K., Kumar, K.S., Babu, C.S.: Predictive modeling for identifying return defaulters in goods and services tax. In: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, pp. 631–637 (2018)
18. Mehta, P., et al.: Big data analytics for tax administration. In: Kő, A., Francesconi, E., Anderst-Kotsis, G., Tjoa, A.M., Khalil, I. (eds.) EGOVIS 2019. LNCS, vol. 11709, pp. 47–57. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-27523-5_4
19. Mehta, P., Mathews, J., Kumar, S., Suryamukhi, K., Sobhan Babu, C., Kasi Visweswara Rao, S.V.: Big data analytics for nabbing fraudulent transactions in taxation system. In: Chen, K., Seshadri, S., Zhang, L.-J. (eds.) BIGDATA 2019. LNCS, vol. 11514, pp. 95–109. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23551-2_7
20. Michalski, R.S., Carbonell, J.G., Mitchell, T.M.: Machine Learning: An Artificial Intelligence Approach. Springer, Heidelberg (2013)
21. Moresi, E., et al.: Metodologia da pesquisa. Brasília: Univ. Católica de brasília **108**(24), 5 (2003)
22. Rad, M.S., Shahbahrami, A.: High performance implementation of tax fraud detection algorithm. In: 2015 SPIS, pp. 6–9 (2015)
23. Silva, D., Carvalho, S.: Data science methods and techniques for goods and services trading taxation: a systematic mapping study (2021)
24. Speelman, D.: Logistic regression. Corpus Methods Semant. Quant. Stud. Polysemy Synonymy **43**, 487–533 (2014)
25. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (2018)
26. de Planejamento e Tributação, I.B.: Brasileiros trabalham 149 dias para pagar os impostos de 2021 (2021). https://ibpt.com.br/brasileiros-trabalham-149-dias-para-pagar-os-impostos-de-2021/. Accessed 16 Feb 2022
27. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining: Practical Machine Learning Tools and Techniques, vol. 3 (2011)
28. Yu, J., Qiao, Y., Shu, N., Sun, K., Zhou, S., Yang, J.: Neural network based transaction classification system for Chinese transaction behavior analysis. In: 2019 IEEE International Congress on Big Data (BigDataCongress), pp. 64–71 (2019)

Hybrid AI to Support the Implementation of the European Directive

Monica Palmirani(✉)

CIRSFID, University of Bologna, Bologna, Italy
monica.palmirani@unibo.it

Abstract. This paper presents a use case of hybrid AI approach applied to the European legislation with the aim to detect and analyse the similarity between some European Directives and the corresponding implementation in the domestic law of the Member States. The approach uses Akoma Ntoso XML standard for supporting the comparison between the different articles. We have conducted this research inside of the project 'Drafting legislation in the era of AI and digitisation' with the support of the EU Commission – Directorate General Informatics Unit B2 – Solutions for Legislation, Policy & HR.

Keywords: Hybrid AI · Akoma Ntoso · Visualization · Implementation of European Directive

1 From LegalXML Standards to AI for Legislative Process

In the last two decades the legal informatics community, using multidisciplinary methodologies, elaborated relevant outcomes that provide solutions for the modelling legal knowledge in the vision of the Semantic Web framework (Casanovas et al. 2016), Open Government Data (Casanovas et al. 2017, Francesconi 2018), Free Access to Law Movement[1] (Greenleaf 2011). Also, the Official Gazettes moved to digital format[2] with a deep transformation of the Legal Sources paradigm. The LegalXML community developed different XML standards (e.g., AkomaNtoso, AKN4UN, AKN4EU; Palmirani 2011, Palmirani and Vitali 2011) for structuring legal texts, metadata legal model (e.g., RDF models for legal domain, like ELI), legal rule modelling languages (e.g., LegalRuleML; Palmirani et al. 2011), URI naming conventions that are persistent over time (e.g., ELI/ECLI), while enhancing legal reasoning through the literature in AI and Law. Machine learning and Legal Analytics extract legal knowledge and predictive models, while legal design proposes new pattern for smart visualization (Ashley 2017; Verheij 2020). The LegalXML approach ranges from the legal official text, approved by institutions (e.g., Parliament, government), to its formal modellization using XML, logic programming and functional coding.

[1] https://www.nyulawglobal.org/globalex/Legal_Information_Institutes1.html.
[2] https://op.europa.eu/en/web/forum_official_gazettes/home.

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 110–122, 2022.
https://doi.org/10.1007/978-3-031-12673-4_8

The MIT Computational Law Development Goals is a July 2020 initiative that aims to research new methods for making the law human-centred, measurable, and computable, machine-readable in the Semantic Web approach, interoperable thanks to international data-models and standards. In some cases, the legal drafting activity is enhanced by the mentioned technologies with the aim to support the law-making process and to favour the paperless creation, consolidation, publication, dissemination, access of the legal source of law using the Web or apps. Additionally, there is a large community of legal drafters that is investigating how to improve the process using RegTech tools in pragmatic ways (International Association of Legislation 2019 Seoul Conference proceedings, Micheler and Whaley 2019).

In 2018, the New Zealand Government started a project named "Rules as Code"[3] and in 2020 it proposed to OECD-OPSI (Cracking the code: Rulemaking for humans and machines; Greenleaf et al. 2020) to codify a new approach: the idea is to use the coding methodology (e.g., UML, flow chart, pseudo-coding) to create a macro-schema of law, legally binding, that as an output produces the legal text in natural language. It is a sort of a *reverse engineering* approach with respect to the predominant method, and it is backed by legal theory, where the digitalization is conducted from the legal text expressed in natural language to the formal-logic representation. Stanford CodeX lab[4], Australia and Canada governments[5] are investigating this new direction also using programming languages (e.g., Java, Python, C++, etc.).

This approach is very fascinating, however, it arises many research questions in the community of legal theory, philosophy of law, constitutional law, and untimely also in the legal informatics area that have dedicated the last 30 years to the Artificial Intelligence and Law analysis. The main questions are the follow:

i) the Law is not only rules, but it includes parts that are hardly reducible in static formulas (e.g., principles and values, Hildebrandt 2018);
ii) fix the norms in a monolithic coding formula does not permit flexible adaptation of the norms to the evolution of the society (open-textured Hart 1961);
iii) artificial languages are a subset of natural language (Chomsky 2006; Marmor 2014) so we need take in consideration this limit and to investigate if some other computational linguistics formal method are more effective in legal domain (Fillmore and Baker 2009);
iv) norms could be intentionally contradictory in order to balance different interests, institutions (*legal pluralism* for managing coexisting legal orders);
v) any prediction is based on the past, so it is limited in the detection of the new concepts or the autopoietic role of the legal language;
vi) any prediction influences the decision-maker and the future human behaviour (Hildebrandt 2020, Diver 2020);

[3] https://joinup.ec.europa.eu/collection/better-legislation-smoother-implementation/discussion/better-rules-and-rules-code-references-australia-nz-mainly; https://www.digital.govt.nz/blog/what-is-better-rules/.

[4] https://law.stanford.edu/projects/computational-linguistics-and-effective-legal-drafting/.

[5] Making Laws in a Post-Modern World: Are You Ready – CIAJ Legislative Drafting Conference https://ial-online.org/legislative-drafting-conference-making-laws-in-a-post-modern-world-are-you-ready-ottawa-september-2020/.

vii) the autonomy of the addressee of the norms is a fundamental element of normativity and lack of transparency of the code does not always permits to maintain this autonomy in the agent both in the creation of the norms (legislative process) and in the execution of the rule (Forst and Günther 2021).

2 Hybrid AI in Legal Domain

With these important research questions in mind, we believe that it is possible to have a good balancing between legal theory and benefits produced by the introduction of the ICT in the legislative process without affect the democratic principles. We propose to use the so-called *Hybrid AI* where *human-in-the-loop*, *human-on-the-loop*, and *human-in-command* principles[6] are combined with different complementary disciplines (law, philosophy, ethics), using symbolic and sub-symbolic AI techniques integrated with Semantic Web findings in order to add context and meanings to the pure *data-driven* or *code-driven* methodology. The *Hybrid AI* is very promising approach especially in legal domain where the context, values, and concepts are fundamental for correctly apply the AI outcomes (AICOL 2021, Fratrič et al. 2020) (Fig. 1).

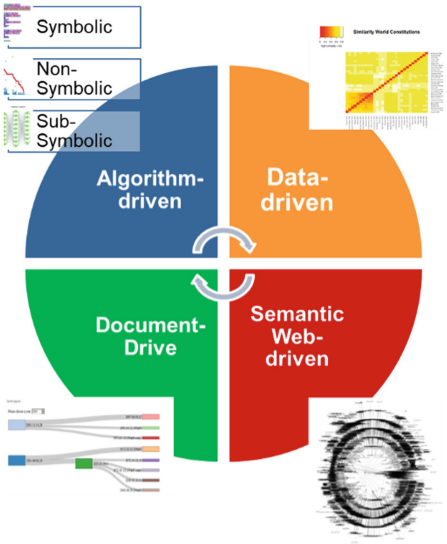

Fig. 1. Hybrid AI model.

Often the different legal informatics techniques are fragmented and each of them isolated could present limits: i) the data-driven approach is more oriented to the probabilistic approach which is based on data extracted by the documents. The description logic could deduct some not perfectly accurate assertions that is not recommendable in

[6] High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence, 2019.

the legal domain; ii) the non-symbolic algorithmic-approach is based on probabilistic methods (e.g., machine learning) that do not often include the semantics, it a good proxy but it should be integrated with symbolic representation (rule-based); iii) the document-oriented approach is oriented to model the structure the parts of the legal text and to data mining; iv) the semantic web approach is focused on capturing the concepts and the relationships. The Hybrid AI intends to use all these aspects together: symbolic AI with logic formalism, non-symbolic AI for extracting hidden legal knowledge from the legal text, document analysis for creating a network of relationships inside of the legal order, the semantic annotation of the meaning of the knowledge extracted and annotated.

Another fundamental element for guaranteeing the legitimacy of the whole *digital law-making process* is the metadata concerning the workflow (see Fig. 2).

Furthermore, we know about some specific critical technical issues that characterise the legislative domain when the AI non-symbolic alone could present problems. For this reason, we use LegalXML Akoma Ntoso standard as the background format of the legal document and on the top of this we add other AI solutions. We present here some examples of critical issues where AKN can help:

- *Granularity vs. Structure*: machine learning works at sentence level and this approach cannot link different parts of the legal speech semantically connected (e.g., obligation-exception, duty-penalty). For this reason, we also need a symbolic AI level (based on rules) for connecting the part of the legal reasoning. AKN provides the document structure to the machine learning;
- *Content vs. Context*: machine learning often works without the context (e.g., jurisdiction, temporal parameters) and can deduct something probabilistically correct insufficient to collocate the information in the correct semantic (e.g., the legal lexicon changes over time according to the evolution of the society, including the concept of European citizen which changed in the last ten years and the machine learning tends to compare similar terms). AKN provides the context to the machine learning;
- *Past vs. Future*: machine learning depends to the past data series (e.g., new brilliant solution has no historical series), so new concepts introduced with the law (e.g., smart working) are not known by the non-symbolic engine. AKN provides a quasi-ontology of concepts expressed in the text and using this information we could create a light ontology for supporting the checking of new emerging legal concepts (e.g., starting from the analysis of the definitions);
- *Internal vs. External information*: machine learning does not consider the normative and juridical citations (normative references), or better it recognises the sequence of characters but not the topic that this citation intends to inject in the legal reasoning. For this reason, AKN provides the correct link, based on permanent Unique Identifier, to the destination text;
- *Static vs. Dynamic*: The normative references evolve over time (e.g., art. 3 is not the same forever) and AKN provides a temporal model of the norms managing versioning and *point-in-time*. In this manner we are capable to discover the norms abrogated, suspended, postponed or retroactive and to use the legal knowledge extracted by non-symbolic AI in effective way.

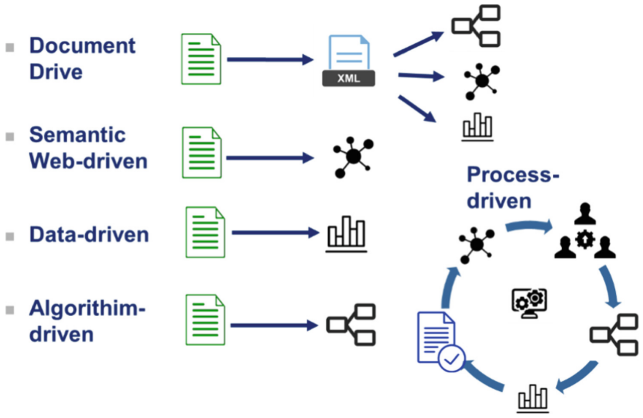

Fig. 2. Workflow based approach.

3 Drafting Legislation in the Era of Artificial Intelligence and Digitisation

The European Commission is recently providing a roadmap for the *digital-ready legislation*[7] with an interdisciplinary approach and it is investigating the "Drafting legislation in the era of artificial intelligence and digitisation" (workshop 2019)[8]. EU Commission, Directorate-General for Informatics is performing with the University of Bologna a study on "Drafting legislation in the era of artificial intelligence and digitisation" that includes three pilot-cases using AI techniques applied to support the legal drafting units. In this study we propose a third way (i.e., *Hybrid AI for Law*) with a legal and technical model for developing computable informatics legal systems compliant by-design (or *Legal Protection by-design* as Hildebrandt defined) with theory of law, intended in the autopoietic role to create new framework never seen before. Legal formalism and logic-positivism (reductionism and textualism), used for decades, are not sufficient for coding law resilient to the passage of time. There is the necessity to maintain flexibility to be applicable to different jurisdictions, context, historical periods, changes of the society. Neither the opposite radical legal hermeneutic nor subjectivism, used in the legal area, are good approaches for the Web of Data (Filtz et al. 2021). For this reason, this project is ground-breaking because nowadays the mentioned communities are silos, and nobody is interested to find a new innovative structure that conciliates legal theory/philosophy of law disciplines with emerging technologies that are deeply modifying the current society.

The application of the AI in legislative domain consists of:

- *Creation of Law*. AI for supporting the legislative drafting and the law-making process in the generative phase of the Law;

[7] https://joinup.ec.europa.eu/collection/better-legislation-smoother-implementation/digital-ready-policymaking.

[8] https://ial-online.org/wp-content/uploads/2019/07/Invitation-EN.pdf.

- *Application of Law*. AI for supporting the decision support process using the Law. We find in this field methods and tools for legal reasoning, checking compliance, trends prediction. These instruments are often applicable to a specific domain (e.g., privacy, taxation, contract law);
- *Monitor the Law*. Legal data analytics for discovering hidden knowledge in the legal system and to monitor the quality of the legal order as complex system.

Creation of Law
Drafting applications
• Assist the drafting activity with templates and patterns
• Management of the amendments of the documents
• Support the cooperative work on the document by internal and external actors
• Contextual drafting (e.g., definitions, acronyms)
Linguistic support
• Help the translation and the linguistic tasks with AI
• NLP tools for extracting legal knowledge (e.g., actors, events, frequent errors)
• Prevent un-fair language (e.g., gender balancing) or support some policy (e.g., privacy anonymization, digital ready act)
Application of law
• Legal Reasoning (e.g., detecting ex-ante legislative inconsistency during drafting);
• Compliance checking (e.g., ex-post analysis GDPR, license);
• Machine Learning Extraction (e.g., clustering, classification, extraction of situations) (e.g., derogations, obligations).
Monitor the Law
• Data analysis on the documents (e.g., network analysis);
• Predict future behaviours (e.g., predictive on bill).

