Advanced Sciences and Technologies for Security Applications

Reza Montasari *Editor*

Applications for Artificial Intelligence and Digital Forensics in National Security

Springer

Advanced Sciences and Technologies for Security Applications

Indexed by SCOPUS

The series Advanced Sciences and Technologies for Security Applications comprises interdisciplinary research covering the theory, foundations and domain-specific topics pertaining to security. Publications within the series are peer-reviewed monographs and edited works in the areas of:

- biological and chemical threat recognition and detection (e.g., biosensors, aerosols, forensics)
- crisis and disaster management
- terrorism
- cyber security and secure information systems (e.g., encryption, optical and photonic systems)
- traditional and non-traditional security
- energy, food and resource security
- economic security and securitization (including associated infrastructures)
- transnational crime
- human security and health security
- social, political and psychological aspects of security
- recognition and identification (e.g., optical imaging, biometrics, authentication and verification)
- smart surveillance systems
- applications of theoretical frameworks and methodologies (e.g., grounded theory, complexity, network sciences, modelling and simulation)

Together, the high-quality contributions to this series provide a cross-disciplinary overview of forefront research endeavours aiming to make the world a safer place.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors.

Reza Montasari
Editor

Applications for Artificial Intelligence and Digital Forensics in National Security

Springer

Editor
Reza Montasari
Swansea, UK

ISSN 1613-5113 ISSN 2363-9466 (electronic)
Advanced Sciences and Technologies for Security Applications
ISBN 978-3-031-40117-6 ISBN 978-3-031-40118-3 (eBook)
https://doi.org/10.1007/978-3-031-40118-3

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

Contents

Bias, Privacy and Mistrust: Considering the Ethical Challenges of Artificial Intelligence

Annie Benzie and Reza Montasari

Abstract The current landscape of artificial intelligence (AI) is complex, and is a source of hope and fear alike. It is a field which is constantly progressing, whilst demonstrating unforeseen challenges for creators and users. The benefits of using AI tools are clear, given it is now commonplace across the globe, finding its way into homes, schools, and workplaces alike. However, as reliant as society has become on advancing technologies, an increasing number of ethical challenges have been emerging, including bias, privacy violations, both leading to lack of trust. This chapter contextualises these issues by first presenting a short history of AI, including some challenges to its development. Following this, bias, privacy and mistrust are discussed, before solutions are suggested for future development and mitigating the stated troubling areas.

Keywords Artificial intelligence · AI · Machine learning · Ethical AI · Privacy · Bias · Trust

1 Introduction

In a world where technology is constantly developing and becoming ever more capable of solving complex issues, it is clear that there is demand for AI solutions in every sector, from finance to healthcare. Perhaps as a caveat to the benefits of enjoying these technological advancements, there are a growing number of emerging ethical challenges that are often perceived to be part and parcel of AI implementation, such as algorithmic bias and privacy violations. This chapter first considers the

A. Benzie (✉) · R. Montasari
School of Social Sciences, Department of Criminology, Sociology and Social Policy, Swansea University, Singleton Park, Swansea SA2 8PP, UK
e-mail: benzieannie@gmail.com

R. Montasari
e-mail: reza.montasari@swansea.ac.uk
URL: http://www.swansea.ac.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_1

evolution of AI, including its definition and how its purpose has been continuously adapted since its conception. A short history of the concept provided, alongside challenges to the development of AI, describing crucial setbacks and lessons learned as a result. Section 2 considers some of the ethical challenges that have subsequently arisen. Firstly, the meaning of bias is considered, presenting that meanings are widely contested and may vary depending on context, as well as types of bias, demonstrating a taxonomy of bias types in the AI pipeline. We consider meaningful metrics for measuring bias and how fairness may be codified. The section then considers privacy violations as a significant challenge to AI, including aspects such as lack of public understanding regarding what user data is being used for, and highlighting the ease with which violations may occur. Finally, the concept of mistrust is presented, both as a consequence of privacy violations and bias, as well as a unique challenge in its own right. Following this, Sect. 3 seeks to address the ethical challenges by considering potential methods which can be used to address them, such as increasing privacy through methods such as machine learning perturbation, and empowering users to trust technology through transparency and bias mitigation. Section 4 concludes and presents ideas for how further research may shape future progression of ethical AI.

2 The Evolution of Artificial Intelligence

2.1 Definitions

Definitions of Artificial Intelligence vary depending on context. For the purposes of this paper, AI will be defined as 'a broad set of methods, algorithms, and technologies that make software 'smart' in a way that may seem human-like to an outside observer' [22]. As discussed by Russell and Norvig [25], the human element of intelligence stems from computers being trained to make decisions in a rational manner by considering available options [25]. Machine Learning (ML) is the study of algorithms that are capable of learning relationships and patterns and using these as a basis on which to make decisions. ML problems may be classified into two categories, supervised and unsupervised learning. Supervised learning is named so due to the data being labeled in order to train the algorithm. For example, when considering face recognition, the machine learns based on images labeled as 'face' and 'non-face'. Training the algorithm like this allows it to predict whether a previously unseen face is a face or not. In an instance of unsupervised learning, images are not labelled.

2.2 A Short History

In order to fully assess the emerging ethical challenges and corresponding solutions, it is first necessary to consider the development of AI over its history, which stems back over 60 years. It is thought that the concept of 'Artificial Intelligence' was coined in 1956 by Professor J. McCarthy at Stanford University, Professor M. L. Minsky at the Massachusetts Institute of Technology, and Professors H. Simon and A. Newell at Carnegie Mellon University, along with C. E. Shannon at Bell Labs, N. Rochester at IBM [23]. At the time, their definition of AI covered machines that had the ability to understand, think and learn in a similar way to human beings, whereby they simulated human intelligence. Interestingly, definitions of AI often compete, as it could be argued that the concept of intelligence itself is subjective. Research over the years has considered related questions, such as 'what constitutes human intelligence?', and 'Is it possible for a machine to display intelligence, even if the process behind it does not emulate that which would be displayed by a human'?

When computers were designed, and during the years that followed, computers followed instructions. Machines did not have the capacity to learn from experience and past behaviors. This 'learning' aspect followed and allowed for prediction and classification. Machine learning was further revolutionised when scientists began implementing algorithms that had their design rooted in the human brain, using powerful neural networks. This was the beginning of deep learning, a powerful subset of machine learning, which has now revolutionised multiple fields, including finance, medicine, and healthcare.

It is no secret that AI has been the focus of much attention from industry, media, government, and consumers over recent years, thanks to the area of deep learning, which has been a game changer for the world of AI, helping tech giants such as Google to refine the accuracy of picture searches, for example [23]. The increased use of AI technologies in most sectors across society has increased expectations for its capabilities. For example, the use of the Watson system which was developed by IBM helped hospital staff to analyse millions of patient records for cancer treatments to help with future diagnosis. Notably increasing expectations of AI ability came in 2016, when AlphaGo algorithm beat the world champion of Go, with a score of 4:1. This came with a new lease of life in terms of global attention which sparked investment and interest in the development of intelligent machines. However, potential aside, AI is no stranger to a setback.

2.3 Challenges to the Development of Artificial Intelligence

According to Pan [23], AI development has been subject to three major setbacks since it was established in the 50s. Firstly, they present that in 1973, a report was published which concluded that research concerning machines, such as robots, held little to no value and the disappointing outputs that had been noted at that stage

indicated that research should therefore be halted. The second setback, Pan presents, is the failure of the Ministry of International Trade and Industry of Japan to develop a viable solution, after spending approximately 850 million dollars to develop a parallel-inference machine, capable of listening and speaking. Despite the project's failure in the early nineties, it could be argued that this project helped to drive AI development in the right direction, focusing on software instead of hardware, which should ultimately be designed to support the internal innovation. The third setback was the attempt to construct an encyclopaedia of knowledge by Stanford University in 1984. With the rise of the search engine, it was determined to be ineffective to attempt to transfer information from human to machine in the hope of it becoming an expert with human intelligence levels. Rather, as was learned from this failed project, machines may automatically learn information from their environment.

It is worth mentioning that it is only through such failures that scientists have come to have a deeper understanding of the functionalities of AI, and as it has advanced over the years, adapt objectives to better suit the outputs that AI has been able to deliver. Rather than persisting with the goal of building machines that can emulate human intelligence, these goals shifted to hybrid intelligence systems, new crowd intelligence systems and more complex intelligence systems [23]. However, regardless of the shift in goals, for decades, users and scientists alike have raised concerns that machine intelligence will one day surpass that of humans. However, differences between the types of intelligence must be noted. For example, the intelligence of a human is, to a degree, biological, as humans learn from their surroundings and are capable of complex emotional intelligence. It is thought that in the near future, advancements will be seen in the field of hybrid-augmented intelligence. We have already seen benefits of such technology, including machines helping in medical diagnostics and surgical procedures. Combining the intelligence of machines and humans is thought to be a huge area of potential future research and development. However, regardless of this potential, its relationship with humans has led to the uncovering of a more modern set of setbacks.

3 Emerging Ethical Challenges

Research in the area of AI has been mapping areas of ethical concern, from privacy violations [7], transparency failings, biases, to mistrust. According to Jobin et al. [11], there are five ethical principles: transparency, justice and fairness, non-maleficence, responsibility and privacy [11]. This section will discuss some of these issues, namely the existence and impacts of privacy, bias, and mistrust which is perhaps both a consequence of the former challenges, as well as a problem in and of itself.

3.1 Bias

The types of bias that may occur in the pipeline are extensive. Srinivasan and Chander [26] present that bias may be introduced anywhere in the AI pipeline, from the creation of the dataset, data analysis, and the evaluation and testing stages. These include sampling bias, framing effect bias, sample selection bias and sample treatment bias [26].

The Fig. 1 demonstrates the taxonomy of bias types and the relationships between them.

It is widely presented that bias is a challenge that is concerning both scientists and users of AI systems. According to Baeza-Yates [1], bias may be introduced at multiple stages throughout the process, i.e. from the algorithm itself, from the feedback loop between the user and the system, or introduction to the initial data [1]. According to McKinsey Global Institute [18], bias, or more specifically, unwanted bias, may be described as "systematic discrimination against certain individuals or groups of individuals based on the inappropriate use of certain traits or characteristics" [18, p. 2]. This may refer to attribute such as race, gender, sexual orientation or disability. But how can fairness be defined? And, importantly, how can it be codified? According

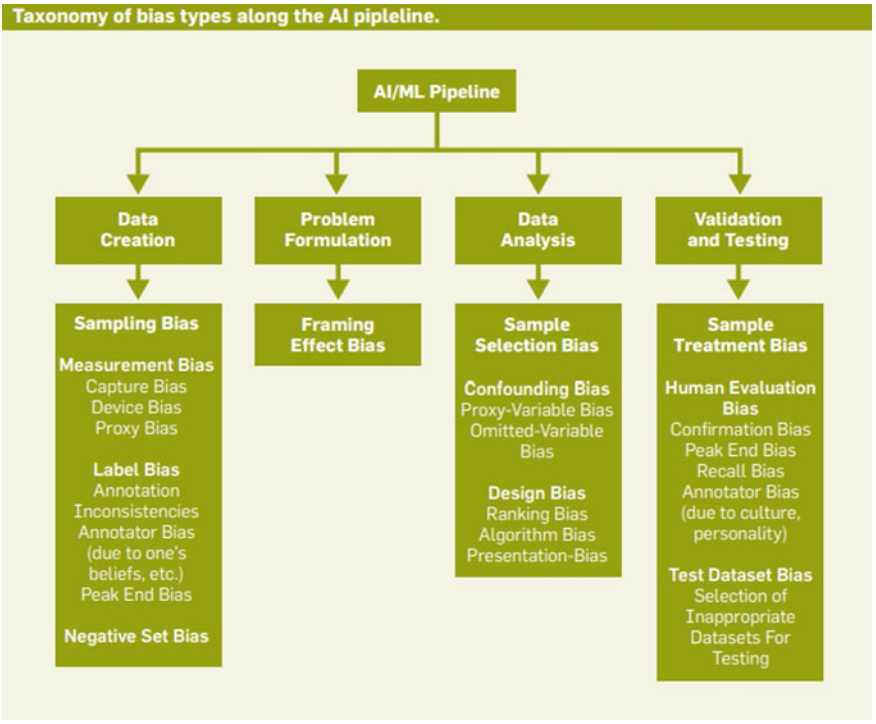

Fig. 1 Bias types within AI pipeline [26]

to a report by Crawford et al. [6], this is a critical consideration. The report refers to CEO image search as an example. Would fairness in this instance be to ensure that an equal percentage (50%) of the CEOs displayed by the search engine are women? Or would fairness represent the current, unequal reality?

Research into fairness metrics is extensive. Narayanan presents 21 distinctive definitions of fairness but described this as "non-exhaustive". Interestingly, some research has suggested that the balance between what is deemed to be fair and unfair is perhaps finer than one may expect. For example, research by Hu and Chen presents a "just" labor market model. However, the meaning of fairness could change over time, as groups may be subjected to further harm in the long term if they are, for example, approving a loan that the individual later cannot repay (Hu & Chen). As the meaning of fairness is highly contested, particularly depending on the individual circumstances, it is most likely that the definition would have to be crafted based on the use case, as patterns of potential unfairness may differ from case to case.

As presented in 'Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective', science always precedes societal wisdom [28]. While AI technology is well-established and founded in excessive research, the ethical aspects surrounding these technologies have only persisted for fewer than 20 years. AI systems are used as experts in a number of fields, and the ways in which they are being used often exceed human ability. Driverless cars, image recognition, transcription and text analysis are all examples of activities in which machines are all capable of outperforming humans. However, the very thing that makes it so useful, is perhaps a reason for strict controls to be enforced—its autonomy. But the meaning of autonomy is ambiguous. According to Noorman and Johnson [21],

> autonomy implies acting on one's own, controlling one's self, and being responsible for one's actions. Being responsible for one's actions in particular requires that the person had some kind of control over the outcome at issue. Thus, framing robots as becoming increasingly autonomous may suggest that robots will be in control and that human actors will, therefore, not be in control. [21]

This concept of 'lack of control' is interesting. Particularly as one of the challenges stems from the fact that AI is often not understood well before it is implemented, and therefore is not within our control as well as it perhaps should be. Without thorough understanding of the technologies, how can they be efficiently and ethically implemented? According to Bostrom and Yudkowski [3], one of the biggest risks taken is the assumption that AI can be implemented without in-depth understanding of the power of intelligent systems (p. 12). According to Zandi et al. [29], it is exactly this lack of understanding surrounding algorithms used to generate the outputs that leads to the conception of ethical questions regarding its governance, data ownership and privacy, among other factors. The paper raises common concerns related to the implementation and usefulness of AI. Concerns include, what contexts are the algorithms designed for? For example, how do we ensure that poorer nations can reap the same benefits as wealthier ones? What safeguards should be considered, particularly in fields such as healthcare? Are there any limitations of the datasets used to train the machines?