4 Hybrid AI Supports the Transposition and Implementation of the Acquis

We have applied the Hybrid AI to several use-cases. One of these is to compare the transposition of some directives into Italian domestic legislation with the original directive to measure the relationships between the different articles and so identify where the two document diverge. Nanda et al. (2019) faced the same problem with unsupervised and supervised similarity techniques. In our approach we use also Akoma Ntoso knowledge for understanding the relationship between articles.

The dataset examined several directives. We focused our attention on Directive 2014/89/EU establishing a framework for maritime spatial planning.[9]

[9] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32014L0089.

4.1 Akoma Ntoso Serialization

We took the FORMEX 4.0 file of this directive from CELLAR database. We converted it into Akoma Ntoso using the Formex2AKN service.[10] We did the same extracting from Normattiva,[11] the Italian legislative portal, using the corresponding implementation Legislative Decree 201/2016,[12] and we converted it into Akoma Ntoso. We then extracted all the articles of the two documents using Xpath queries.

Art. 3 EU Directive 2014/89/UE	Italian Legislative Decree 201/2016
<article GUID="003"> <num>Articolo 3</num> <heading>Definizioni</heading> <list> <intro> Ai fini della presente direttiva si intende per: </intro> <point> <num>1)</num> <content> "<def eId="ref_1" refersTo="#politicaMarittimaIntegrata">politica marittima integrata</def>" </content> </point> <point>	<article eId="art_3"> <num> Art. 3 </num> <heading> Definizioni </heading> <paragraph eId="art_3__para_1"> <num> 1. </num> <list> <intro eId="art_3__para_1__intro_1"> Ai fini del presente decreto si intende per: </intro> <point eId="art_3__para_1__point_a"> <num> a) </num> <list> <intro eId="art_3__para_1__point_a__intro_1"> «<def eId="ref_1" refersTo="#acqueMarine">acque marine</def>» : </intro>

The Akoma Ntoso serialization allows to compare articles with each other, to also evaluate the qualified part of the legal text like definitions and normative citations (e.g., <def> and <ref>). In particular, the references to other norms are interesting information for verifying if the EU legislation is properly addressed in the Italian transposition.

The EU directive 2014/89 is cited three times in the Italian law: i) the first time is the legal basis citation that justify the transposition; ii) the second and third times the citations state the obligations of transparency defined by the EU Commission.

[10] bach.cirsfid.unibo.it/formex2akn-v2/

[11] https://www.normattiva.it/.

[12] https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2016-10-17;201.

Italian Legislative Decree 201/2016

```
<citations eId="cits_2">
        <citation eId="cits_2__cit_1">
          <p> Vista la <ref href="/akn/eu/act/directive/ep/2014/89/!main
" eId="ref_4">direttiva 2014/89/UE </ref> del Parlamento europeo e del Consiglio
del 23 luglio 2014 che istituisce un quadro per la pianificazione dello spazio
marittimo; </p></citation>
```

```
<point eId="art_8__para_2__point_b">
                <num> b) </num>
                <content>
                  <p> invia alla Commissione europea e agli altri Stati membri
copia dei piani di gestione dello spazio marittimo, compreso il pertinente materiale
esplicativo esistente sull'attuazione della
<ref  href="/akn/eu/act/directive/ep/2014/89/!main" eId="ref_64"> direttiva
2014/89/UE </ref>, entro tre mesi dalla loro approvazione, nonche' gli aggiornamenti
successivi dei piani entro tre mesi dalla pubblicazione; </p>
                </content>
```

The Akoma Ntoso serialization allows a more accurate analysis of the similarity and a better final interpretation of the results by the legal expert. He/she can access to the semantic annotation and so to identify some further explanatory information (e.g., references, definitions) for deducting the final conclusions.

4.2 Similarity Indexes

We created an experiment using KNIME, producing a Cartesian product between each article of the directive and each article of the national law: 17 articles of the EU directive combined for each of the 12 articles in the National Law, for a total of 204 rows. We created pairs that we fed to different similarity AI algorithms (e.g., Levenshtein, Jaro–Winkler, e-gram overlap distance, etc.) for measuring the distance between the articles. We then selected the pairs with the maximum value of similarity, creating a matrix. The matrix is visualized using different graphs connecting on one side the Member State's implementation number of the article and on the other side the article of the directive that has the highest similarity index. In the middle we find the similarity index. We can notice that we have a connection between art. 3 and art. 6 with the lower similarity index (see Fig. 3). Art. 3 in the Italian transposition act states the legal concept definitions with peculiar customization for Italian jurisdiction (e.g., the definition of the Mediterranean Region). The art. 6 of the EU Directive defines the minimal requirements that each member state should regulate according to its specific context (e.g., maritime coasts). The algorithm discovers that in art. 3 of the Italian act some definitions are related to the specific requirements described in the art. 6. Additionally, since art. 6 is a general high-level rule it is not connected to any other article of the Italian law. On the contrary, the art. 1 of both laws is very similar considering that they both describe the purpose of the regulation that evidently should converge otherwise the transposition is out of scope. Art. 7 relates to three articles of the EU Directive: art. 5, 10, 12. This is because art. 7 aggregates several points coming from the three different EU Directive articles. This analysis is very useful for lawyers, for the legislator, and for citizens, and additionally

to compare the different levels of compliance to EU Directives from several Member States. This tool could be very substantial for the EU institutions for monitoring the effectiveness of EU Directives and the different levels of its implementation in each Member State (Fig. 3).

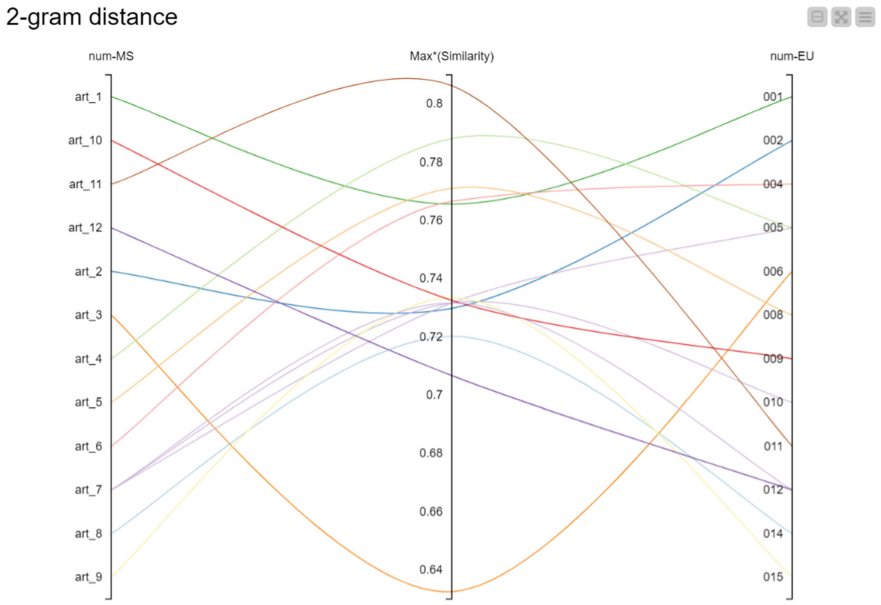

Fig. 3. 2-g Distance between the EU directive and its national implementation

Additionally, we correlated the similarity index (see Fig. 4) in relation of the articles. The blue colour means a higher similarity, red a low similarity. We noticed that there is a concentration of similarity around articles 5, 6, 7, and 8, but not on the final part of the Directive. This is quite normal, considering that a European directive presents the main principles and values in the first part of the provisional norms, and in the final part it usually provides only recommendations with less bindingness. In the last part of EU Directives, we find also norms devoted to delegate to the Member States' domestic law some specific topic to regulate. For this reason, some articles of the EU Directive are not connected with the national transposition.

Another example is Directive 2014/53/EU, implemented in Italy by Legislative Decree 128/20164.[13] Both documents, the Directive and its Italian implementation, are composed of 52 articles. In the following figure we can see the analysis of the similarity index using a Cartesian product. The index of similarity is calculated for each article of the Italian transposition in relation of each article of the 52 of the EU Directive. In this way we can select the max of the similarity index and thus find the best option of correlation. It is evident that large portion of the text converses with the original Directive:

[13] https://www.gazzettaufficiale.it/eli/id/2016/07/14/16G00137/sg.

the index of similarity falls for all the articles in the interval 0.80–0.95 (see the box of Fig. 5).

Similarly to the previous case, we can find that the most similar part of the implementation is on the first part of the normative provisions, and the last articles do not find

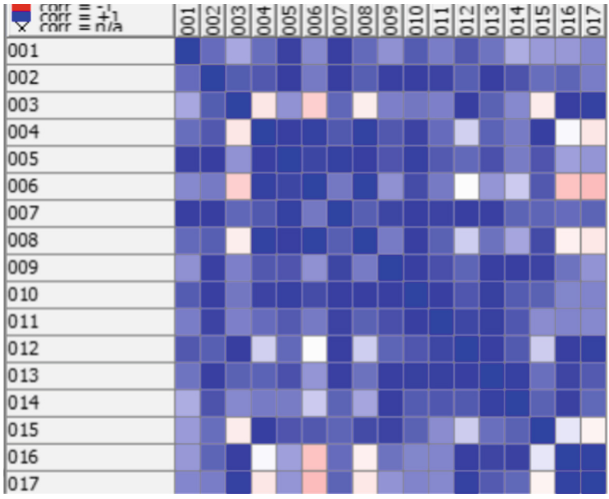

Fig. 4. Linear correlation between the similarity index related to the articles of the EU directive

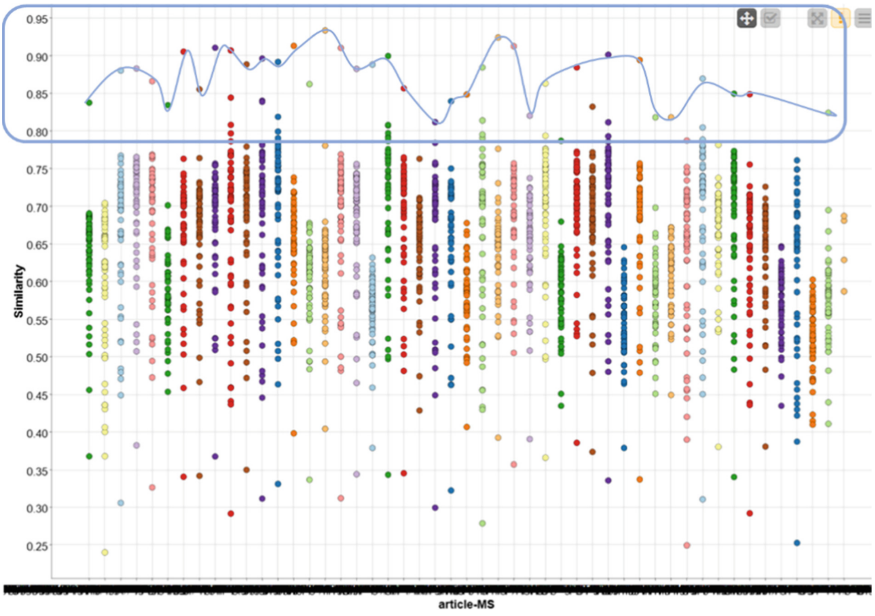

Fig. 5. Distribution of the similarity index using the Cartesian product of 52 articles of the EU directive and 52 articles of the Italian transposition law

a real correspondence in the Italian implementation. It is very evident that art. 6 is not implemented in the Italian legislation with the same normative essence. In fact, art. 6 delegates the appropriate norms (like arts. 35, 37, and 38) to each Member State.

5 Conclusions and Interpretation of the Results

The Digital Era is entering also in the legislative process creating a deep digital transformation and new legal theory research questions are pushing the urgent need to define a new theoretical framework of the Smart Legal Order. This framework designs the permitter where the emerging AI technologies could operate preserving the constitutional principles and the democratic values, while tackling the ethical issues. On the other hands the Hybrid AI methodology could mitigate some risks and weaknesses produced adopting isolated AI non-symbolic techniques. In this light, LegalXML standards, in particular Akoma Ntoso (AKN), could guarantee a solid background for combining rule-based approaches, semantic web knowledge, document structure information and non-symbolic AI (Palmirani et al. 2014a, 2014b, 2017, Sovrano et al. 2020). AKN is capable also to manage the workflow of the legislative process and the temporal model of the diachronic evolution of the norms over time, for building a solid Smart Legal Order.

The similarity index alone produces some very interesting results but without a legal interpretation we might conclude something wrong.

Directive	TP	FP	TN	FN	Precision	Recall	Accuracy
Directive 2014/53/EU	36	4	10	2	90%	94.74%	88.46%
Directive 2014/89/EU	5	4	4	2	55.55%	71.42%	60%

Some articles diverge because the Member State is delegated by the EU Commission to regulate. Some articles are regulating transitory norms and they depend on domestic law (e.g., abrogations). Finally, some other articles diverge because they statue the connection between national and European level of norms (e.g., process of notification, certification, authorization).

Finally, the analysis allows to detect where the EU norms have been implemented in the national law even if the original provisions are aggregated or split in different partitions of the structure. It is the case of Directive 2014/89/EU in which, even if the accuracy is quite low, experiments provide relevant information about the position of the norms in the implementation. We can conclude that use of Akoma Ntoso representation with this hybrid approach, and in general the semantic and the LegalXML annotation, produces important inputs for implementing an explicable and transparent law-making system even if supported by AI, avoiding the black-box effect (Pasquale 2015).

References

Ashley, K.: Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press (2017)

Casanovas, P., Palmirani, M., Peroni, S., Van Engers, T., Vitali, F.: Semantic web for the legal domain: the'next step. Seman. Web **7**(3), 213–222 (2016)
Casanovas, P., et al.: A linked democracy approach for regulating public health data. Heal. Technol. **7**(4), 519–537 (2017)
Chomsky, N.: Language and Mind. Cambridge University Press (2006)
Fillmore, C.J., Baker C.F.: A frames approach to semantic analysis. In: Heine, B., Narrog, H. (eds.) In The Oxford Handbook of Linguistic Analysis, pp. 313–340. Oxford University Press (2009)
Filtz, E., Kirrane, S., Polleres, A.: The linked legal data landscape: linking legal data across different countries. Artif. Intell. Law **29**(4), 485–539 (2021)
Forst, R., Günther, K.: Normative Ordnungen. Suhrkamp Verlag (2021)
Francesconi, E.: On the future of legal publishing services in the semantic web. Future Internet **10**(6), 48 (2018)
Fratrič, P., Sileno, G., van Engers, T., Klous, S., et al.: Integrating agent-based modelling with copula theory: preliminary insights and open problems. In: Krzhizhanovskaya, V.V. (ed.) Computational Science – ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part III, pp. 212–225. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-50420-5_16
Greenleaf, G.: Free access to legal information, LIIs, and the free access to law movement. In: Danner, R., Winterton, J. (eds.) The IALL International Handbook of Legal Information Management. Ashgate (2011)
Greenleaf, G., Mowbray, A., Chung, P.: Strengthening development of rules as code. Submission to the OECD's OPSI on Cracking the Code (2020)
Hildebrandt, M.: Law as computation in the era of artificial legal intelligence: speaking law to the power of statistics. Univ. Toronto Law J. **68**, 12–35 (2018)
Hildebrandt, M.: Code-driven law: freezing the future and scaling the past. In: Deakin, S., Markou, C. (eds.) Is law Computable? Critical Perspectives on Law and Artificial Intelligence. Hart Publishing (2020)
Marmor, A.: The Language of Law. Oxford University Press (2014)
Micheler, E., Whaley, A.: Regulatory technology: replacing law with computer code. Eur. Bus. Organ. Law Rev. **21**(2), 349–377 (2019)
Palmirani, M.: Legislative change management with Akoma-Ntoso. In: Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M. (eds.) Legislative XML for the Semantic Web. Law, Governance and Technology Series, vol. 4, pp. 101–130. Springer, Dordrecht (2011). https://doi.org/10.1007/978-94-007-1887-6_7
Palmirani, M., Vitali, F.: Akoma-Ntoso for legal documents. In: Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M.A. (eds.) Legislative XML for the Semantic Web, pp. 75–100. Springer Netherlands, Dordrecht (2011). https://doi.org/10.1007/978-94-007-1887-6_6
Palmirani, M., Vitali, F., et al.: Swiss Federal Publication Workflow with Akoma Ntoso. JURIX, pp. 179–184 (2014a)
Palmirani, M., Vitali, F., Bernasconi, A., Gambazzi, L.: Swiss federal publication workflow with akoma ntoso. In: Proceedings of JURIX, pp. 179–184 (2014b)
Palmirani, M., Governatori, G., Rotolo, A., Tabet, S., Boley, H., Paschke, A.: LegalRuleML: XML-based rules and norms. In: Olken, F., Palmirani, M., Sottara, D. (eds.) Rule-Based Modeling and Computing on the Semantic Web: 5th International Symposium, RuleML 2011– America, Ft. Lauderdale, FL, Florida, USA, November 3–5, 2011. Proceedings, pp. 298–312. Springer Berlin Heidelberg, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24908-2_30
Palmirani, M., Governatori, G., Athan, T., Boley, H., Paschke, A., Wyner, A.: Legalruleml core specification version 1.0. OASIS Standard (2017)
Pasquale, F.: The Black Box Society: The Secret Algorithms that Control Money and Information. Harvard University Press (2015)

Sovrano, F., Palmirani, M.,Vitali, F.: Deep Learning Based Multi-Label Text Classification of UNGA Resolutions. ICEGOV2020 (2020)
Verheij, B.: Artificial intelligence as law. Artif. Intell. Law **28**(2), 181–206 (2020)
Nanda, R., et al.: Unsupervised and supervised text similarity systems for automated identification of national implementing measures of European directive. Artif. Intell. Law **27**(2), 199–225 (2019)

Derogations Analysis of European Legislation Through Hybrid AI Approach

Monica Palmirani(✉) and Davide Liga

CIRSFID-ALMA AI, University of Bologna, Bologna, Italy
{monica.palmirani,davide.liga2}@unibo.it

Abstract. This paper presents a use case of hybrid AI approach applied to the European legislation with the aim to detect the derogations in the norms and to extract the main components. The result is modelled in Akoma Ntoso XML standard for supporting further applications, open data sharing and interoperability between different tools. We have conducted this research inside of the project 'Drafting legislation in the era of AI and digitisation' with the support of the EU Commission – Directorate General Informatics Unit B2 – Solutions for Legislation, Policy & HR.

Keywords: Hybrid AI · Akoma ntoso · Visualization

1 Drafting Legislation in the Era of Artificial Intelligence and Digitisation

The European Commission is recently providing a roadmap for the *digital-ready legislation*[1] with an interdisciplinary approach and it is investigating the "Drafting legislation in the era of artificial intelligence and digitisation" (workshop 2019)[2]. EU Commission, Directorate-General for Informatics is performing with the University of Bologna a study on "Drafting legislation in the era of artificial intelligence and digitisation" that includes three pilot-cases using AI techniques applied to support the legal drafting units. In this study we propose a third way (e.g., *Hybrid AI for Law*) with a legal and technical model for developing computable informatics legal systems compliant by-design (or *Legal Protection by-design* as Hildebrandt defined) with theory of law, integrating also Semantic Web approach and LegalXML annotation (Filtz 2021, Robaldo 2019).

2 Hybrid AI Approach and Methodology

We propose to use the so-called *Hybrid AI* where *human-in-the-loop*, *human-on-the-loop*, and *human-in-command* principles[3] are combined with different complementary

[1] https://joinup.ec.europa.eu/collection/better-legislation-smoother-implementation/digital-ready-policymaking.

[2] https://ial-online.org/wp-content/uploads/2019/07/Invitation-EN.pdf.

[3] High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence, 2019.

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 123–137, 2022.
https://doi.org/10.1007/978-3-031-12673-4_9

disciplines (law, philosophy, ethics), using symbolic and sub-symbolic AI techniques integrated with Semantic Web findings in order to add context and meanings to the pure *data-driven* or *code-driven* methodology. The *Hybrid AI* is very promising approach especially in legal domain where the context, values, concepts are fundamental for correctly apply the AI outcomes (Rodríguez-Doncel 2021, Fratrič 2021, Verheij 2020, Ashely 2017). Another important element that is fundamental for guaranteeing the legitimacy of the whole *digital law-making process* is the metadata concerning the workflow. Secondary we know about some specific critical technical issues that characterise the legislative domain when the AI non-symbolic alone is used: i) normative references knowledge and connected destination text are often not investigated, ii) article-level approach is fundamental respect the document-oriented method, iii) the temporal analysis of the documents permits to manage the point-in-time, iv) legal language peculiarities must be considered respect a sole frequency approach (e.g., TF-IDF term frequency-inverse document frequency method), v) contextual and semantic enrichment in legal domain is crucial (e.g., jurisdiction). For this reason, we use LegalXML Akoma Ntoso standard (now abbreviated in AKN) as background format of the legal document and on the top of this we add other AI solutions. In the light of the previous considerations, we have adopted a hybrid methodology mixing different techniques: unsupervised for discovering new knowledge to offer to the legal experts during the supervised phase. Additionally, we have also used symbolic and non-symbolic AI techniques. Finally, Akoma Ntoso is used as skeleton for creating the context and the semantic annotation. We have proceeded as following (Table 1).

Table 1. Methodology of hybrid AI.

A. Legal analysis	1. Legal analysis using Eur-Lex
B. Preparation of dataset	2. Selection of the document using Sparql end point of CELLAR[a]
	3. Use of metadata of ELI/CDM/Eurovoc (the ontologies defined at European level for managing legal metadata) from CELLAR
	4. Raw conversion of the documents in AKN
	5. Legal analysis of random sample of partitions and metadata
	6. Definition of the taxonomy of derogations
C. Exploration of the knowledge	7. Extraction of the partitions (article level)

(*continued*)

Table 1. (*continued*)

	8. Using KNIME[b] + Python[c] for exploring new knowledge
D. Experiment	9. Annotation of the fragments
	10. NLP analysis of the derogations (RegEx[d], tree banks) and AI (tree kernel) for extracting the component of the derogations (e.g., destination, jurisdiction, scope, temporal elements, etc.)
	11. AKN conversion of the legal knowledge extracted
E. Evaluation	12. Visualization
	13. Validation
	14. Interpretation

[a]CELLAR: Publications Office, *Cellar: the semantic repository of the Publications Office*, Publications Office, 2018, https://data.europa.eu/doi/10.2830/028321.
[b]https://www.knime.com/
[c]https://www.python.org/
[d]See the standard specification of the language RegEx: ISO/IEC/IEEE 9945:2009 *Information technology – Portable Operating System Interface (POSIX®) Base Specifications, Issue 7.*

3 Derogations

A derogation is a legislative tool that makes it possible to create particular subcases starting from a basic obligation, permission, or right. The derogation is frequently connected with the action of 'disapplication' (e.g., "disapplication provision"[4]), limited to a specific interval of time, or related to some special categories of addresses regulated with the specifications of conditions. As we know from legal theory, this instrument is very relevant in preserving rule-of-law principles all the while making for flexibility in some circumstances like the COVID-19 pandemic. However, derogations are difficult to retrieve in the text and across the common legal databases. It is more difficult to track modifications over time. Legal experts and the legal drafters therefor struggle to follow the chain of derogations and thus obtain clear and transparent legal information. Here is an example of a modification of a derogation introduced in response to COVID-19:

Article 1
Regulation (EU) No 223/2014 is amended as follows: `(omissis) in Article 13(1), the following subparagraph is added:`[a]`'By way of derogation from the first subparagraph, the deadline for the submission of the annual implementation report for the year 2019 shall be 30 September 2020.';`

[a]Art. 1 of Regulation EU 2020/559, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32020R0559&from=EN. Visited in June 2022.

[4] See here an example: https://www.legislation.gov.uk/ukpga/2008/7/section/6/enacted.

3.1 Preliminary Taxonomy of Derogations

Using EUR-Lex[5], a legal expert defines 15 categories of derogations along four axes: i) Frequent linguistic formulations; ii) Temporal parameters; iii) Relationships between EU legislation and Member States; iv) Relationships between primary legislation and delegated acts. Then the legal informatics has extracted the fragment of the documents and the legal experts have grouped them in the following categories using the legal meaning:

• Frequent linguistic formulations:

by way of derogation from/to <partition>
without prejudice to the obligations defined by <partition>
derogation applied in accordance to/with
derogation from/to the <partition> referred to in <partition>
derogation from
derogation to application
derogation applied in accordance with

- Derogation and internal times

Derogation that includes temporal parameters that are part of the deontic norms:

By way of derogation from the first subparagraph, the deadline for the submission of the annual implementation report for the ***year 2019 shall be 30 September 2020***

- Derogation and external times

Derogation that are limited by temporal parameters in the applications, enter into operation, or enter into force:

Upon Commission's initiative or in response to a request from a beneficiary country, a beneficiary country may be granted a ***temporary derogation*** from the provisions of this section where:
`By way of derogation from the date of application referred to in the second paragraph of Article 61, Article 46 shall apply from` ***`17 June 2018`*** `insofar as necessary in order to allow a timely recognition of control authorities and control bodies.`[a]

[a] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018R0848&from=EN Visited in June 2022.

[5] EUR-Lex "is your online gateway to EU Law. It provides the official and most comprehensive access to EU legal documents. It is available in all of the EU's 24 official languages and is updated daily" https://eur-lex.europa.eu/content/welcome/about.html?locale=en. Visited in June 2022.

- Derogation and jurisdiction

Derogation that are limited in the application to some countries:

By way of derogation from paragraphs 1 and 2, ***in Cyprus, Croatia, Malta and Slovenia***, the amount referred to in those paragraphs may be set at a value lower than EUR 500, but not less than EUR 200 or, in the case of Malta, not less than EUR 50

- Exception

Exception is a special derogation where the subcase is strongly distinguished from the regular norm:

In any case, all operators and groups of operators, with the ***exception*** of those referred to in Articles 34(2) and 35(8), shall be subject to a verification of compliance at least once a year

- Shall not apply

'Shall not apply' is used as negative formulation:

Article 19 of this Directive ***shall not apply*** where a bundle within the meaning of Directive (EU) 2018/1972 includes elements of an internet access service as defined in point (2) of Article 2 of Regulation (EU) 2015/2120 of the European Parliament and of the Council

- Member States delegation of derogate

This formulation is used for delegate to the Member State the regulation of some subcases:

Member States shall regularly review derogations under this paragraph taking into account good practices in separate collection of waste and other developments in waste management

- Request of derogation from Member States

This formulation is used when the Member State requests more room for applying a derogation:

> ***At the request of a Member State,*** the Commission may allow a derogation from the prohibition set out in Article 13(1) of Regulation (EC) No 1967/2006, provided that a number of conditions set out in Article 13(5) and(9) are fulfilled

- Delegated Acts

This formulation is used for adopting delegated Acts. This is a derogation using different level of source of law:

> ***The Commission shall adopt delegated acts*** in accordance with Article 264 concerning the special rules referred to in paragraph 1 of this Article regarding derogations from the requirements provided for in Article 229(1) and Articles 233 and 237 and imposing additional requirements for the entry into the Union of the following:

Finally, we frequently find the so-called *reflexive derogation* that act internally to the same act often correlated to temporal period of efficacy.

> By way of derogation from ***point (a)***, from *1 January 2021 to 31 December 2021* the obligation to decommission the unique identifier of medicinal products which the wholesaler intends to distribute outside of the Union shall not apply to products which he intends to distribute in the United Kingdom

The *external derogation* has the destination another legal document:

> 2. By way of derogation from ***Article 13(6) of Regulation (EC) No 1235/2008***, at the verification of a consignment the relevant Member State's competent authority shall endorse the certificate of inspection by validating box 20 in TRACES[a]

[a] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32020R0977.

3.2 Dataset of Derogations

The dataset is made up of legislative acts in the span of time from 2010 to 2020 for a total of 15,328 documents. All the documents were provided in Formax[6] format by the European Publication Office with also the related metadata (e.g., ELI[7]/CDM[8]/Eurovoc[9]). Each document is a package with a main document and possible further attachments and annexes. The documents are converted into Akoma Ntoso and using the taxonomy we came up with 13,587 partitions involved in the derogation, using a preliminary 'indicator' taxonomy for extracting the text involved in the experiment. The goal should detect several important elements in the text: destination of the derogation, action proposed, jurisdiction, temporal elements, conditions, quantities. In Akoma Ntoso we can model directly the text of the derogation and integrate the semantic role of each element in the metadata specifically designed for hosting the information concerning the modifications, including derogation.

3.3 Akoma Ntoso Conversion

During the conversion in AKN from Formex we detected the part of the wording involved in the derogation: citations of the main obligation that is derogated from, temporal parameters, and the particular scope being derogated from. Additionally, we have modelled the derogation in Akoma Ntoso metadata in order to reuse them for further statistical elaborations.

---

[6] Formex is the acronym of "Formalized Exchange of Electronic Publications"."Formex describes the format for the exchange of data between the Publication Office and its contractors. In particular, it defines the logical markup for documents which are published in the different series of the Offical Journal of the European Union.Formex V4 is based on the international standard XML (Extensible Markup Language—W3C Recommendation, February 10, 1998). It entered into force on May 1st, 2004." https://op.europa.eu/it/web/eu-vocabularies/formex.

[7] ELI is the acronym of "European Legislation Identifier" that defines Unique Naming convention for legislation, common data model of metadata, interoperable mechanism (API) for query the information between the member states. It has been introduced with the "Council conclusions of 6 November 2017 on the European Legislation Identifier" in all the member states and in the European institutions. The ELI patterns for the European legislation is defined here: https://eur-lex.europa.eu/eli-register/eu_publications_office.html, https://op.europa.eu/it/web/eu-vocabularies/eli. Visited in June 2022.

[8] CDM is the acronym of "Common data model" that is used by the European Publication Office for creating a sharable ontology: https://op.europa.eu/it/web/eu-vocabularies/cdm.