An interesting example of an area of AI raising ethical questions, is recommender systems. According to Milano et al. [19], recommender systems may be described as:

> functions that take information about a user's preferences (e.g. about movies) as an input, and output a prediction about the rating that a user would give of the items under evaluation (e.g., new movies available), and predict how they would rank a set of items individually or as a bundle. [19]

According to Burr et al. [5], recommender systems shape our individual experiences of using technology. For this reason, it is a valuable example of AI to scrutinise, given that the success of such systems depends on the use of user personal data. Example of ethical challenges of recommender systems include: the practices of user profiling, data publishing, algorithm design, user interface design, and online experimentation or A/B testing [5]. This could lead to privacy breaches, anonymity breaches, behaviour manipulation and bias in the recommendations given to the user, content censorship, exposure to side effects, and unequal treatment [24]. In this way, it could be said that recommender systems highlight the three challenges discussed in this paper- bias, privacy, and mistrust. One of the main concerns behind the use of these technologies is potential bias that could impact the accuracy of the technology. For example, Buolsmwini and Gebru [4] found that leading facial recognition technology performed much more accurately when identifying male white face, compared to women and people of colour.

3.2 Privacy

Privacy of personal information is protected by data security measures, ensuring that there are steps in place designed to prevent access to such data by those to whom it does not concern. Machine learning algorithms feed on big data, and therefore require large datasets in order to increase accuracy. This requirement for huge amounts of data, often from multiple sources, increases the probability of tracing data, which in turn may defeat privacy objectives. However, the size of the data aside, the availability of sensitive data, (for example, from social media, mobile phones, and credit card transactions), allows machine learning algorithms to learn and make inferences from data that may not have been disclosed. Looking closer at an example of this, a paper from Kosinski et al. [13] is composed of data from a survey undertaken by 58,466 volunteers from the United States. The information retrieved during the survey included the respondents' Facebook profile information, including their Likes and psychometric test scores. Using dimensionality reduction to pre-process the Likes dataset, before a linear regression model predicts individual profiles from the data, Kosinski et al. present that a multitude of personal attributes, such as sexual orientation to happiness, can be predicted at an accuracy of, in some cases, as high as 0.95 (when discriminating between Caucasian and African American volunteers) [13]. Additionally, the paper presents that Likes attract users who

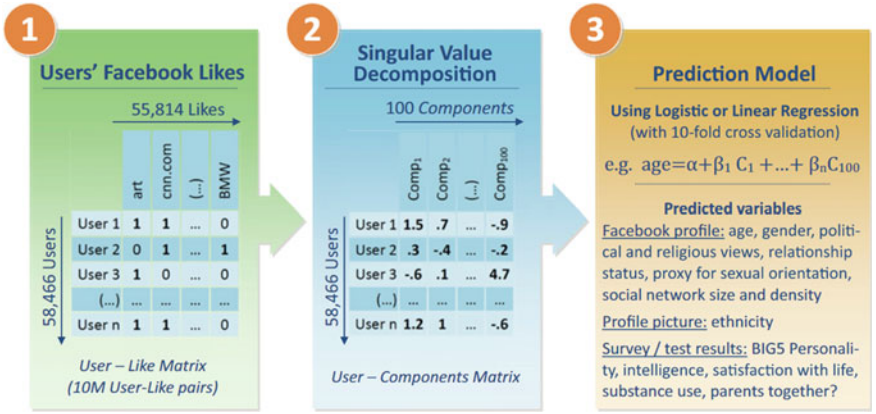

Fig. 2 The user-like matrix, the singular value decomposition, and the prediction model and variables [13]

possess a different average personality profile and are more likely to belong to a specific demographic, which is, in turn, used to predict user details. It also suggests similarity between Facebook Likes and other types of data such as search histories and online purchases, which only adds to the possibility of revealing further personal attributes.

On one hand, the ability to accurately predict demographic data can prove useful for marketing and commercial purposes [13]. However, this 'power' held by commercial companies may bear substantial negative consequences for users, as private information is collected and used without consent. And in an online world where internet users increasingly leave digital footprints, sometimes even unknowingly, it can be difficult to keep track of the personal information being revealed [13].

In an attempt to protect user privacy, data protection regulations are becoming increasingly strict. For example, established in May 2018, the General Data Protection Regulation (GDPR) is perhaps one of the most restrictive privacy laws and requires EU companies and partners to have a duty of truth—that is they must be transparent regarding how their data will be used, and may also only collect necessary data. With transparency at its core, the regulation enforces that data may only be kept for as long as necessary. Additionally, the individual has the right to ask how their data is being used, and for it to be deleted [9].

However, according to Kosinski et al. [13], this may only be "superficially reassuring", as it makes the promise of de-identification realistic, despite this not necessarily being the case. There are many interesting examples of how privacy violations may occur. An interesting example is Amazon Alexa, which revolutionised the home—allowing users to control elements of their surroundings using voice commands. Given it, and similar products, are increasingly commonplace in the average home, it also means that companies such as Amazon also store growing amounts of, often personal, data. The Alexa Voice Service uses a series of APIs to communicate between the product and the cloud. According to a study by Krueger and McKeown

[14], as these devices are often operating, this "makes this data a valuable source of evidence for investigators performing digital forensics" [14]. Interestingly, the study presents that although there are multiple options for users to delete data, it was concluded that most to note remove 100% of user data. Another interesting example is data captured on human emotions online, e.g., using the TimeDepth camera system, which is made up of components such as proximity sensor, infrared camera, speaker and microphone. According to Hodl and Gremsl [10], this combination of features allows the camera system to capture over 50 facial muscle movements, which goes beyond what one would generally consider to be personal data. In this way, it is thought that facial recognition technologies will continue to cause privacy concerns as it continues to advance.

3.3 Mistrust

Both bias and privacy concerns are challenges that may lead to mistrust among users, which may also be considered a challenge in its own right. For example, bias and commercial involvement impact public trust [17]. According to Kerasidou [12], trust can be broken down into the following three concepts: vulnerability, voluntariness, and good will. The type of trust is dependent on to whom it is directed. For example, it may be personal, or institutional. Interestingly, Kerasidou's paper considers that trust is detrimental to the standing of research, and a lack of it poses great threat. However, it also presents that this threat may be false, given that, depending on the field (for example, medicine), actions of the public may not align with complete mistrust.

According to a public survey conducted by YouGov, for the British Computer Society, 53% of the 2000 respondents had "no faith in any organisation to use algorithms when making judgements about them" [2]. The survey followed the high-profile UK exams incident which saw an algorithm implemented to predict student grades be removed in favour of predictions made by teachers. Perhaps influenced by this event, the survey highlights that a mere 7% of respondents trusted algorithms implemented in the education sector [2]. In addition to the education sector, the level of mistrust of the algorithms employed by social media companies is similarly high. According to the survey, levels of trust in algorithms to "serve content and direct user experience was similar at 8%" [2]. Interestingly, trust in AI decision-making tools employed by the big tech companies, including Apple and Google, was higher at 11% [2]. In 2019, 'Ethics Guidelines for Trustworthy AI' was published by the European Commission's High-Level Expert Group, a publication that trust is crucial in order to fully reap the benefits of AI technologies. However, it also calls for the development of trustworthy AI, which would focus on two core elements:

> (1) respect for fundamental rights, applicable regulations and core principles and values, ensuring an 'ethical purpose' and (2) technical robustness and reliability, to avoid unintentional harm caused by lack of technological mastery (EU Commission 2019b: I). Trustworthy AI is, according to the Guidelines, ethical, lawful, and robust AI. [27]

According to Sutrop [27], the guidelines are vague when it comes to defining trust, despite this being an essential consideration when seeking to build and maintain it. As a baseline, the guidelines suggest that trustworthiness is equal to giving trust, however, that is not the case, as trusting that technology is worth of our trust does not mean that it indeed is. Perhaps interestingly, is the distinction made in the paper between 'trust' and 'reliance'. In literature, the term 'trust' is generally used to refer to a relationship that exists between people, whereas 'reliance' is used to refer to the relationship we have with an inanimate object. The difference being that a person you trust has the power to betray you, but an inanimate object, such as a television, can, at most, only cause disappointment [12, 27]. In research on the subject of AI ethics, AI is often discussed in terms of its trustworthiness, or lack thereof. With this frame of thinking in mind, it is interesting to consider whether, as users of these technologies, when we refer to its trustworthiness, we are considering the capabilities of the solution itself, or the creators behind such technology.

4 Solutions and Considerations for the Future

As discussed, ethical issues, such as privacy concerns and algorithmic bias may cause mistrust and taint the public perception of AI solutions. Mistrust then may open up criticism and anxiety, which then feeds deepening levels of mistrust. In order to avoid such issues, it is critical to consider how ethical challenges may be mitigated.

Figure 3 demonstrates the structure of a human-centric approach to AI, which mitigates common risks such as black-box models, privacy violations and bias and discrimination, by introducing the following requirements: algorithmic transparency and fairness, understandable explanations, privacy-preserving algorithms and data cooperatives [15]. With this in mind, it is important to consider each of these requirements. Srinivasan and Chander [26] argue that in order to tackle bias, it is vital that the structural dependencies of dataset features is fully understood to effectively pinpoint the root of potential bias. Determining which aspects of the data may be deemed sensitive depending on the given application. Datasets must be representative and great care should be taken to ensure that data fully reflects the current landscape, and consistency when annotating data is crucial [26]. Further consideration should be made when conducting A/B testing, for example, treatment bias should be avoided by ensuring that the test conditions are not restricted to a certain demographic, for example.

Additional solutions for bias include the following: user-centered solutions e.g., adjustable tools for users so they can control how their personal data is used, including the ability to opt out of online experiments.

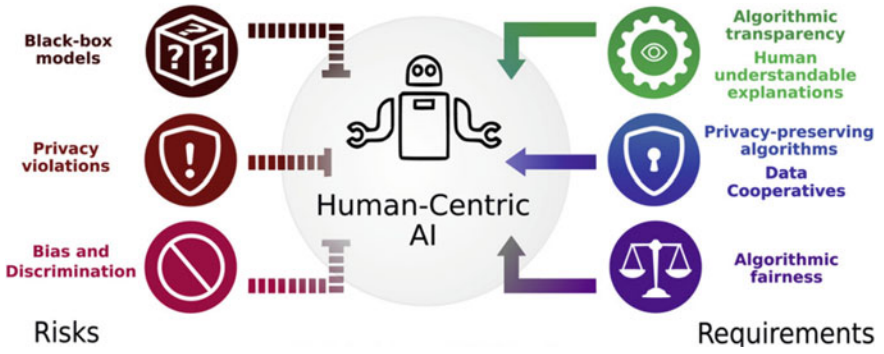

Fig. 3 Human centric AI [15]

When working towards the final goal of privacy, there are many angles from which the problem may be addressed. According to Liu et al. [16], one method may be algorithmic perturbation in deep learning, to ultimately train a model while protecting individuals' data by adding noise. Research in this area presents that this noise may be added to the initial data, parameters, or the output, although is mostly focused on adding to the gradients. The main challenge with adding noise at this stage is that the number of iterations and the amount of noise that can be injected go hand in hand, which often means over-injection of noise. At both the input and output stages, there are also challenges with this method, which may tie into big data protection concerns unrelated to deep learning. Interestingly, perhaps the most successful ways to add noise to protect data is to add to the 'hidden layer of an autoencoder'. Interestingly, this concept of using machine learning for privacy protection is gaining traction, such as using GNN to create synthetic datasets. In this way, it is thought that further research on these techniques, among others, may lead to advancements when seeking to solve the challenge of AI privacy concerns.

Additionally, it could be argued that another way to challenge bias, privacy, and mistrust, is by expanding transparency, and when discussing transparency in AI systems, it is difficult to avoid mentioning explainable AI (XAI). According to Ehsan et al. [8], one of the main challenges with XAI is that explanations are usually not fit for the intended audience, but rather the tech-minded creators, rendering it ineffective. The paper delves into Social Transparency (ST), and presents four design categories in an attempt to rectify blind spots in current XAI architecture, namely: what, why, who, when. However, the paper also highlights that further work is required to unpack how these insights may be transferred, although acknowledges that viewing XAI through the lens of ST is vital in order to progress and challenge how it is designed and reflect on how it can be improved to benefit users.

5 Conclusion

Section 1 presents a short evolution of AI, from its conception to the current landscape. It then presents the challenges to the development of AI including examples of how the goals of implementing machine learning have had to change over time, such as focusing efforts on software instead of hardware, and understanding that machines learn from their environment rather than manual information transferal. Following on from this, Sect. 2 delves into some of the emerging ethical challenges that have stemmed from the common use of AI technologies, namely bias, privacy, and mistrust as a result. Distinct types of bias are discussed, such as sampling and framing bias, as well as how bias may be measured. Examples were then provided as to how privacy infringements may occur, such as by accessing, sometimes unknowingly, personal or identifiable data without consent, such as via Facebook Like data, as well as audio data from Amazon Alexa, for example, which may be used as digital evidence. The topic of mistrust is then handled, as an issue in its own right, perhaps stemming from the existence of the above. The European Commission's 'Ethics Guidelines for Trustworthy AI' was discussed, as well as public survey information, revealing that, in some sectors, trust in organisations using algorithms for decision-making purposes was as low as 7%.

Despite the increasingly challenging nature of the problems faced by those seeking to ethically implement AI solutions, there are methods that should be the subject of further research in order to meet the common goal of ethical AI. With this in mind, Sect. 3 considers how privacy violations and bias can be mitigated through algorithmic transparency and fairness, human understandable explanations, and privacy-preserving algorithms. Above all, the glaring issues of AI adoption are those affecting humans, thus implementing more human-centric approaches to combat these challenges is critical. In order to fully tackle emerging and continuous ethical issues, further research should be conducted into the areas described above, such as tethering XAI and SI together to ensure that human issues are solved in the appropriate way, that is, by not only focusing on the technological impact but by understanding the consequences on users and building solutions with them in mind.

In addition, increased levels of transparency surrounding how technologies operate, increase education among users and may be empowered to use AI solutions. Perhaps, as discussed in relation to EU guidelines, it is perhaps not that users should blindly trust AI, but rather the aim should be that users feel empowered to trust AI systems. This feeling of empowerment may only be satisfied through increased levels of user education, not only on the current AI landscape but its benefits along with challenges such as those raised in this chapter. Education occurs in many settings, and may be carried out more formally, or by simply making technology more accessible, through increased levels of transparency and fairness, and ensuring it can be understood by those who use it and may be impacted by potential adversity. Although ethical challenges cannot be solved overnight, this chapter merely seeks to bring more context to the subject by highlighting interesting points of future research.