[9] Eurovoc "is the EU's multilingual and multidisciplinary thesaurus. It contains keywords, organized in 21 domains and 127 sub-domains, which are used to describe the content of documents in EUR-Lex." https://eur-lex.europa.eu/browse/eurovoc.html?locale=en. Visited in June 2022.

Formex	Akoma Ntoso
<ALINEA>By way of derogation from the second paragraph, Member States may choose not to apply the provisions of point ORO.FTL.205(e) of Annex III to Regulation (EU) No 965/2012 and continue to apply the existing national provisions concerning in-flight rest until <DATE ISO="20170217">17 February 2017</DATE>.</ALINEA>	<alinea eId="body__art_2__al_3"> <content eId="body__art_2__al_3__content"> <mod eId="body__art_2__al_3__content__mod_1"> By way of derogation from the second paragraph, Member States may choose not to apply the provisions of <ref eId="ref_1" href="href="/akn/eu/act/regulation/2012-02-17/965-2012/!main/>annex_III"> point ORO.FTL.205(e) of Annex III to Regulation (EU) No 965/2012 </ref> and continue to apply the existing national provisions concerning in-flight rest until<date date="2017-02-17" refersTo="#derogationTime">17 February 2017</date>. </mod> </content> </alinea>

We first define all elements that in the sentence relate to derogation action. Thus, we extract the following element that find an adequate representation in:

- destination of the derogation is detected in the text using <ref href > (normative references connected with the derogation in order to produce a graph of all the derogations and the relative norms). This information is reported also in the metadata in the tag <destination> inside of the modification information <scopeMod>;
- conditions (e.g., only for the bank, only for the COVID-19 pandemic situation) are detected in the text and then reported in the metadata tag <condition>;
- jurisdiction (e.g., only for Denmark) is detected in the text using <location> even if it is not really a physical place but the concept of the state sovereignty and the consequent jurisdiction;
- temporal parameters (e.g., for six months) are detected in the text and stored in different tags of the metadata according to the role played by this information: <duration> for specifying how long the derogation acts, <force> for recording the enter into force time, <efficacy> for recording the time of enter into operation;
- limitation in the application of some specific domain (e.g., only for tax law) we can store this information in the <domain> tag.

We then model everything in AKN in order to fix the knowledge and to reuse it for the search engine, the semantic web filter, or other sophisticated application.

```
<scopeMod type="exceptionOfScope">
<source href="body__art_2__al_3__content__mod_1"/>
<destination href="/akn/eu/act/regulation/2012-02-17/965-2012/!main/annex_III"/>
<force>
<date date="2014-02-20"/>
</force>
<duration>
<date date="2017-02-17" refersTo="#endDate"/>
</duration>
<condition> Member States </condition>
<domain> continue to apply the existing national provisions concerning in-flight
</domain>
</scopeMod>
```

3.4 Technical Aspects

To retrieve the relevant part of the text we combined the use of regular expressions together with the powerful SpaCy library, in Python. More precisely, a software has been created which is capable of reading Akoma Ntoso structures by using two python libraries dedicated to the navigation of xml files (i.e. lxml and xml.etree), searching for all the elements of the body of the legal document to match all those elements containing the most frequent linguistic formulas by which derogations appears in the text. This first part of the process employed regular expressions and aims at extracting the sentences where derogations are found. After the navigation of the Akoma Ntoso files and the matching process of the derogation linguistic formulas, a SpaCy[10] pipeline has been feed with all retrieved sentences. This pipeline (which is currently under development) is built to detect as much information about the derogations as possible (e.g. time, places, conditions). After having found derogations, and having extracted the relative information, the software returns an Akoma Ntoso file which is marked up with the information about derogations.

Regarding the first phase, the matching process performed by using regular expressions produces a list of matches for the content of each element found in the xml structure of the body of the Akoma Ntoso files. It is important to note that this first part has been designed to deal with the physiological fragmentation of the content (fragmentation which can be more or less relevant depending on the verbosity of the information within the xml structure). In fact, those xml tags which can contain textual content sometimes have inner inline elements which can add a further layer of complexity during the extraction of the sentences. In this part, we made sure that all elements' texts and tails were considered, in order to reconstruct the content of the sentences as much as possible. This choice of design makes it possible to work also with irregular structures, or with structures which might possibly be not well-formed, or whose internal text is fragmentised by inner inline tags. In this way, we tried to feed the regular expressions with sentences that are as much complete as possible.

[10] SpaCy is an open-source library of Natural Language Processing techniques capable to manage the task called Named Entity Recognition that permits to detect in the text persons, organizations, roles, dates, locations, concepts. For applying SpaCy to legal language domain it is necessary a customizaiton. https://spacy.io/.

In the second phase, the sentence containing the retrieved provision is transferred into a SpaCy pipeline in order to find as much information about the derogation as possible. Also in this case, having a sentence as much complete as possible is crucial to achieve better results. In fact, the steps of the pipeline are performed by navigating the dependency tree, the part of speech, and the tags of the sentences. The pipeline, which is currently under development, is partially completed, and has been designed as follows:

1 Searching for normative references (i.e. references to legal documents, such as "Regulation 96/2016") and add them as "act_reference" entities;
2 Searching structures of legal documents (i.e. references to portion of legal documents, such as "Second paragraph of Article 3"), adding them as "part_reference" entities, and connecting them to any related normative reference possibly found in the first step;
3 Searching for EU member states and adding them as entity;
4 Searching for time references, and adding them as temporal entities;
5 Adding as entities the derogation formulas (which, as said before, are found using regular expressions, an example of derogation formula is "by way of derogation from");
6 Adding deontic operators such as "should" (which can be crucial to determine the presence of other elements related to the derogation within the same sentence, and which can be used to take into consideration the deontic and argumentative dimension of the derogation);
7 Connecting the derogation formula (found in step 5) to its corresponding "destination": this means connecting the derogation formula to the relative part_reference (including the act_reference, if any);
8 Finding any delegation and adding this information into the SpaCy document (i.e. finding all those linguistic formulas which imply a delegation of power within the derogation);
9 Find conditions (i.e. finding as many conditions as possible, related to the derogation) by navigating specific linguistic indicators (such as "only if", "unless", and so on).

In other words, all the provision extracted in the first phase are then passed to this SpaCy pipeline to search for the information about the derogation itself. The result of the SpaCy pipeline is a complex object which contains all the information we need to finally markup the original Akoma Ntoso with the new information about the derogation (including all the metadata which have been found). Although the pipeline is still under development, we managed to extract hundreds of derogations and produce marked up Akoma Ntoso accordingly.

3.5 The Derogation Graph

Starting from the Akoma Ntoso files marked up with annotations (as described before), we created a graph to represent and visualize derogations using d3js, a powerful library which is capable of generating interactive visualizations. Graphs are generally composed of two elements (van Kuppevelt 2020, Coupette 2021, Chalkidis 2021): nodes

and links connecting nodes. It is not the first time that the scholars create graphs for representing the legal knowledge, but concerning the legislation there are few experiments that stressed three main problems: i) the computational issues because the calculation of the network and the relationships between nodes need a specific algorithm capable to work with hundreds nodes (Bommarito 2019, Katz 2014); ii) the visual model for managing the complexity of the graph with large numbers of nodes (Koniaris 2021); iii) the transparency in the navigation that could be useful for the legal experts and in meantime understandable and auditable with valuable legal metadata (Alschner 2020; de Jong 2022).

For coping with these three main issues, we extracted a graph where each node is a legal act (which stores the information of the act itself, and where the unique ID is the FRBR Akoma Ntoso URI) and where each link is a relationship of derogation which can be established either with other nodes or with the same node (for reflexive relations). We used the AKN annotation and specifically the metadata modification <scopeMod> elements. The destination of the derogation is used for creating a connection between derogation and document derogated. The other elements (which are not necessarily present) are considered as information about the single connection, so their information is stored within the relative link. More precisely, "duration" provides information about the duration of the derogation, "place" provides information about the geographical zones where the derogation is valid, "domain" provides information about the domains of application of the derogation, "condition" provides information about the conditions for which the derogation is valid.

For each node of the graph (i.e., for each legal document) we also collected its relative Eurovoc classifications, which can be found within the AKN tag <classification>, and more precisely in the "value" attribute of any tag <keyword> whose "dictionary" attribute is equal to "eurovoc". In the image above, we stored seven pieces of information for each node. The *unique AKN ID*, the *CELEX number*[11], the *title* of legal act, and all the times in which the single node derogates others or itself (*derogator_times* and *selfderogated_times*, respectively), or is derogated by others (*derogated_times*). Finally, we stored the Eurovocs related to each legal document. Regarding the links, they can store not only the *source* and *destination*, but also all the information that we extracted from the files as explained before, namely: *places*, *conditions*, *domain*, *duration*. Also, it is important to underline that the information of single nodes are displayed whenever the mouse is over them, as described in the following image:

Importantly, each node has been assigned a *color* and a *weight* depending on 3 variables:

1) how many times the node is derogating other nodes (not considering itself);
2) how many times the node is derogating itself;
3) how many times the node is derogated by other nodes.

[11] *CELEX number* "is a unique identifier assigned to a document. It is independent of the language of the document." It is the unique identifier adopted before the ELI and AKN naming convention and it is a legacy number for detecting legislative document in CELLAR independently to the various generations of the technology evolution. See https://eur-lex.europa.eu/content/help/eurlex-content/celex-number.html.

The weight, which determines the dimension of the node, is achieved by following formula:

$$w = k * (d + r + t)$$

Fig. 1. Network analysis of the derogations.

where k is a fixed coefficient intended to produce the most appropriate graphical result, d is the times in which the node is derogating other nodes (excluding itself), **r** is the times in which the node is self-derogated, and t is the times in which the node is derogated (by other nodes or by itself). We call d + r + t "centrality".

The color is decided as follows: *yellow* nodes are nodes which are mainly derogated by other nodes (t is greater than r and d), *green* nodes are nodes which are mainly derogating themselves by other nodes (r is greater than t and d), *blue* nodes are nodes which are mainly derogating other nodes (d is greater than t and r). Finally, *grey* nodes are nodes which are not derogating nor being derogated ($d = t = 0$). To make the graph less verbose, we included in the grey nodes only those whose r is greater than 4 (i.e., they derogated themselves more than 4 times).

By clicking on the single node, it is possible to visualize the relative xml file (if it is available), which will be visualized in a new tab, within the same browser with all the metadata for a better explicability to the legal expert (see Fig. 1). More precisely, we used eXist-db to store and visualize each AKN document.

4 Conclusions

The derogation analysis provided much important information that could be used for better regulation:

- 6% of derogations are delegated to Member State legislation. This is interesting because we can detect the parts of derogations from the EU legislation that involve national law (relationship between EU and Member State law);
- A minimal part is connected with temporal conditions;
- Another minimal percentage is connected with 'delegated acts'. Also, this is relevant in the relation between primary and secondary law in EU sources of law.

Three patterns (see Fig. 2) win out, linguistically speaking, in virtue of how useful they turn out to be in LEOS[12] (a specialized editor developed by the EU Commission for legal drafting) modelling for harmonizing derogations and markup (this is useful for the search engine):

- "By way of derogation from/to" - 26%;
- "shall not apply" - 20%;
- "By way of exception" - 11%.

We have also used Eurovoc classification to understand the topics which are more affected by derogations. We have discovered that the domains with more derogations are: import license, tariff, food and agriculture. The most frequent self-derogation falls in the financial/market domain regulation. Finally, the most frequent derogating acts belongs to the food, agriculture, regulation of the market topics, even if the recent emergency due to the COVID records a significant percentage of derogations.

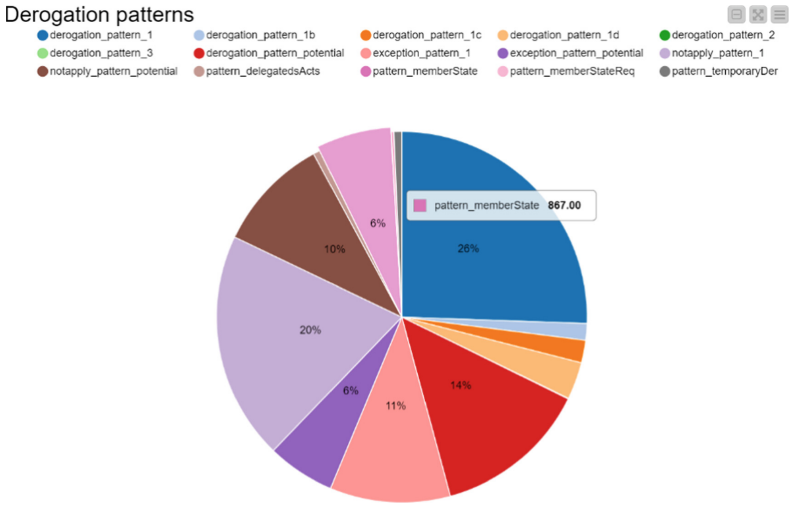

Fig. 2. Statistical about the pattern.

[12] LEOS "(Legislation Editing Open Software) is an open source software designed to help those involved in drafting legislation, which in itself is a complex knowledge-intensive process, by supporting efficient online collaboration." https://joinup.ec.europa.eu/collection/justice-law-and-security/solution/leos-open-source-software-editing-legislation. Visited in June 2022.

Acknowledgements. This work was supported by the 'Drafting legislation in the era of AI and digitisation' with the support of the EU Commission – Directorate General Informatics Unit B2 – Solutions for Legislation, Policy & HR. The European Publication Office provided the documents and the datasets.

References

Alschner, W.: Legal Analytics. AI and the Law in Canada. In: Scassa, T., Martin-Bariteau, F. (eds.) LexisNexis Canada (2020). Available at SSRN: https://ssrn.com/abstract=3662196

Ashley, K.: Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge University Press (2017)

Bommarito II, M.J., Katz, D.M.: A mathematical approach to the study of the United States Code. Physica A: Statistical Mechanics and its Applications, 389 (2019)

Boulet, R., Mazzega, P., Bourcier, D.: A network approach to the French system of legal codes, part I: analysis of a dense network. Artificial Intelligence and Law, 19 (2011)

Chalkidis, I., et al.: LexGLUE: A Benchmark Dataset for Legal Language Understanding in English (13 Oct 2021)

Coupette, C., Beckedorf, J., Hartung, D., Bommarito, M., Katz, D.M.: Measuring law over time: a network analytical framework with an application to statutes and regulations in the United States and Germany. Front. Phys. **9**, 658463 (2021)

de Jong, H., van Dijck, G.: Network analysis in legal history: an example from the Court of Friesland. Tijdschrift voor Rechtsgeschiedenis-Revue d Histoire du Droit-The Legal History Review, 1–13 (2022)

Filtz, E., Kirrane, S., Polleres, A.: The linked legal data landscape: linking legal data across different countries. Artificial Intelligence and Law **29**(4), 485–539 (2021). https://doi.org/10.1007/s10506-021-09282-8

Fitsilis, F.: Artificial Intelligence (AI) in parliaments – preliminary analysis of the Eduskunta experiment. The Journal of Legislative Studies **27**(4), 621–633 (2021)

Forst, R., Günther, K.: Normative Ordnungen. Suhrkamp Verlag (2021)

Francesconi, E.: On the future of legal publishing services in the semantic web. Future Internet **10**(6), 48 (2018)

Katz, D.M., Bommarito II, M.J.: Measuring the complexity of the law: the United States Code. Artificial Intelligence and Law, 22 (2014)

Koniaris, M., Anagnostopoulos, I., Vassiliou, Y.: Network analysis in the legal domain: a complex model for European Union legal sources (v3) (2021). https://arxiv.org/abs/1501.05237

Liga, D., Palmirani, M.: Classifying argumentative stances of opposition using Tree Kernels. In: ACAI 2019: Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, pp. 17–22 (2019)

Palmirani, M.: Legislative change management with Akoma Ntoso. In: Legislative XML for the semantic Web, pp. 101–130. Springer (2011)

Palmirani, M., Vitali, F.: Akoma-Ntoso for legal documents. In: Legislative XML for the semantic Web, pp. 75–100. Springer (2011a)

Palmirani, M.: Akoma Ntoso for making FAO resolutions accessible. In: Knowledge of the Law in the Big Data, pp. 159–169. IOS Press (2019)

Palmirani, M., Cervone, L.: Measuring the Complexity of the Legal Order over Time. In: Casanovas, P., Pagallo, U., Palmirani, M., Sartor, G. (eds.) AICOL -2013. LNCS (LNAI), vol. 8929, pp. 82–99. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45960-7_7

Robaldo, L., Villata, S., Wyner, A., Grabmair, M.: Introduction for artificial intelligence and law: special issue "natural language processing for legal texts." Artificial Intelligence and Law **27**(2), 113–115 (2019). https://doi.org/10.1007/s10506-019-09251-2

Rodríguez-Doncel, V., Palmirani, M., Araszkiewicz, M., Casanovas, P., Pagallo, U., Sartor, G. (eds.): AICOL/XAILA 2018/2020. LNCS (LNAI), vol. 13048. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-89811-3

Sovrano, F., Palmirani, M., Vitali, F.: Deep Learning Based Multi-Label Text Classification of UNGA Resolutions. ICEGOV (2020)

van Kuppevelt, D., van Dijck, G., Schaper, M.: Purposes and challenges of legal network analysis on case law. In: Whalen, R. (ed.) Computational legal studies: The promise and challenge of data-driven legal research Edward Elgar Publishing (2020)

Verheij, B.: Artificial intelligence as law. Artif Intell Law **28**, 181–206 (2020)

Good Practices of User Experience and Design Research for Mobile and Electronic Governmental Services

Rachelle Sellung[1](✉), Marie Hölscher[1], and Lisa Burgstaller-Hochenwarter[2]

[1] Fraunhofer IAO, Nobelstraße 12, 70569 Stuttgart, Germany
{rachelle.sellung,marie.hoelscher}@iao.fraunhofer.de
[2] Technikon Forschungs- Und Planungsgesellschaft mbH, Burgplatz 3a, 9500 Villach, Austria
BURGSTALLER@technikon.com

Abstract. This paper summarizes desk research conducted on User Experience and Design Research for Mobile Governmental Services and presents Good Practices. As mobile services are the next step for governmental services, it is necessary that as this transition occurs that the integration of user needs, and user experience is considered from the start of the development process. This paper summarizes the methods, key findings, challenges that were discovered from UX and Design literature research on mobile and electronic governmental services. As a result, this paper derives Good Practices as a reference point for technical developers and other interested stakeholders to include the necessary user experience and design needs for market success.

Keywords: Mobile services · Mobile government · mGovernment · eGovernment · Governmental services · User experience · Design research

1 Introduction

It is becoming increasingly important today that the electronic services provided by the government offer policy information and governmental services. However, the design and management of electronic government (eGov) apps is in some cases in great need of improvement, which leads to problems in their use (Cahyono and Susanto 2019). The technologies developed in recent years in the field of mobile and electronic services for governmental services, all faced the problem of having to meet different requirements of digitalization. These requirements are in security, privacy and numerous other requirements but also around user friendliness. Unfortunately, when technologies are developed, the focus is often on one of the requirements and far too rarely on the user experience (Kureerung and Ramingwong 2019b). Only if the services are accepted by the users, they can benefit from the usefulness of the services. This then leads to the ability to tailor the government's services to its target audience, thereby reaching the broader population (Kö et al. 2018).

The motivation of this work was to conduct a literature review of existing User Experience and Design Research that has been done on mobile governmental (mGov)

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 138–149, 2022.
https://doi.org/10.1007/978-3-031-12673-4_10

services and electronic governmental (eGov) services. After reviewing which methods have been used and results that have been the outcome of a general consensus of what is lacking and how it could be improved, this paper presents ten good practices that other key stakeholders could follow to improve their User Experience of their services along their development process. The research and work presented in this paper has contributed and will be ongoing work in the EU funded MGOV4EU. This project focuses on creating pilots that showcases a citizen-centric approach that enables eIDAS and the Single Digital Gateway in mobile cross-border governmental services in Europe.

This paper has three following chapters. Chapter two will highlight key related work on User Experience (UX) and Design Research that has been conducted on mGov and eGov Services. This chapter will have two subsections where the first it will explain which methods, models, and approaches have been used. Second, it will explore the key insights from those studies. The third chapter will present the Good Practices which have been extracted from the related work that has been done. Lastly, this paper finishes with a conclusion and future work section.

2 Related Work for UX and Design Research on mGov and eGov Services

This section expands on desk research conducted as a basis for the good practices. There was a reviewing process of selected key words in relation to User Experience and Design Research on Mobile and Electronic Government Services in Scopus and Google Scholar. After this process was conducted, then the most relevant studies were chosen to be presented. First, this section describes an overview of methods that include interviews, surveys, and case studies on mGov and eGov applications and websites regarding User Experience or Design Research. Second, there is a summary of some key results that were concluded in of how some of the research solved various challenges in User Experience and Design Research. The studies highlighted here aimed at gathering feedback about the usability, design, and user experience from mGov developers, designers, and users which were then used to present UX principles and improvements for the development of mGov services.

2.1 User Experience and Design Research Methods Used on mGov and eGov Services

This section summarizes and highlights the related work of five different approaches that were used on evaluating the mobile government services in the user experience and design research.

First, this approach analyzes existing user experience models and uses them to develop a framework for user experience designs that promote awareness of information distribution via mGov applications (Kureerung and Ramingwong 2019a). Several usability models used in the topics of mobile applications evaluation, mGov, and user interface design were identified and analyzed. The authors examined the recurrency of the usability characteristics of usability models like Nielsen's usability model (Nielsen 1993), the ISO 9241–11 usability norm (Abran et al. 2003), and the Software Usability

Measurement Inventory (SUMI) (Kirakowski 2018) along with MGQM (Hussain and Kutar 2009), PACMAD (Harrison et al. 2013), Extend PACMAD (Saleh et al. 2015), QUIM (Seffah et al. 2006), MUSiC (Macleod et al. 1993). Usability factors for mGovernment applications are then collected and clustered. The usability design framework is put forward to promote the effective use of usability factors in the development process. Based on the most recurrent characteristics such as efficiency and ease of use, the authors then developed a framework for designing mGov applications (Kureerung and Ramingwong 2019a).

The second approach that involved the use of usability model which is the Norman's Interaction model aimed at studying the Interaction Design Patterns in governmental apps with the help of the 76 mobile interaction design patterns developed by Hoober and Berkman in 2011 (da Silva and Freire 2020); (Hoober and Berkman 2011). The 13 categories of interaction design patterns developed by Hoober and Berkman are Composition, Display of Information, Control and Confirmation, Revealing more Information, Lateral Access, Drilldown, Labels and Indicators, Information Controls, Text and Character Input, General Interactive Controls, Input and Selection, Audio and Vibration, Screens & Lights and Sensors. These were then combined by Silva and Freire into the following six categories: User Action, Help & Feedback, Search & Filter, Content Design, Input, and Navigation. The authors then evaluated 27 governmental applications in order to find the implemented design patterns in each one of them (da Silva and Freire 2020; Hoober and Berkman 2011).

The third kind of approaches employed empirical research methods to examine the design of gov applications by conducting surveys, interviews, and case studies with citizens and mGov designers and developers (Chang et al. 2020a, 2020b; Isagah and Wimmer 2018; Lönn et al. 2016). A similar approach has been employed for the examination of the collaboration through citizen sourcing for the improvement of the development of mobile government applications (Lönn et al. 2016). The research follows the six steps of the Design Science Research Process by Peffers: Problem identification and motivation, Definition of objectives of the solution, Design and Development, Demonstration, Evaluation, Communication. For the improvement of mobile government applications, workshops were conducted with municipality officials from multiple municipalities and discussed the potential process that can be implemented or invented that would allow citizens to send complaints directly to the government. Based on the results of the workshops, the authors developed three prototypes: An App, an app integrated with an ePlatform, and a final solution integrated with a case management system. The prototypes were then demonstrated to the municipality officials and evaluated via simulations, informed arguments from the government and citizens, tests, and a survey with 35 (Lönn et al. 2016).

In order to determine the most important factors in mGov applications for the elderly, a fourth approach has been identified in their study, they used a mix of quantitative and qualitative methods, including an mGovernment prototype, to investigate key acceptance factors. Research approach used the IGUAN framework, which is a user-driven method. The approach consisted in deriving some design factors from the TAM such as perceptions and attitudes towards the system (Kö et al. 2018). The factors derived were then used to develop an After Scenario Questionnaire (ASQ) that has been filled out by

elderly citizens after getting a demonstration of a conceptual model of an mGov service. Based on those results, a prototype has been developed and demonstrated and a survey has been conducted to assess if the improvements were effective (Kö et al. 2018).

Another method by Cahyono and Susanto examined the user design of mobile eGovernment websites. This involved using eye-trackers and MindWave to record retinal activity and brain waves while e participants searched for specific information on 9 government websites. Then, a 28 statement interview was conducted with the participants to determine the impact of human-computer interaction aspects and mobile website design on the efficiency of mGov services in Indonesia (Cahyono and Susanto 2019). Overall, this section shows that different methods and approaches can be utilized for the evaluation of the user experience in mGov websites and applications. The variety of the methods used does not only display the difficulty of properly evaluating mGov services, but also proves that different target groups require different research approaches. This section first provided a look on the usage of existing usability models and how mobile application characteristics are relevant for the evaluation of mGov services. Case studies were also a popular evaluation method, as multiple authors employed different techniques in conducting these studies, including ASQs, workshops, and the analysis of the brain activity.

2.2 Key Findings from Related Work

These key findings present a set of factors, solutions, and methods that have been identified and developed for the improvement of mGov and eGov design. First, this section provides an overview of design framework for mGov applications, that focuses on satisfying the primary requirements of the user interface. Second, there is an overview of a set of design pattern categories and usability requirements found in empirical research carried out with mGov developers and designers. These findings were found to be the most impactful in the improvement of the usability in mGov applications. Third, UX principles that derived from empirical research from a citizens and mGov users perspectives are presented. These principles can be used for the improvement of the information layout and user interface in governmental apps.

Overview of design framework for the interface of for mGovernmental applications.

Focusing on meeting the primary user interface requirements, this is an overview of some of the design framework results that related work came up with for mGovernmental applications.

The analysis of usability models such as Nielsen's heuristics and the SUMI resulted in the development of a framework for the development of user interface design for mGov applications (Kureerung and Ramingwong 2019a). Their framework consists first in defining the inputs required and the goals that need to be met. The findings are then used to raise questions and start with the design process. In this process, security, privacy, simplicity, learnability, memorability, and satisfaction are the most important factors in use to be improved. Within the factor of use, factor requirements, that are based on specific criteria, need to be fulfilled in order to present an improved user interface for the mGov service (Kureerung and Ramingwong 2019a). This method allows mGov developers to first identify and describe the main functionality of the user interface, which in return provides a way to determine the main requirements of the application.

From there, mGov developers can work on developing mobile applications and websites that are focused on providing the main functionalities in the most usability-satisfying and user-friendly way. These findings also align with other studies, that emphasize the importance of user interface design (Chang et al. 2020a 2020b); (Isagah and Wimmer 2018); (Kureerung and Ramingwong 2019a).

Design pattern categories and usability requirements gained from mGov developers.

This section provides an overview of a set of design pattern categories and usability requirements found in empirical research with mGov developers and designers. These findings proved to be particularly important for improving the usability of mGov applications.

The analysis and categorization of Hoober and Berkman's interaction design patterns resulted in the identification of 6 interaction design categories, that are most impactful in the design of mGov services (da Silva and Freire 2020); (Hoober and Berkman 2011). These categories contain different components that allow mGov application designers to provide a cleaner user interface, an effort free user experience, and an overall more user-friendly product (da Silva and Freire 2020). The 76 interaction design patterns from which these categories were derived can be further studied in (Hoober and Berkman 2011). As for the pattern categories, a brief explanation and some examples of interaction design patterns are presented in the following figure.

Content Design	The Content Design category describes the way content is presented to users in a mobile application. It includes Patterns like Titles, Icons, Lists, Pop-ups, and grids along with the dynamic arrangement of graphical or textual information.
User Actions	The User Actions define the set of methods the app provides to the user to interact with the application and execute the wished-for operation. This category contains interaction strategies such as gestures, touch, voice, and sign on.
Input	The Input patterns category consists of patterns that concern data entry activities. These patterns include Keyboards, Input areas, form selections, voice inputs, and mode switches along with other data entry related patterns.
Navigation	The Navigation category incorporates all activities related with the navigation, access, and movement of the user between content pages. The patterns included in this pattern category are tabs, revealable menus, scrolling, links, and fixed menus along with many more patterns.
Help & Feedback	The Help & Feedback patterns category consists of patterns that informs the users about the status of the operations they are conducting. Wait indicators, notifications, and haptic output are some of the patterns that belong to this category.
Search & filter	The Search & filter patterns category includes search engines and filters that aim at optimizing the search for data and functionalities. These patters are also employed to limit the data volume that the user is exposed to.

Fig. 1. The most impactful interaction design pattern categories for the design of mGov applications (da Silva and Freire 2020); (Hoober and Berkman 2011).