References

1. Baeza-Yates R (2022) Ethical challenges in AI. In: International conference on web search and data mining, Virtual Event, AZ, USA, Association for Computing Machinery, New York, NY, USA, pp 1–2, Feb 2022
2. BCS (2021) The public don't trust computer algorithms to make decisions about them, survey finds, BCS. Retrieved from https://www.bcs.org/articles-opinion-and-research/the-public-dont-trust-computer-algorithms-to-make-decisions-about-them-survey-finds/. Accessed on 9 Dec 2022
3. Bostrom N, Yudkowski E (2018) The ethics of artificial intelligence
4. Buolamwini J, Gebru T (2018) Proceedings of the 1st conference on fairness, accountability and transparency. PMLR 81:77–91
5. Burr C, Cristianini N, Ladyman J (2018) An analysis of the interaction between intelligent software agents and human users. Mind Mach 28:735–774. https://doi.org/10.1007/s11023-018-9479-0
6. Crawford K, Dobbe R, Dryer T, Fried G, Green B, Kaziunas E, Kak A, Mathur V, McElroy E, Sánchez AN, Raji D, Rankin JL, Richardson R, Schultz J, West SM, Whittaker M (2019) AI Now 2019 report. AI Now Institute, New York. https://ainowinstitute.org/AI_Now_2019_Report.html
7. de Montjoye Y, Hidalgo C, Verleysen M, Blondel V (2013) Unique in the crowd: the privacy bounds of human mobility. Sci Rep 3(1376):1–5
8. Ehsan U et al (2021) Expanding explain ability: towards social transparency in AI systems. In: Proceedings of the 2021 CHI conference on human factors in computing systems [Preprint]. https://doi.org/10.1145/3411764.3445188
9. Goddard M (2017) The EU general data protection regulation (GDPR): European regulation that has a global impact. Int J Mark Res 59(6):703–705. https://doi.org/10.2501/ijmr-2017-050
10. Hodl E, Gremsl T (2022) Introduction to the special issue: questioning modern surveillance technologies: ethical and legal challenges of emerging information and communication technologies. Inf Polity 27(2):121–129. https://doi.org/10.3233/ip-229006
11. Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1(1):389–399
12. Kerasidou A (2017) Trust me, I'm a researcher!: the role of trust in biomedical research. Med Health Care and Philos 20:43–50. https://doi.org/10.1007/s11019-016-9721-6
13. Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behaviour. Proc Natl Acad Sci 110(15):5802–5805. https://doi.org/10.1073/pnas.1218772110
14. Krueger C, McKeown S (2020) Using Amazon Alexa APIs as a source of digital evidence. In: International conference on cyber security and protection of digital services (Cyber Security) 2020:1–8. https://doi.org/10.1109/CyberSecurity49315.2020.9138849
15. Lepri B, Oliver N, Pentland A (2021) Ethical machines: the human-centric use of artificial intelligence. iScience 24(3):102249. https://doi.org/10.1016/j.isci.2021.102249
16. Liu B et al (2022) When machine learning meets privacy. ACM Comput Surv 54(2):1–36. https://doi.org/10.1145/3436755
17. McKay F, Williams BJ, Prestwich G, Bansal D, Hallowell N, Treanor D (2022) The ethical challenges of artificial intelligence-driven digital pathology. J Pathol Clin Res 8:209–216. https://doi.org/10.1002/cjp2.263
18. McKinsey Global Institute (2019) Notes from the AI frontier: tackling bias in AI (and in humans). Retrieved from https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/MGI-Tackling-bias-in-AI-June-2019.pdf. Accessed on 9 Dec 2022
19. Milano S, Taddeo M, Floridi L (2020) Recommender systems and their ethical challenges. AI Soc 35:957–967. https://doi.org/10.1007/s00146-020-00950-y
20. Narayanan A 21 Definitions of fairness and their politics. In: Tutorial presented at the first conference on fairness, accountability, and transparency (FAT∗)

21. Noorman M, Johnson DG (2014) Negotiating autonomy and responsibility in military robots. Ethics Inf Technol 16(1):51–62. https://doi.org/10.1007/s10676-013-9335-0
22. Noyes K (2016) 5 things you need to know about A.I.: cognitive, neural and deep, oh my! Computerworld, 10 Nov. Retrieved at www.computerworld.com/article/3040563/enterprise-applications/5-things-you-needtoknow-about-ai-cognitive-neural-anddeep-oh-my.html
23. Pan Y (2016) Heading toward artificial intelligence 2.0. Engineering 2(4):409–413
24. Paraschakis D (2017) Towards an ethical recommendation framework. In: 2017 11th international conference on research challenges in information science (RCIS), pp 211–220. https://doi.org/10.1109/RCIS.2017.7956539
25. Russell S, Norvig P (1995) Artificial intelligence a modern approach. Simon & Schuster, New Jersey
26. Srinivasan R, Chander A (2021) Biases in AI systems. Queue 19(2):45–64. https://doi.org/10.1145/3466132.3466134
27. Sutrop M (2019) Should we trust artificial intelligence? J HumIties Soc Sci 23(4):499. https://doi.org/10.3176/tr.2019.4.07
28. Tzimas T (2021) Legal and ethical challenges of artificial intelligence from an international law perspective. Springer, Thessaloniki
29. Zandi D et al (2019) New ethical challenges of digital technologies, machine learning and artificial intelligence in public health: a call for papers. Bull World Health Organ. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/279412/PMC6307511.pdf. Accessed on 9 Dec 2022

A Balance of Power: Exploring the Opportunities and Challenges of AI for a Nation

Shasha Yu and Fiona Carroll

Abstract Artificial intelligence is having a profound impact on the development of human society. It is improving—in some case, re-inventing—our economic, political, cultural, educational and medical sectors, to name a few. For many it is a cost effective solution that makes processes more effective, more intelligent and often more independent. However, in doing this, it is also having a deterministic influence on the dynamics of our societies. From an economic perspective, AI technology could be a game-changer, giving emerging markets the opportunity to outpace more developed markets. In fact, it has the ability to change the balance of global power, so much so that many countries are now striving for a national strategy on AI. And this goes to the very heart of a nation's security where AI can also create significant implications for the protection and defence of it's citizens, and economy. This chapter presents how Artificial Intelligence technology is extremely important in how it can shape the strength and power of a nation. Moreover, it highlights how AI can both positively and negatively impact a nation's security. In summary, the chapter will provide a detailed overview of AI, it will analyse the direct and indirect effects of AI on national security and will present some potential solutions.

Keywords Artificial intelligence · National security · AI · Machine learning · Big data

S. Yu (✉)
School of Professional Studies, Clark University, Worcester, MA, USA
e-mail: ShaYu@clarku.edu

F. Carroll
Cardiff School of Technologies, Cardiff Met University, Cardiff, Wales
e-mail: fcarroll@cardiffmet.ac.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_2

1 Introduction

Research into artificial intelligence (AI) dates back to World War II. Alan Turing, a British mathematician, was most probably the first to decide that artificial intelligence was best studied through computer programming rather than by building machines [47]. In fact, Alan Turing's 1950 paper in *Computing Machinery and Intelligence* argued that if a machine can successfully pretend to be human in front of a knowledgeable observer, then it should be considered intelligence [69]. The term 'Artificial Intelligence' was first used in 1955 by John McCarthy et al. in the *Dartmouth Summer Research Project on Artificial Intelligence*, establishing Artificial Intelligence as a research discipline [48]. According to John McCarthy, AI includes, but is not limited to, the following branches: Logical AI, Search, Pattern recognition, Representation, Inference, Common sense knowledge and reasoning, Learning from experience, Planning, Epistemology, Ontology, Heuristics, and Genetic programming [47].

Today, artificial intelligence has been widely used in all walks of life and has had a profound impact on the development of human society. It has brought about fundamental changes in the development of countries in the economic, political, cultural, educational and medical spheres. According to PwC research, artificial intelligence could transform the productivity and GDP potential of the global economy [35]. By 2030, AI could contribute up to US$15.7 trillion to the global economy, more than the current output of China and India combined [35]. From a macroeconomic perspective, AI technology could be a game-changer, giving emerging markets the opportunity to outpace more developed markets.