The empirical studies carried out with mGov designers and developers that aimed at identifying the requirements of mGovernment services from designers' perspective, showed that most mGov designers and developers prioritized the usability over other requirements of mGov services such as security, privacy, interoperability, integration, compatibility, and scalability (da Silva and Freire 2020) (Isagah and Wimmer 2018,

2017). A detailed overview of the usability requirements expressed by mGov developers and designers, as well as their characteristics is presented in Table 1:

Table 1. Usability requirements from the perspective of mGov application developers and designers (Isagah and Wimmer 2017, 2018)

Learnability	Easy to learn, Easy to use, Easy to remember
Recognizability	Meets Service Goals, Convenient to user environment, Easy to understand
Operability	Suitability for the device, Conformity of the device with user expectation
User Error Protection	Error tolerance
User Error Aesthetics	Clear and attractive interface
Accessibility	The use of multichannel, The use of multi-language, The use of common and cheap channel

As displayed in Table 1, the usability requirements of mGov applications are categorized into Learnability, Recognizability, Operability, User Error Protection and Aesthetics, and Accessibility (Isagah and Wimmer 2017, 2018). Each of these requirements vary in importance according to the kind of service and kind of users it attracts. For the elderly, for example, the ease of use is the most important requirements, since most of the elderly lack experience with new technologies (Kö et al. 2018); (Susanto et al. 2017); (Talukder et al. 2020). The main challenge in satisfying these requirements consists in finding or developing principles, methods, or frameworks that could cover these concerns but can also be employed across different mGov applications. Furthermore, most designers and developers rely on the existing agile methods used as a substitute for traditional software development methods. These agile methods differ in practices and tactics and do address service requirements in different ways, which shows that there is a lack of standards in the development of mGov services (Isagah and Wimmer 2018). Most designers were also found to employ design approaches that do not involve the user, nor do they use development frameworks that address all usability requirements (Isagah and Wimmer 2018). Cross-platform frameworks, as an example, address compatibility requirements but do pose some challenges regarding the user interface and performance of an app. In addition, the employed guidelines such as material design guidelines, SMS guidelines, and mobile operating systems guidelines are very different for each kind of device. With the technological development happening in the smartphone industry, it becomes hard for mGov designers and developers to agree on a specific guideline. Therefore, standardized principles, guidelines, and best practices should be developed, that address the requirements of m-government services, regardless of the type of device the service is provided on (Isagah and Wimmer 2018).

UX principles derived from citizens for improvement.

Now, UX principles derived from empirical research from the perspective of citizens and mGov users are presented. These principles can be used to improve information layouts and user interfaces in government apps.

Regarding the usability of mobile government applications, the results from case studies and empirical work with citizens can be matched to the Content Design and Search & Filter categories in Fig. 1. (Chang et al. 2020a, 2020b; Isagah and Wimmer 2018). The layout of information is the first design aspect that has a large potential of having a usability reducing effect on mGov apps. To reduce the efforts that the user must make when using the app's functionalities, repeated entrances should be avoided. A proper display of information and functionalities according to their importance also showed signs of increased usability in our findings (Chang et al. 2020a, 2020b; da Silva and Freire 2020; Isagah and Wimmer 2018, 2017). In addition, mixed or large quantities of displayed information proved to be discouraging for potential mGov users (Kureerung and Ramingwong 2019a, Chang et al. 2020a, 2020b).

The next design aspects that showed some negative effects on the usability were colors and icons. The effect of different colors on the human brain has been largely documented, thus making the choice of color a somewhat important decision in the appearance of the interface (Chang et al. 2020a, 2020b; Kö et al. 2018). The colors used should therefore not be too heavy or occupy too much space, nor should the background color be too close to the colors of the buttons. In some cases, where an administration possesses a logo, the colors used in their mobile website or applications should match the colors of the logo (Chang et al. 2020a, 2020b). The icons used in mobile government applications also proved to affect the user-experience, since icons that are hard to recognize or to understand lowered the usability of the app. Inconsistencies in design styles were also found to be negatively affecting the mGov app (Chang et al. 2020a, 2020b; Isagah and Wimmer 2018). Regarding the aforementioned usability aspects, the following principles have been applied and have been proven to improve mobile government applications from an user-experience point of view (; Isagah and Wimmer 2018):

- Multiple entries for the same functions need to be simplified.
- Redundant entries need to be removed.
- Important services need to be placed at high priority positions in the layout.
- The size of visual elements needs to be adapted to their importance.
- The popular services identified in surveys and user experiments need to be added.
- The information architecture needs to be changed to show the most important information first.
- Government applications should need to have a uniform design across administrations to reduce the cost of training users.

Challenges

As a reflection of this desk research, there were five main challenges that can be summarized. The first challenge is that different smartphones and interfaces have different requirements for mGov application development (Isagah and Wimmer 2018, 2017). The second challenge is that there are no standardized approaches that meet the usability requirements of mGov services (Chang et al. 2020a, 2020b; Isagah and Wimmer 2018; Lönn et al. 2016). Third, many developers and designers of mGov use approaches to develop working mobile applications. This requirement usually takes precedence over usability (Isagah and Wimmer 2018; Kureerung and Ramingwong 2019a). The fourth

challenge is that demographics, political status, familiarity with technology, trust, and the nature of the service being offered have the most impact on the usability of mGov services (Chang et al. 2020a, 2020b; da Silva and Freire 2020; Isagah and Wimmer 2018; Kö et al. 2018; Kureerung and Ramingwong 2019a; Lönn et al. 2016).The final challenge is that many mGov solutions do not involve citizens in the development process, making it even more difficult to make them user-friendly (Kö et al. 2018; Lönn et al. 2016).

3 Good Practices of User Experience and Design Research for mGov and eGov Services

As a result of the literature review, summarizing the different methods and results from the relevant results, this section summarizes ten good practices for mGov and eGov services to follow in pursuing a higher level of user experience and design for their services. These good practices should be seen as guidelines or a tool kit in the design of new or improving existing mobile government and egovernment services.

1. Learnability

According to (Isagah and Wimmer 2018) learnability is characterized that the user is able to easily learn, use and remember. In the context of both eGovernment and mGovernment applications, learnability means that the user would easily learn how to use the app or service, not have any difficulties using it and finally, easily remember how to use the app or service or how to find certain information within the application. Learnability contributes to the increase of user-friendliness in the short term and user acceptance in the long term. Thus, learnability is also a way to overcome the third challenge, which states that many developers and designers of mGov use approaches to develop working mobile applications. This requirement usually takes precedence over usability.(Isagah and Wimmer 2018; Kureerung and Ramingwong 2019a), where developers would only focus on technical aspects rather than on usability aspects of the solution.

2. Minimalistic and simple design

The need for a minimalistic and simple design of the service that allows the user to focus on the important functions of the services. Simplicity automatically increases accessibility, which means that no user groups are excluded because they lack certain capabilities. Therefore, this good practice is to provide a barrier-free and user-friendly solution in order to overcome the fourth challenge which is that demographics, political status, familiarity with technology, trust, and the nature of the service being offered have the most impact on the usability of mGov services which highlights the inevitable different characteristics of users (Chang et al. 2020a, 2020b; da Silva and Freire 2020; Isagah and Wimmer 2018; Kö et al. 2018; Kureerung and Ramingwong 2019a; Lönn et al. 2016) that make it difficult to develop a service that fulfills the needs of different user groups.

3. Language

Language refers to the language in which the eGov or mGov service is offered. The goal of this good practice is to ensure that the language used in the service is one that is understood by a broad user group. This can be fulfilled by offering one common

language or many. However, this good practice goes beyond offering various languages for the service, but to ensure that the language used is presented in a simple and clear way that any average user would understand despite their technical or legal understanding. Similarly, to a simple design, language can help to overcome the differences between user groups.

4. User readable terminology

The User Readable Terminology good practice implies that all the terminology (labels, buttons, messages, etc.) is understandable for users with little technical understanding. This should also include users that are new to the software and the subject. This good practice does not only enhance usability, but it also guarantees that no user group is excluded. This is particularly important regarding overcoming the forementioned challenge about how demographics, political status, familiarity with technology, trust, and the nature of the service being offered have the most impact on the usability of mGov services.

5. Help & feedback

This good practice implies offering a "helpdesk" for users that answers any questions that may arise during the user experience. Whenever the user is not able to proceed within the application, he or she must be able to get assistance. This assistance can be provided either by means of direct interaction with a team or a software in the background e.g. through a chatbot, or by means of simple clickable "i" that provides the user with additional information. Feedback in mobile or web-based applications refers to patterns that "inform the users about the status of the operations they are conducting". Such patterns include for example notifications or haptic output (da Silva and Freire 2020; Hoober and Berkman 2011). However, feedback can also mean that the user provides feedback to the developer. Both help and feedback contribute to an enhanced user-friendliness.

6. Error handling

Error handling is an important step within the development of any application and has been described in many user design studies. This good practice implies the involvement of end users to ensure that all predictable cases the system hinders the user to make mistakes. But the system should not just block an operation. Instead, it should explain to the user why this operation is not available at the moment. If there is an error, or the user makes a mistake the system should provide clear and understandable cause, also giving the user clear instruction on how to fix it. It shows that there is a strong interdependency between error handling and feedback and therefore, both must be installed.

7. Search & filter

One of six interaction design patterns defined by Hoober and Berkman is the search & filter design pattern (da Silva and Freire 2020; Hoober and Berkman 2011). As in any other web-based or mobile application, there must be a way for the user to search for certain information, data or functionality through a search engine implemented in the application. Even if all good practices described in this chapter are fulfilled by the developer, some users might prefer to look for data through a search engine instead of using other functionalities that are already there. Another add-on that comes along with that search engine is a filter. Such a filter makes it much faster for the user to find what he or she is looking for.

8. Operability

Operability or adaptability is a good practice where the User Interface developed must be adaptive, where the content is presented to users in a high quality way despite the size or device. According to (Isagah and Wimmer 2018, 2017) operability stands for (a) suitability for the device, and (b) conformity of the device with user expectation. Independently of the definition, the developer must make sure that the service can be accessed through any device. This good practice addresses another challenge that many developers of mGovernmental services have faced is the huge variety of devices available on the market, through which the user can access a service. This challenge requires that any solution is operable and supported by all mobile devices available on the market.

9. Placement of information

Two good practices for user interaction and design of mGov services have been already studied intensively in case studies and empirical work with citizens by (Kureerung and Ramingwong 2019a), (Chang et al. 2020a, 2020b; da Silva and Freire 2020; Isagah and Wimmer 2018, 2017). One important factor is the right placement of information within the application. It has been shown that a straight-forward layout and arrangement of instructions and functionalities is crucial for the usability of the service. Also, overlaps and replications of text and generally, large quantities of text should be avoided. The user must always be clearly directed to the most important functionalities.

10. Use of colors

This good practice is one that has reflected its importance in case studies and empirical studies with citizens is the right choice of colors (Chang et al. 2020a, 2020b; Kö et al. 2018). Not only the wrong choice of colors can negatively impact the user, but also the inconsistent use of logos or corporate identity of a service provider. Moreover, icons play an important role in the user experience. Well-designed and well-placed icons can be a benefit for the user (Chang et al. 2020a, 2020b; Isagah and Wimmer 2018). All in all, the "look and feel" of the application for the user must be as appealing as possible. The aforementioned studies show, that the involvement of citizens in the user design process can help a lot to design a user-friendly service that attracts many users.

4 Conclusion and Future Work

As a result of exploring related work, there has been a variety of different methods or frameworks that have been used to study the User Experience and Design research of mGovernment and eGovernment services. They have explored and described different design approaches and characteristics that are used for various eGov or mGov services and portals. Many previous research conducted combined multiple usability evaluation models during their process.

After summarizing which methods, models, and approaches were used in the most suitable related work, a summary of their results was also made and their suggestions. This process of establishing an overview set the foundation for creating the ten good practices for developers and other interested stakeholders in establishing User Experience and Design Good Practices in a simple and efficient way. While User Experience and Usability are often thought of at the end of the development process, this is something that should be integrated throughout the entirety of development. With integrating or having good practices in mind of how to make future services with a higher level of user experience, could assist greatly in user adoption of these services.

These good practices also set a foundation for future work for research. These good practices should be set to be tested, evaluated, refined, or expanded in an academic field as well as in a practice. Future work could include a greater amount of User Experience and Design models in its consideration, including ones that have not already been studied with in regard to mGov or eGov services.

Overall, this is a continuous work to assist in improving the integration of User Experience and Design Research into practice of future mGov and eGov services. While this paper has been focused on mGov and eGov services, these good practices are broad enough to apply to many different situations regarding improving services and their User Experience.

References

Abran, A., Khelifi, A., Suryn, W., Seffah, A.: Usability meanings and interpretations in ISO standards. Softw. Qual. J. **11**(4), 325–338 (2003)

Alcaraz-Quiles, F., Urquia-Grande, E., Muñoz-Colomina, C., Rautiainen, A..: E-government implementation: transparency, accessibility and usability of government websites, In: International E-Government Development, pp. 291–306 (2018)

Cahyono, T.A., Susanto, T.D.: Acceptance Factors and User Design of Mobile e-Government Website (Study Case e-Government Website in Indonesia). Indonesia (2019)

Chang, D., Li, F., Huang, L.: A User-centered Evaluation and Redesign Approach for E-Government APP (2019) (2020a)

Chang, D., Li, F., Huang, L.: A User-centered Evaluation and Redesign Approach for E-Government APP (2020b)

da Silva, L.F., Freire, A.P.: An investigation on the use of interaction design patterns in brazilian government mobile information systems. In: Brazilian Sympostum on Information Systems, pp. 3–6. Brazil

Harrison, R, Flood, D., Duce D.: 2013. Usability of mobile applications: literature review and rationale for a new usability model. J. Interact. Sci. **1**(1), 1 (2013). (2020)

Hoober, S., Berkman, E.: Designing mobile interfaces: patterns for interaction design (2011)

Hussain A., and Kutar, M.: Usability metric framework for mobile phone application. In: 10th Annu. Postgrad. Symp. Converg. Telecommun. Netw. Broadcast, no. June 2009, pp. 1–5 (2009)

Isagah, T., Wimmer, M.A.: Mobile government applications: challenges and needs for a comprehensive design approach. In: ICEGOV. Presented at the ICEGOV'17, pp. 423–432. New Delhi, India (2017)

Isagah, T., Wimmer, M.A.: Addressing requirements of M-government services: empirical study from Designers' perspectives. In: ICEGOV '18. Galway, Ireland (2018). https://doi.org/10.1145/3209415.3209469

Kirakowski, U.E. Solutions: SUMI: the Software Usability Measurement Inventory, no. October 2006 (2018)

Kö, A., Molnar, T., Matyus, B.: A User-centred Design Approach for Mobile-Government Systems for the Elderly. Hungary (2018)

Kureerung, P., Ramingwong, L.: A Framework for Usability Design to Promote Awareness of Information Disseminated via Mobile Government Applications. Thailand (2019)

Lönn, C.-M., Uppström, E., Nilsson, A.: Designing an M-government solution: enabling collaboration through citizen sourcing, In: ECIS 2016, p. 68. Turkey (2016)

Macleod, M., Bowden, R., Bevan, N.: The MUSiC Performance Measurement Method (1993)

Nielsen, J.: Enhancing the explanatory power of usability heuristics. In: Proc. SIGCHI Conf. Hum. factors Comput. Syst. Celebr. Interdepend. - CHI '94, pp. 152–158 (1993)

Saleh, A., Bintiisamil, R., Fabil, N.B.: Extension of pacmad model for usability evaluation metrics using goal question metrics (Gqm) approach. J. Theor. Appl. Inf. Technol. **79**(1), 90–100 (2015)

Seffah, A., Donyaee, M., Kline, R.B., Padda, H.K.: Usability measurement and metrics: A consolidated model. Softw. Qual. J. **14**(2), 159–178 (2006)

Susanto, T.D., Diani, M.M., Hafidz, I.: User acceptance of e-government citizen report system (a Case Study of City113 App). In: 4th Information Systems International Conference 2017, pp. 560–568. Bali, Indonesia (2017)

Talukder, S., Chiong, R., Corbitt, B., Bao, Y.: Critical factors influencing the intention to adopt m-Government services by the elderly. J. Glob. Inf. Manag. **4**, 74–78 (2020)

Smart Solutions in AgriTech: Research Trajectories in the Digital Transition

Salvatore Sapienza(✉)

Department of Legal Studies, University of Bologna, Bologna, Italy
salvatore.sapienza@unibo.it

Abstract. The following study investigates and identifies research trajectories pertaining to the digital transition of agriculture and food production. While a vast amount of research aims to discover new technologies, or to apply them in novel ways, their large-scale implications as regards data ownership and data governance are relatively overlooked. Regulatory interventions are demanded to steer data ownership and data governance towards the 'common good'. It is thus necessary to identify how research can contribute to the discussion on sensitive areas of policymaking that have been the object of environmental regulation, including the EU Green Deal and UN Sustainable Development Goals. In the light of this necessity, this paper identifies issues with ethical relevance emerging from the adoption of new technologies in agritech, including Artificial Intelligence techniques and Internet of Things applications. To do so, this study attempts to systematise and categorise existing research trends by clearly identifying their scope and understanding the relationships that exist among them. The results of this enquiry show that five interconnected research trajectories - namely, technical solutions, data governance, data ownership, ethics and law - can foster the discussion on agritech transition. The connections between these research areas can be understood in terms of a descriptive and a prescriptive perspective.