The authors of this chapter highlight that Artificial Intelligence technology is extremely important in how it can shape the strength and power of a nation. In doing so, it can positively and negatively impact its national security. The following sections will give an overview on AI, they will analyse the direct and indirect effects of AI on national security whilst also, they will present some possible solutions.

2 Artificial Intelligence Overview

According to John McCarthy, artificial intelligence "is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable" [47, p. 2]. *UNESCO's World Commission on the Ethics of Scientific Knowledge and Technology (COMEST)* describes artificial intelligence as "involving machines capable of imitating certain functionalities of human intelligence, including such features as perception, learning, reasoning, problem solving, language interaction, and even producing creative work" [16, p. 6]. Furthermore, Stuart Russell and Peter Norvig

summarise previous definitions of AI and distil four possible goals to pursue in artificial intelligence: Acting humanly, Thinking humanly, Thinking rationally, and Acting rationally [62].

In 1997, IBM's supercomputer *Deep Blue* caused an uproar when it defeated one of the greatest chess masters of the time, Garry Kasparov. It became the first computer system to beat the world chess champion within the standard tournament time limit. Moreover, in 2015, a robot named *Eugene Goostman* won the Turing Challenge for the first time. In this challenge, human raters chatted with unknown entities using text input and then guessed whether they were talking to a human or a machine. *Eugene Goostman* tricked more than half of the human evaluators into thinking they had been talking to a human. Nowadays, the main application areas of AI include Natural Language Processing (NLP), Speech Recognition, Image Recognition and Processing, Autonomous Agents, Affect Detection, Data Mining for Prediction, and Artificial Creativity [70].

3 Artificial Intelligence in National Security

3.1 The Impact of Artificial Intelligence on Economy

Thanks to the Internet and computer technology, we are living in an era of rapid growth in the amount of data available. Analysis shows that Internet users have more than doubled in the last decade, climbing from 2.18 billion at the beginning of 2012 to 4.95 billion by early 2022 [42]. In addition, the 'typical' global Internet user now spends nearly seven hours a day using the Internet (consuming information and generating new information) on all devices [42]. IBM calculates that more than 2.5 quintillion bytes of data is generated every day [70]. The Internet has long surpassed radio, TV and newspapers as the largest source of information for people, and is also the main place for major companies to compete for potential users and customers.

Back in the 1970s, Nobel Prize winning economist Herbert Simon introduced the concept of the 'attention economy'. He said, "In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention" [64, p. 37]. Goldhaber helped popularize the term 'attention economy', arguing that "the currency of the new economy is not money, but attention" [30, p. 1].

In today's era of big data, huge amounts of data is being produced every second and the rate of data production is still growing exponentially. Moreover, individuals have a limited attention span, which means that when they pay attention to something, they do not pay attention to other things. In contrast to the decreasing cost of information brought about by technological progress, attention is even more scarce under the impact of information overload. The scarcity of attention makes it highly valuable and

can lead to additional resources, such as money, fame, and power. Indeed, influencers can get a lot of money through live marketing, even the average everyday person can gain a large fan following by posting novel videos. Even politicians can gain public attention and support by going around giving speeches. And behind all of this, there is an increasing number of artificial intelligence algorithms.

In the fierce war for users' attention, major companies conduct *Psychographic analysis* of users. *Psychographic analysis* aims to analyse the subjective characteristics of users in order to understand or predict their behaviour. It analyses areas such as personality traits, activities, interests, opinions, needs, goals, values and attitudes [79]. It is usually information obtained by analysing a range of behavioural data. For example, researchers can uncover users' opinions, emotions, and attitudes based on the social relations between users on social networks and the interactions of user sentiment orientation [10]. AI-based *psychographic analysis* can accurately predict a user's preferences and likely future behavior based on previous behavior or the behavior of other users belonging to the same group. Furthermore it can apply targeted solicitations such as pushing ads, news, or even recommending people that they are likely to connect with.

Most AI relies on learning knowledge from past data, and the larger the data sample, the better it performs, so data is considered the oil of the new age. Technology giants and business giants have the advantage of strong capital and technology to access more data resources and create more advanced AI products. This can significantly reduce costs and improve service quality, which in turn helps them to have greater access to more data resources. In contrast, the majority of small to medium-sized enterprises are at a disadvantage in terms of access to data resources. There are only an exceptional few small to medium-sized enterprises that are able to excel in their niche areas, resulting in a polarised situation. Indeed, it is only a few companies that dominate data resources thus become monopolies and undermine the healthy ecology of the national economy.

As the gatekeeper of the Internet monopoly, Google, for example, has billions of users and countless advertisers around the world. With its vast amount of data and powerful AI algorithms, it has become one of the world's richest companies with a market capitalization of $1 trillion [19]. Its influence even transcends national borders to become a global monopoly [19]. However, in 2017, Google was convicted and fined €2.42 billion by the EU for its monopolistic practices in advertising, the largest such antitrust fine issued by the European Commission [56]. Google accounts for almost 90% of all search queries in the U.S. and uses anti-competitive tactics to maintain and expand its monopoly on search and search advertising. In 2020, the U.S. Department of Justice sued monopolist Google for violating antitrust laws [32]. Indeed, Google has used its dominant position to engage in a series of anti-competitive practices that harm competition and consumers. It also has reduced the ability of innovative new companies to grow, compete, and discipline Google's behavior.

In fact, the impact of AI algorithms (driven by big companies) on the economy is much broader than that. When large companies gain a monopoly on access to user data, they can in fact control their users' lives in many ways. For example, short-form video platforms can decide what information to push to users; social media

platforms recommend from time to time what influencers they might be interested in promoting; and commercial websites push users merchandise, movies, and books that might appeal to them. This influence is so subtle that when a user who is influenced by it makes a decision, he/she may not even realize that he/she is influenced by it. For example, they may have chosen a product from brand A over brand B when shopping because they saw more push messages about brand A on the website, and that was the result of an AI algorithm. That is, when AI algorithms can influence a person's attention, they can influence how people spend their money. In this sense, business giants with AI algorithms at their disposal can more-or-less control the flow of money.

In the manufacturing industry, the adoption of AI can replace a portion of the work that would otherwise be performed by humans. It can limit the number of hours worked, significantly reduce the cost of products and make them more competitive in the marketplace. As a result, countries and companies that invest in AI earlier will benefit more from this, gaining accelerated economic growth. Less developed countries and regions, on the other hand, are likely to be left far behind. According to PwC research, the rate of adoption of AI technologies significantly impacts economic development potential, showing an imbalance across regions [57]. By 2030, the largest economic gains from AI are expected to occur in China (26% GDP growth) and North America (14.5% growth), totalling US$10.7 trillion and accounting for nearly 70% of the global economic impact [57]. They are followed by Southern Europe, developed Asia, Northern Europe and Latin America with 11.5%, 10.4%, 9.9% and 5.4% respectively, while for the rest of the world, the total impact is 5.6% [57].

3.2 The Impact of Artificial Intelligence on Employment

According to UN DESA, the global population is expected to reach 9.8 billion by 2050, of which more than 6 billion will be of working age [71]. At that time, people of working age may face significant employment pressures. At the same time, with the rapid development of artificial intelligence and automation, intelligent devices are replacing human work in various fields. The increased utilization of technology has reduced the number of jobs, and more and more jobs are being replaced by automated machines and software. A study by PwC shows that by the mid-2030s, 30% of jobs and 44% of workers with low levels of education will be at risk of automation as AI advances and becomes more autonomous [35]. In a similar research, the *World Economic Forum* estimates that, as a result of artificial intelligence and automation, 85 million jobs will be replaced by 2025, while 97 million new jobs will be created in 26 countries [78]. However, the report also sets out a list of jobs that are growing and decreasing in demand, with the growth being concentrated in highly skilled jobs such as artificial intelligence, data-related jobs, information security and the Internet of Things, while the decrease in demand is mainly for manual labour or simple skilled jobs [78].