Keywords: Agritech · Smart farming · Green transition · Green deal · Data governance · Data ownership

1 Introduction

The European Union (EU) Green Deal[1] aims at fostering the transition towards eco-friendly and sustainable economic models of agriculture and food production.

[1] Communication From the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions "The European Green Deal" COM/2019/640 final.

S. Sapienza—This study has been funded by Programma Operativo Nazionale (PON) "Ricerca e Innovazione" 2014–2020 CCI2014IT16M2OP005, by the Italian Ministry of University and Research approved with Decision CE C(2015)4972 of 14 July 2015, in the light of D.M. 1062 of 10 August 2021.

A. Kö et al. (Eds.): EGOVIS 2022, LNCS 13429, pp. 150–162, 2022.
https://doi.org/10.1007/978-3-031-12673-4_11

This intervention is aligned with the United Nations Sustainable Development Goals 2 ("Zero Hunger"), 8 ("Decent Work and Economic Growth"), and 12 ("Responsible Consumption and Production"). The 'great promise' of Big Data and Artificial Intelligence (AI) is to foster this transition by rationalising a wide set of operations. They range from public decision-making, to data availability for consumers, 'smart' solutions for food business operators, and so forth. Implementing policies seem urgent to implement the new political agendas set in the aftermath of COVID-19 pandemic, at national (e.g. the Italian National Plan Resistance and Resilience), regional and international level.

The digital transition of agriculture and food production is placed within the so-called Fourth Revolution [1] and raises interesting questions as regards the priorities to be set. Regulators should steer such "Green & Blue" [2] transition toward the 'common good' and their decisions should keep into account not only the economic growth, but also other factors such as food safety and long term sustainability. However, conflicting interests among the involved stakeholders - including consumers, the food industry, local farms, small-medium enterprises - might hamper the finding of shared solutions and policies.

Research could foster the discussion about such policy-making by identifying the goals and the means of agricultural policies aimed at digitalising farming. However, research areas focused on the adoption of agritech solutions seem quite fragmented. In particular, a gap can be observed between the discussion on 'smart' methods and the general policies set by the political agendas and supported by the legislation. Therefore, this paper aims to cast light on possible research trajectories in agritech with the goal of providing an overview of research topics and identifying their relationships.

The study adopts a cross-disciplinary method of enquiry, especially when establishing a connection between research areas with the goal of filling the aforementioned gap. Rather than a complete and systematic literature review, this paper identifies general trends across several research areas. The qualitative systematisations provided in this study are thus meant to identify common patterns in agritech that raise questions relevant to policymaking. Primary sources of investigations have been extensive reviews in agritech, identified by certain keywords[2]. Other sources have been selected from these reviews to investigate the use of established and newly-adopted technologies identified as such by the reviewers. In some of the extracted reviews, emerging trends have also been linked to existing issues in new technologies and discussions about their ethical and moral dimensions (e.g., AI ethics). From these premises, the overview of research trajectories has been developed by systematising these findings. The

[2] [["Big AND Data" OR "Artificial AND Intelligence"] AND ["Smart AND Farming" OR "Digital AND Agriculture" OR "Agritech"] AND "Review"] was the query used to identified papers that contain literature reviews on the topic under scrutiny. Research was performed on academic databases (namely, Scopus, Web of Science, Google Scholar) using title-based and topic-based queries and refined to avoid duplicates.

proposed categorisation consists of five intertwined research areas, namely technical solutions, data governance, data ownership, ethics and law.

The paper is divided in four sections following this Introduction. Section 2 briefly summarises the previous works in the area of agritech and data governance, whereas Sect. 3 contextualises these works within emerging trends in new technologies. Section 4 identifies the research avenues (or trajectories) for future studies in this area and briefly discusses their possible implications. Concluding remarks summarise the main findings of this work and illustrate the next steps for this research.

2 Current Developments in Agritech

Reviewers (e.g. [3–5]) have identified several ongoing trends in agriculture 4.0, also known as precision agriculture (PA) or smart farming. These trends are not characterised by the use of a single technology, but they are made possible by the combination of multiple ICTs aimed at improving the profitability and sustainability of farming [6], e.g., by increasing the degree of automation in certain tasks or by improving decision systems [7]. Despite being dated back to 1980 s [8], digitalised agriculture is now scalable due to lower costs in microprocessors and new technologies such as cloud computing or mobile applications.

Geographic Information Systems (GISs) combine spatial data with soil information [9]. While this technology is not entirely new since it was originally proposed by FAO in the 1990s [10], its deployment has been recently proven to be successful in several case studies discussed by the literature in the fields of urban extension, deforestation and climate change [11]. Soil sensors aggregate these data with satellite images [12]. Recent data acquisition trends also relate to the use of Unmanned Arial and Ground Vehicles (UAVs and UGVs respectively) [13].

The deployment of Internet of Things (IoT) technologies - smart objects connected to each other through a wireless network infrastructure [14] - has been fruitful in crop and resource management and monitoring, with increases in quality and quantity of the crop yield [15]. Data acquisition is also related to environmental information, such as moisture, temperature, and light [12].

Furthermore, the proliferation of mobile applications for Agritech has been observed by reviewers [16], also in developing countries [17]. Mobile applications allow data access and visualisation, thus contributing to the decision-making processes regarding crop nutrition, fertilisation and irrigation. Smartphone cameras can be used for crop protection and diagnosis purposes.

Blockchain technologies are deemed to be game-changing for traceability across the whole food chain [18]. Use cases conducted by IBM and Walmart [19] and Carrefour [20] show the potential of this technology in food logistics. In smart farming, blockchain can be used in the early stages of the food chain by the seed provider and the producer [21]. Given the presence of several actors involved in the agricultural supply chain and the international nature of modern trades, blockchain can reduce the time required for manual checks and document processing [22], ultimately lowering production costs.

These technologies are, indeed, multifunctional. While they seem mainly oriented to support business activities of the food industry, they can also ease consumers' decision-making. The integration of these solutions goes under the names of "Traceability 4.0", "Smart Packaging" or "Smart Labelling". Recent studies [23] show that mobile applications equipped with augmented reality tools, including QR-codes, impact consumers' perception on product quality and origin, thus bringing consequences on the international food market. Traceability solutions are quite diverse and applicable to heterogeneous goods, including olive oil [24] and wine [25].

The integration of these technologies into interoperable models represent a noteworthy challenge. Different data sources should be integrated to extract meaningful information from them. Existing research has proposed an integration at the semantic level by means of Semantic Web Technologies [26], with possible benefits on data management and automated systems building [27]. However, these reviewers have identified a relative paucity of research in this area, which remains underutilised despite great potential.

The use of these technologies is also related to agricultural policymaking. Together with food producers, suppliers, and consumers, also policymakers can benefit from technologies such as remote sensing and data analytics to shape information-based governance models [28]. In particular, targeted policies for specific areas and real-time interventions become feasible when ICTs are spread throughout the food chain. Together with agricultural policies meant to support economic activities, deploying data-driven policies should also mean to foster sustainability [29].

To summarise, this section has highlighted how digital solutions are reshaping agriculture under several perspectives. The three main pillars of these developments consist in fostering agribusiness, empowering consumers, and sustaining decision-making processes. While existing research shows how each of these pillars can individually benefit from digital transition in agriculture, it is still unclear what direction the involved stakeholders should follow and, eventually, how to find a unified strategy for the 'common good' to be adopted by decision-makers. The next section illustrates some of the issues emerging from this transition to be investigated by academic research.

3 Agritech and Emerging Issues

The following section is aimed at identifying current and novel issues that emerge from the digital transformation of agriculture specifically related to data and information used for decision-making processes. As it will be discussed, the identified issues can be subsumed within two general categories, namely data ownership and data governance.

Reviewers in [5] observe that "[d]ata ownership, protection, and security are perceived as not sufficiently close to farmers' needs, thus becoming threats to be mitigated, if not completely avoided. In more words, nowadays, digital solutions for [smart farming, *ed.*] are under-utilised because practitioners fear data misuse

and the loss of control over their business". Similarly, scholars have discussed data ownership in farming under the perspective of finding the entity entitled to exploit the economic value of information [30,31] or in terms of the relation and the distribution of power between companies providing digital infrastructures *vis-a-vis* farmers [3,32].

A certain degree of lexical ambiguity can be found in this literature, possibly due to the diverse background of the scholars in this field. For instance, 'data privacy', 'data confidentiality', 'data protection', and similar expressions are used interchangeably. Most of them, however, regard different issues in the realm of personal data rather than other information that does not refer to an identified or identifiable individual[3]. While it is true that some data in this domain could potentially qualify as personal data for the purposes of data protection law - for instance, food consumption data [33] or farmers' location - the majority of the digital information produced in farms is constituted by the data on temperature, humidity, nitrogen levels, geographical information, water use, vehicle data, ecc. captured by sensors deployed in the fields.

When focusing on non-personal information from a data-centric perspective, the clashes between confidentiality and data re-use, the preservation of competition, and data sovereignty can be *prima facie* observed as the most relevant concerns emerging from digitalised agriculture. While these issues have been widely discussed in fields such as data protection, data security and international data transfer with regards to personal information, little discussion has been made over non-personal data, and even less about farm-related information.

Furthermore, the centrality of (big) data in the current debate shall not nudge researchers, practitioners and policy-makers into believing that algorithms should be relegated outside the discussion. In fact, while the crucial role of data-as-an-asset has been fruitfully captured by the contemporary debate in agritech information, the way we 'make sense' of data - i.e., algorithms - seems absent from the current literature.

However, scholars have progressively shifted from the data-centric to the algorithm-centric [34] level of abstraction [35]. Agritech is not different from other domains. With AI and automated decision systems made possible by the abundance of data provided by the applications discussed in Sect. 2, issues already identified by the technical and ethical literature [36] will eventually be detected in agritech. While AI systems can fruitfully enhance decision-making at every level (food business operators, consumers, policy makers), it might be the case that some risks outweigh the opportunities offered by such novel solutions. In particular, two possibilities can occur.

On the one hand, high-level and horizontal issues (i.e., pertaining to all AI applications meant to support decision-making) can be identified. *Inter alia*, algorithmic explainability can be seen a paramount requisite for AI systems meant to support public decision-making [37], whereas human oversight and accountability frameworks are necessary to attribute the responsibility for the

[3] That is the definition of 'personal data' under Article 4(1) of the EU General Data Protection Regulation (Reg. 2016/679).

actors involved, especially in the fields of product liability and in the business-to-business relationship between farmers and ICT suppliers.

On the other hand, applied and vertical issues (i.e., specifically pertaining to the context of agritech AI-powered solutions) raise attention. For instance, it might be the case that disparate access to 'smart' solutions amplify the diversity between different areas and regions, or that AI-suggested solutions prioritise profit over environmental concerns (e.g., in the case of fertilisation and the use of plant protection products vìs-á-vìs environmental concerns).

To summarise, two main areas of potential investigation have been identified in addition to technical advancements. On the one hand, the quasi-proprietary relationship between a legal entity and some information - identified as 'data *ownership*' - has been detected as one of the key issues. As a research trend data ownership investigates the possibility to access, analyse, and store (i.e., 'use') data, both from technical and legal perspectives. In fact, the rules governing the usage of data by a qualified entity can be expressed in design requirements (e.g. via APIs, data portability instruments, etc.) as well as in legal terms (e.g. in 'hard laws', contracts, codes of conducts). On the other hand, data *governance* expresses the procedures that govern the creation of data-related rules, the regulatory instruments adopted to create, modify and to amend these rules, and the objectives that they are meant to fulfil.

Data ownership and data governance interplay significantly. Any entity qualified to use a given dataset is also entitled to determine the purposes of the such usage (e.g. business analysis) and the rules that govern it (e.g., restricting access to third parties). In other words, who owns the data usually decides how to use it. This faculty is subject to changes when a coercive regulatory intervention, such as 'hard law', determinates from the outside the behaviour of the data owner by restricting or enlarging its possibilities. Who shall determinate such behaviour, by what means, and for what purpose is the research trajectory that goes under the name of data governance.

4 Possible Research Trajectories and Their Connections

The previous section has identified data ownership and data governance as two central research areas. A combined discussion about these topics is necessary to understand digital transformations across the whole food chain. In particular, it has to be noted that, without regulatory interventions, a discrete margin of appreciation is left to the involved stakeholders in determining governance choices that have consequences on ownership (e.g., allowing or restricting data access to third parties).

Research has discussed how ethics play a significant role in determining governance and, consequently, ownership choices in new technologies. For instance, re-thinking data ownership in agriculture has been considered necessary, also on the basis of ethical considerations [38]. In the field of algorithmic governance, it has been observed that "[e]thics plays a key role in this process by ensuring that regulations of AI harness its potential while mitigating its risks" [39], when AI

solutions are meant to promote a general interest and the social good [40], as with the case of agritech. In sum, the digital transition of farming and related activities require a careful balance of risks and opportunities.

On the contrary, it might be argued that law is sufficient in addressing conflicting interests (e.g., between data confidentiality, openness and re-use). In the European legislation, data flows might even seem over-regulated: the contextual application of the General Data Protection Regulation[4], the Non-personal Data Regulation[5], the Data Act[6], the Data Governance Act[7], and the forthcoming Artificial Intelligence Act[8], jointly with sectoral food law legislation, e.g., in the field of food safety[9] suggests that EU decision-makers have already decided for a clear direction on how to solve ownership and governance issues.

Some clarifications on the role of ethics are then necessary. First, ethics is not necessarily meant to replace the normative role of existing laws. Following Floridi's interpretation of the role of ethics in this debate [41], it shall be deemed either as a challenge to the existing legislation to be used in a *de iure condendo* perspective ("hard ethics") or as what ought and ought not to be done over and above the existing regulations ("soft ethics", among which we can include 'data ethics' [35]). Furthermore, principles can be placed between law and design specifications [42].

As argued above, several pieces of legislation regulate the use of data in the European Union. Therefore, data ownership and data governance are (also) co-designed by such several provisions and, in turn, technical solutions are impacted by these rule-shaped ownership and governance models. While the law sets the "how" ownership and governance should be shaped, ethics contributes to identify the "what-for" certain choices are made. With new technologies and applications emerging and in a transition phase, ethics (as a discipline) has the capability to receive inputs emerging from the current status of ownership and governance and

[4] 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 [2016] OJ L 119/1.

[5] 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 [2018] OJ L 303/59.

[6] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data.

[7] Proposal for a Regulation of the European Parliament and of the Council on European data governance - COM/2020/767 final.

[8] Proposal for a Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - COM/2021/206 final.

[9] Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety [2002] OJ L 31/1. This Regulation has been deeply amended by the 2021 Transparency Regulation to provide more public access to food safety information.

support legal operators (hence, not only lawmakers but also judges and scholars) in identifying the best option among the many possibilities offered by technical solutions.

This is the case of agriculture and food production. As a market sector, it constitutes a field of research dense with opportunities and risks. Their balance determines the quality of life of consumers and impacts the whole population of a given region, let alone economic consequences for food business operators. Moreover, data-centric regulations should match the goals of regional or international green transition programmes such as the EU Green Deal and UN Sustainable Development Goals. Ethics can contribute to the identification of a general trend, or solid pillars, to support the interplay between data-oriented provisions and digital transition programmes.

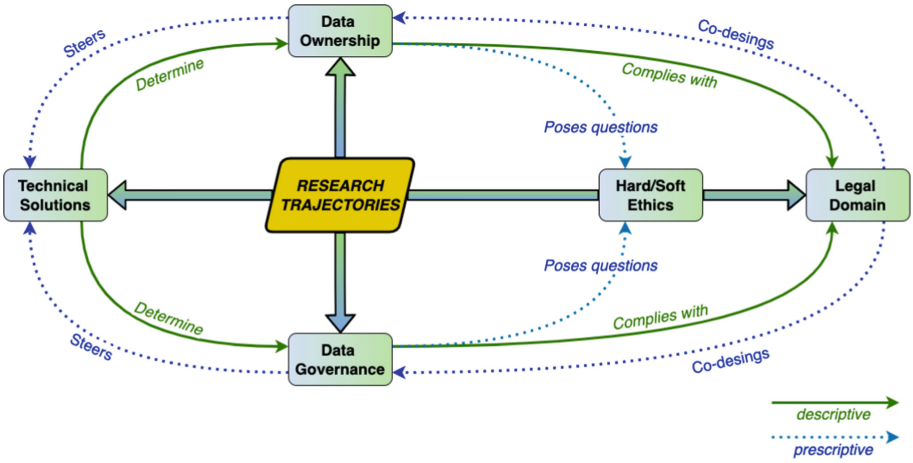

Fig. 1. Possible research trajectories: technical solutions, data ownership, data governance, ethics and law. Descriptive and prescriptive relationships are identified by the arrows

Figure 1 displays the research trajectories identified by this paper. They are not intended to be ranked or classified hierarchically and the order of presentation in this study is only for descriptive purposes. First, the domain of technical solutions is necessary to correctly understand what is (or can be) done with emerging technologies and applications in the realm of Big Data and AI. Such correct understanding is necessary also to prevent, in a later stage, that legal solutions impose excessive or impossible burdens on ICT providers or users.

Then, data ownership and data governance are *prima facie* determined by technical solutions previously discovered. As a research trajectory, data ownership poses the question "Who owns the data?" and investigates the quasi-proprietary relationship between stakeholders and personal or non-personal information, access conditions, data transfers, data processing, data warehousing, and so forth.

Instead, data governance seeks answers to the question "How is data managed?". In particular, it focuses on the allocation of decision-making powers in data-related questions, what regulatory instruments are adopted in the definition of these rules, how they are enforced, what are the general goals of these rules, etc.

Then, data ownership and data governance have to comply with legal requirements, in particular the complex legislative scheme described above. Therefore, a scrutiny on their level of fitness to the current regulation is required. This first set of questions regards any scenario at any given time and a *descriptive* goal.

While being compliant with legal requirements, ownership and governance models pose questions that have an ethical relevance, as they demand morally-relevant choices. As argued above, technologies can be used to prioritise food production, lower costs for consumers, increase access to food, perhaps at the expense of increased land usage and environmental concerns. Ethical dilemmas are not new to agriculture and food production, especially in bioethical research [43].

Today, these trajectories require the inclusion of approaches capable of casting light on novel technologies. After all, and on a positive note, a certain degree of consistency on high-level bioethical principles and AI principles has already been identified by qualified working groups [44], thus paving the way for a progressive integration at applied levels such as agritech. Moreover, this method has been proved to be correct elsewhere, when applied to food safety [33,45]. Furthermore, ethical contributions might be necessary to foster the current debate on forthcoming pieces of legislation in the EU (Data Act and Artificial Intelligence Act).

Therefore, it might be necessary to identify, alongside the legal domain, an ethical research trajectory capable of capturing instances of 'what for' discussions emerging from the debate on data ownership and data governance. Therefore, differently from the other relationships, the one between ownership and governance trajectories is not descriptive, but it aims at discovering "hard ethics" or "soft ethics" approaches to the existing regulation, hence being *prescriptive* or ought-oriented.

Such prescriptive relationship between research trajectories is not isolated. Besides setting general principles and constituting powers [46], law is also inherently *prescriptive*, thus mandating certain behaviours that shall be followed when implementing data ownership and data governance models. Therefore, legal provisions co-design these models, together with other factors such as economic considerations, global policies, power relationships, and so forth.

In turn, ownership and governance models steer the development and the design of the technical solutions deployed in a given environment and, ultimately, are a contributing factor to their adoption. By means of this prescriptive relationship, technical solutions eventually mirror the regulations and contribute to achieve policy-makers goals.

Let us briefly discuss the main implications of the proposed classification. First, the interdependences of the identified research trajectories entail that their

integration is necessary to solve the complexities of agricultural transition. In other words, decision-makers should rely and be informed on these five areas to identify the pros and cons of digital transition policies that are relevant to agritech. Secondly, ethics play a key role in mediating different positions and being, either as a 'hard' or a 'soft' ethics, the pivot of the research trajectories. In this regard, agritech is consistent with other research domain, including bioethics and AI ethics. Finally, the circularity of the proposed model implies that, while no research area is more important than the others, new advances in each area have implications for the others. This map can contribute to identify the nature of such consequences from a theoretical point of view.

5 Concluding Remarks

This study has contributed to identify possible research trajectories in agritech data-related issues. A systematisation of the existing research trends and methodologies was necessary to identify a common ground for further discussion. In summary, the paper has identified five main research tracks in agritech data-related issues: technical solutions, data ownership, data governance, ethics and law. The existence of descriptive and prescriptive relationships among them has also been detected. In this sense, this study in an original contribution to an ongoing, yet highly fragmented, discussion.

However, this study is also limited because it only provides a short description of each research trajectory, without specifying specific methodologies to tackle the peculiar issues of each research area. While each of them remains independent, it is necessary to contextualise them within a broader framework to verify the technical feasibility, the connections with data ownership and data governance, the ethical implications, and the compliance to legal requirements of every newly-developed research product that can contribute to the 'common good' in the digital transition of the agrifood sector. The next steps seem likely to put the research trajectories into practice. Naturally, it might the case that research products do not cover all the aforementioned issues in detail. However, a generic assessment of their implications or, vice versa, their check under a different perspective seems ultimately beneficial for the interplay between research community and to policy-makers.

References

1. Floridi, L.: The Fourth Revolution: How the Infosphere is Reshaping Human Reality. OUP Oxford, London (2014)
2. Floridi, L.: The green and the blue: naïve ideas to improve politics in a mature information society. In: Öhman, C., Watson, D. (eds.) The 2018 Yearbook of the Digital Ethics Lab. DELY, pp. 183–221. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17152-0_12
3. Wolfert, S., et al.: Big data in smart farming-a review. Agric. syst. **153**, 69–80 (2017)

4. Miles, C.: The combine will tell the truth: on precision agriculture and algorithmic rationality. Big Data Soc. **6**(1), 2053951719849444 (2019)
5. Bacco, M., et al.: The digitisation of agriculture: a survey of research activities on smart farming. Array **3**, 100009 (2019)
6. van Es, H.M., et al.: Digital agriculture in New York State: report and recommendations. Cornell University, Ithaca, NY (2016)
7. Sonka, S.: Big data and the ag sector: more than lots of numbers. In: International Food and Agribusiness Management Review, vol. 17, no. 1030-2016-82967, pp. 1–20 (2014)
8. Mulla, D.J.: Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. Biosyst. Eng. **114**(4), 358–371 (2013)
9. Lytos, A., et al.: Towards smart farming: systems, frameworks and exploitation of multiple sources. Comput. Netw. **172**, 107147 (2020)
10. Rossiter, D.G.: ALES: a framework for land evaluation using a microcomputer. Soil Use Manage. **6**(1), 7–20 (1990)
11. Bielecka, E.: GIS spatial analysis modeling for land use change. a bibliometric analysis of the intellectual base and trends. Geosciences **10**(11), 421 (2020)
12. Balafoutis, A.T., et al.: Smart farming technologies – description, taxonomy and economic impact. In: Pedersen, S.M., Lind, K.M. (eds.) Precision Agriculture: Technology and Economic Perspectives. PPA, pp. 21–77. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68715-5_2
13. Vu, Q., Raković, M., Delic, V., Ronzhin, A.: Trends in development of UAV-UGV cooperation approaches in precision agriculture. In: Ronzhin, A., Rigoll, G., Meshcheryakov, R. (eds.) ICR 2018. LNCS (LNAI), vol. 11097, pp. 213–221. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99582-3_22
14. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming competition. Harvard Bus. Rev. **92**(11), 64–88 (2014)
15. Dagar, R., Som, S., Khatri, S.K.: Smart farming-IoT in agriculture. In: 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, pp. 1052–1056 (2018)
16. Mendes, J., et al.: Smartphone applications targeting precision agriculture practices? A systematic review. Agronomy **10**(6), 855 (2020)
17. Baumüller, H.: Towards smart farming? Mobile technology trends and their potential for developing country agriculture. In: Handbook on ICT in Developing Countries, pp. 191–201. River Publishers, Delft (2017)
18. Tian, F.: An agri-food supply chain traceability system for China based on RFID & blockchain technology. In: 2016 13th International Conference on Service Systems and Service Management (ICSSSM), pp. 1–6. IEEE (2016)
19. Kamath, R.: Food traceability on blockchain: Walmart's pork and mango pilots with IBM. J. Br. Blockchain Assoc. **1**(1), 3712 (2018)
20. Carrefour. Food blockchain. https://www.carrefour.com/en/group/food-transition/food-blockchain (2018). Accessed 05 May 2021
21. Kamilaris, A., Fonts, A., Prenafeta-Boldú, F.X.: The rise of blockchain technology in agriculture and food supply chains. Trends Food Sci. Technol. **91**, 640–652 (2019)
22. Hoffman, A., Munsterman, R.: Dreyfus teams with banks for first agriculture blockchain trade (2018)
23. Penco, L., et al.: Mobile augmented reality as an internationalization tool in the "Made In Italy" food and beverage industry. J. Manage. Gov. **25**(4), 1179–1209 (2021)

24. Violino, S., et al.: Are the innovative electronic labels for extra virgin olive oil sustainable, traceable, and accepted by consumers? Foods **8**(11), 529 (2019)
25. Acuti, D., et al.: The effects of QR delivered content on perceived product value. J. Strateg. Mark. 1–23 (2020)
26. Hitzler, P., et al.: Advancing agriculture through semantic data management. Semant. Web **12**(4), 543–545 (2021)
27. Drury, B., et al.: A survey of semantic web technology for agriculture. Inf. Process. Agric. **6**(4), 487–501 (2019)
28. Ehlers, M.H., Huber, R., Finger, R.: Agricultural policy in the era of digitalisation. Food Policy **100**, 102019 (2021)
29. Garske, B., Bau, A., Ekardt, F.: Digitalization and AI in european agriculture: a strategy for achieving climate and biodiversity targets? Sustainability **13**(9), 4652 (2021)
30. De Beer, J.: Ownership of open data: governance options for agriculture and nutrition. Wallingford: Global Open Data for Agriculture and Nutrition (2016)
31. Rasmussen, N.: From precision agriculture to market manipulation: a new frontier in the legal community. Minn. JL Sci. Tech. **17**, 489 (2016)
32. Carbonell, I.: The ethics of big data in big agriculture. Internet Policy Rev. **5**(1), 13 (2016)
33. Sapienza, S., Vedder, A.: Principle-based recommendations for big data and machine learning in food safety: the P-SAFETY model. AI Soc. 1–16 (2021). https://doi.org/10.1007/s00146-021-01282-1
34. Yeung, K.: Algorithmic regulation: a critical interrogation. Regul. Gov. **12**(4), 505–523 (2018)
35. Floridi, L., Taddeo, M.: What is Data Ethics? Phil. Trans. R. Soc. A **374**(2083), December 2016
36. Mittelstadt, B.D., Floridi, L.: The ethics of big data: current and foreseeable issues in biomedical contexts. Sci. Eng. Ethics **22**(2), 303–341 (2015). https://doi.org/10.1007/s11948-015-9652-2
37. Sapienza, S.: Explanations in risk analysis: responsibility, trust and the precautionary principle. Forthcoming in the Proceedings of the 4th International Workshop on eXplainable and Responsible AI and Law (XAILA2021@ICAIL) at the 18th International Conference on Artificial Intelligence and Law (ICAIL 2021) (2021). https://www.geist.re/_media/xaila:xaila2021icail_paper_6.pdf
38. van der Burg, S., Wiseman, L., Krkeljas, J.: Trust in farm data sharing: reflections on the EU code of conduct for agricultural data sharing. Ethics Inf. Technol. **23**(3), 185–198 (2020). https://doi.org/10.1007/s10676-020-09543-1
39. Taddeo, M., Floridi, L.: How AI can be a force for good. Science **361**(6404), 751–752 (2018)
40. Floridi, L., Cowls, J., King, T.C., Taddeo, M.: How to design AI for social good: seven essential factors. In: Floridi, L. (ed.) Ethics, Governance, and Policies in Artificial Intelligence. PSS, vol. 144, pp. 125–151. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-81907-1_9
41. Floridi, L.: Soft Ethics and the Governance of the Digital. Philos. Technol. **31**(1), 1–8 (2018). https://doi.org/10.1007/s13347-018-0303-9
42. Pagallo, U., Casanovas, P., Madelin, R.: The middle-out approach: assessing models of legal governance in data protection, artificial intelligence, and the Web of Data. Theory Pract. Legislation **7**(1), 1–25 (2019)
43. Jasanoff, S.: The Ethics of Invention: Technology and the Human Future. WW Norton & Company, New York (2016)

44. AI4People. AI4People — Atomium. (2018). https://www.eismd.eu/ai4people/
45. Sapienza, S., Palmirani, M.: Emerging data governance issues in big data applications for food safety. In: Kő, A., Francesconi, E. (eds.) EGOVIS 2018. LNCS, vol. 11032, pp. 221–230. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98349-3_17
46. Searle, J.R.: How to derive 'ought' from 'is'. In: Hudson, W.D. (ed.) The Is-Ought Question. CP, pp. 120–134. Palgrave Macmillan UK, London (1969). https://doi.org/10.1007/978-1-349-15336-7_13

Author Index