By that time, more new jobs are being demanded by technology updates, placing higher skill requirements on the workforce. Workers who used to work in labor-intensive industries will have to face structural unemployment, and they will be forced to update their knowledge in order to adapt to the new demands. For those less developed countries and regions with low levels of education, they will be at an even greater disadvantage. The uneven economic development brought about by the development of AI technology will also trigger an uneven distribution of talent by association. Higher investments in AI in developed economies will yield higher returns, thus attracting more highly educated and skilled young people from less developed regions. Due to the siphon effect of talent aggregation [46], the talent resource gap between developed economies and less developed regions will be exacerbated. Especially, when more professionals flow to developed economies that are more suitable for their individual development.

In addition, with the rapid development of AI technology in recent years, computers have not only replaced human workers in many repetitive labor fields, but have even achieved good performance in creative fields. These include areas such as music, literature and painting, and in highly technically demanding fields such as medicine and architecture. For example, previous studies have shown that AI is more accurate than many doctors in diagnosing breast cancer from mammograms [75]. As a result, even the well-educated senior personnel are under pressure to update their knowledge and skills. According to the *World Economic Forum*, by 2025, half of all workers will need to upskill or reskill to prepare for job changes and new jobs [78].

3.3 The Impact of Artificial Intelligence on Education and Culture

In recent years, especially since COVID-19, artificial intelligence has been used in many applications in the education industry. On the one hand, it can provide personalized tutoring and 24/7 accessibility to students. This can help students from different backgrounds to get equal (and equitable) access to education. On the other hand, it helps educators automate tasks such as administration, assessment, grading, and repetitive question-answering, so they can focus on more innovative work. The rapid development of artificial intelligence technology also gives students more hands-on opportunities to implement their ideas. Some functions that previously required complex code can now be easily implemented with codeless AI [65]. For example, Microsoft's *lobe.ai* allows anyone to train computer image classification models to recognize objects and is developing a codeless object detection and data classification platform for the general public [51]. Another codeless training platform, *Teachable Machine*, can be used to recognize user-defined images, sounds, and poses [66].

As a result of natural language processing (NLP) technology, artificial intelligence is also increasingly being used in a variety of writing tasks to help students improve their writing. For example, students can use platforms such as *Grammarly* for grammar checking, *Wordtune* for sentence touch-ups, *Quillbot* for proofreading, plagiarism checking and citation, and even software such as *Rytr* to generate text on a variety of topics. This software learns from a large library of texts and can mimic the generation of texts that are almost indistinguishable from natural human language, even with the option of different languages and styles. Moreover, in the field of computer programming education, artificial intelligence is playing an amazing role. AI-powered code generators, such as *OpenAI Codex*, *DeepMind AlphaCode* and *Amzon CodeWisperer*, can convert natural language representations of tasks into computer-runnable code [5]. Most of them are trained on the GitHub codebase and can generate code based on various major computer programming languages and can convert them to each other. These code generators can help beginners understand various approaches to problem solving and develop their thinking.

In the field of art creation, artificial intelligence also has had an amazing performance. In September 2022, an artwork generated with the AI drawing tool *Midjourney* won the top prize in the digital category at the Colorado State Fair Art Competition in the United States. These AI-generated applications gain popularity at a far lower cost and quicker response than human artists [59]. OpenAI's image generator *DALL-E 2*, released in spring 2022, has more than 1.5 million users and creates more than two million images per day. *Midjourney*, another popular AI image generator released the same year, has more than three million users on its official *Discord* server [61]. Applications such as *DALL-E 2* and *Midjourney* are built by crawling millions of images from the open web, then teaching algorithms to recognize patterns and relationships in those images and generate new images in the same style. This means that artists who upload their work to the Internet may unwittingly help train their algorithmic competitors. *DALL-E 2*, *Midjourney*, and *Stable Diffusion* - enable amateurs to create complex, abstract, or realistic artwork [60]. These AI-created artworks are generated by *Adversarial Generative Network* (GAN). GAN can be seen as such a system. By adding a discriminatory model to the generative model, GAN mimics the mechanism by which humans judge pictures in the real world. Thus, transforms the hard-to-define sample differences into a game problem. Similar to this is *AlphaZero*, which accumulates a large amount of data in the form of self-play and then explores a more optimal strategy from it. In this new research paradigm, the model changes from a tool for analysis to a 'factory' of data.

From a higher perspective, the success of GAN essentially reflects the fact that AI research has entered deeper waters. The focus of research has shifted from perceptual problems such as vision and hearing to solving cognitive problems such as decision making and generation. Compared with machine perception problems, these new problems are often not well solved by humans either, and the solution to such problems must rely on new research methods. Generative AI not only analyzes existing data, but creates new text, images, videos, code snippets and more. On the one hand,

these AI tools help people learn to master various skills better and expand the range of their abilities. On the other hand, they bring new challenges and even potential risks to education, such as ethical issues, bias and bad habits, and over-reliance [5].

Firstly, these AI tools are based on learning from publicly available databases of data derived from the intellectual output of others. As a result, those who use the work generated by these AI tools inevitably face issues of academic integrity. Secondly, they have the potential to make students overly dependent on them [11], or to develop such overconfidence that they neglect their own individual learning and training. And as a result, struggle to obtain the expected performance once they leave these tools. Thirdly, even seemingly correct computer generated code can hide undetectable errors [45] that can be risky and even costly if adopted without full understanding by the user [55]. Fourthly, as AI algorithms learn from data, algorithmic bias can occur when data sampling bias is in the source data and this can result in the under representation of a portion of the population. These biased models may generate codes that impact gender, race, class, and other stereotypes [11].

3.4 The Impact of Artificial Intelligence on the Security of Our Society

In recent years, artificial intelligence and big data technologies have been increasingly used for public security and in police departments. For example, many countries have introduced AI-powered police assistance systems that work well in various areas such as crime prediction, police dispatch, scene investigation, and case solving [84]. Even some cases that have not been solved for years have now been solved with the help of AI technology to unearth and discover new clues. For example, in May 2022, Dutch police have received dozens of leads after using *Deepfake* technology to bring a teenager virtually back to life nearly 20 years after he was murdered [2].

Not only the police, but in fact ordinary people can benefit from AI technology in the judicial process. For example, facial recognition technology is often used by police to identify suspects and witnesses. It can also be used by public defenders to find witnesses to prove a defendant's innocence [37]. However, artificial intelligence technology can be like a double-edged sword, whilst protecting society's security, it's rapid development also continues to generate new challenges to public security. *DeepFake* is a good example here. There artificial intelligence-generated video clips can be used with a variety of techniques to create worlds in which the reality has never happened. In 2021, *DeepFake* creators uploaded a fake Tom Cruise deepfake video on *TikTok* that drew two and a half million views, and even the commercial tools used to identify deepfakes cleared the clips as "authentic" [36].

The artificial intelligence company *DeepTrace* discovered 15,000 deepfake videos online in September 2019, nearly doubling in nine months. A staggering 96% were pornographic content, 99% of which were mapped faces from female celebrities to porn stars [33]. Deep forgery can mimic biometric data and potentially spoof

systems that rely on facial, voice, vein, or gait recognition. The more insidious effect of deepfakes and other synthetic media and fake news is the creation of a zero-trust society where people can't or don't bother to distinguish between truth and falsehood. And when trust is eroded, it becomes easier to question particular events. Deepfake videos about business moguls or politicians can trigger stock price fluctuations and even political events.

Another area of concern is the privacy risk posed by the large datasets used to train AI, which crawl the internet for a variety of available data. Especially, as the data subjects may not even know their data has been proliferated across sites and used to train AI tools. The *LAION-5B* dataset, for example, has more than five billion images, which include artwork by living artists, photographs of photographers, medical images, and photos of people who may not believe that their images will suddenly become the basis for AI training. Worryingly, people cannot opt out of being included in these datasets. Even more, these datasets include photo-manipulated celebrity porn, hacked and stolen non-consensual porn, and graphic images of ISIS beheadings [83].

Artificial intelligence-enabled "nudity" technology makes it easy for people without any expertise to create images of people appearing to be naked. In fact, image-based abuse that creates or alters a person's image without their consent disproportionately affects women. The damaging experience for victims of these realistic images can be devastating, affecting their personal and professional lives, as well as their physical and mental health [15]. With the addition of facial recognition, biometrics, genomic data and artificial intelligence predictive analytics, the uncontrolled proliferation of data puts people's privacy at risk. For example, facial recognition company *Clearview AI* collected 20 billion faces from social media sites like *Facebook*, *LinkedIn* and *Instagram*, as well as other parts of the web, to build an application designed to mine every public photo of people online. These images, while publicly available on the Web, are collected without people's consent. The tool mines photos that people don't post of themselves and may not even realize are online. *Clearview AI* has been the target of multiple lawsuits, and its database has been declared illegal in Canada, Australia, the United Kingdom, France, Italy and Greece. It also faces millions of dollars in fines in Europe [37].

More and more research is now using artificial intelligence techniques to make analyses and predictions about data. However, this process can easily be exploited by the unsuspecting for their nefarious purposes. By injecting specially crafted adversarial data into a target training dataset, an attacker may achieve the goal of manipulating machine learning results [76]. For example, by adding perturbed data to the training dataset, researchers altered the results of selfdriving cars' recognition of traffic signs [20]. This is known as *data poisoning*, and this poisoned data is often difficult to detect, potentially leaving a back door for unscrupulous individuals to manipulate AI models if adopted [12]. Therefore, the results of AI predictions based on the use of poisoned data in research may be distorted, misleading, and even inaccurate.

3.5 *The Impact of Artificial Intelligence on Science and Technology*

With sophisticated sensing devices, vast amounts of data, powerful computing capabilities, and 24/7 working hours, AI has greatly expanded the limits of human capabilities. Indeed, it can perform specialized tasks better than many experienced human experts. Taking medicine as an example, studies have documented that AI systems can correctly classify suspicious skin lesions better than dermatologists [21]. Furthermore, the algorithms of AI can be continuously improved and applied to related fields on a large scale and over a short period of time. As a result, once the breakthrough is achieved, the results can often be seen as spectacular. We have seen similar examples many times in fields such as biology and medicine.

Over the past few decades, research on antibiotics has been slow to develop, with few new antibiotics being developed and most newly approved antibiotics being slightly different variants of existing drugs [68]. In 2020, researchers at MIT used deep learning algorithms to analyze more than 100 million compounds in a few days. They discovered a new antibiotic capable of killing 35 potentially deadly bacteria [4]. In July 2022, AI lab DeepMind's *AlphaFold* announced that it had predicted the structures of nearly all of the more than 200 million proteins known to science and made them freely available [18]. This scientific leap, which covers almost all proteins of known organisms in the DNA database, is thought to have the potential to have a huge impact on global problems such as famine and disease [29]. Just a few months later, in November 2022, researchers at another tech giant, *Meta*, used AI to predict the structure of more than 617 million proteins from bacteria, viruses, and other as-yet-uncharacterized microbes in just two weeks [9]. This macrogenomic database reveals the structures of hundreds of millions of the planet's least-known proteins, promising to accelerate advances in medicine, renewable energy and green chemistry [49]. Moreover, many areas of cutting-edge scientific research have, traditionally, required large financial investments in laboratories, equipment and huge amounts of manpower and time to make some progress. However, now artificial intelligence technologies are changing the game. Elements of the scientific process will increasingly be driven by intelligent agents, especially for processes that do not rely on creativity and abstract thinking [6]. Therefore, countries that invest more in artificial intelligence will benefit more from it and will gain higher rates of scientific and technological development.

Although AI technology has all these advantages in scientific research, it still has some shortcomings. The difficulty in using AI technology for scientific research is how to interpret the data. While AI can often make very accurate predictions, it cannot itself explain why and how it makes such predictions. The AI's processing is unknown to humans, which is known as the algorithmic 'Black Box'. Despite the recent academic and industry efforts on 'Explainable AI' (XAI) [17], the results are hardly satisfactory [39]. This means that while advanced technologies can help

researchers to complete the process of data collection and analysis, more efforts are needed to interpret the data and translate the analysis results into knowledge. Some of the findings accomplished by AI are still to be verified and interpreted by human scientists.

In addition to technology, limiting the development of AI in science and technology is the availability of data. In the age of artificial intelligence and big data, more and more data is being collected by emerging digital methods, such as online communities, eye-tracking and wearable technology. On the one hand, this can be great for some developed countries. But on the other hand, some less developed countries and regions are not benefiting from the same technological development. In fact, collecting data in these places is more difficult than in other regions. And as a result, the availability of data is leading to the polarization of market research. That is, in developed countries and regions, the large amount of data allows AI to be better applied to various scientific studies, thus driving rapid growth in science and technology. In contrast, in less developed countries and regions, there is rarely enough data available for research [34]. Take COVID19 related services as an example. Developed countries and regions have near real-time access to a wide range of COVID-19-related data. This allowed for timely analysis and adjustments to vaccine supplies, financial assistance, travel policies, medical aid, and more. In contrast, in underdeveloped regions with more vulnerable populations, there is still not enough data available to study and provide targeted help to people [50].

One possible way to address the difficulties in obtaining data in underdeveloped areas, which affects the conduct of research, is to use artificial intelligence to infer data. *Facebook* has already done a good demonstration of this. Since 2017, *Facebook* has applied artificial intelligence to satellite data to map roads around the world that are unmapped and missing due to insufficient data. Facebook has named the project *Map With AI* [28], and these AI-mapped maps have helped during COVID-19 vaccine deliveries [50]. Similarly, more AI techniques can be used to infer data from underdeveloped regions for research.

4 Challenges of AI in National Security

As mentioned above, AI is bringing disruptive changes in the economy, employment, education and culture, public security, science and technology. As we have read, every country in the wave of this technological revolution is facing new challenges.

4.1 Disinformation Undermines National Security

In a healthy society, institutions, groups, and individuals make informed decisions based on reliable information received from a variety of sources on a daily basis. This is the foundation of a well-functioning society. However, when people are misled

by erroneous, false, or even maliciously falsified information, it may influence their rational judgment. They may even act contrary to their true intentions and interests, which can create many problems for our security on a local and national level. In truth, the ease of access to AI technology today has lowered the threshold for creating false information, posing risks and hazards to society and national security. Interpersonal interactions are often based on confirmation and trust in each other's identities. However, when identities can be falsified, this trust can be harmed, even with serious consequences, and AI-generated photos are one example of this.

Photos of faces generated with generative adversarial networks (GAN) are so realistic that it is difficult for the average person to tell if they are real photos or computer-generated. This is a technique that actually poses the risk of identity forgery. For example, many GAN-generated photos can be accessed from the *thispersondoesnotexist.com* website [40]. Unsuspecting people can use these photos and fictitious personal information to gain the trust of others and commit criminal acts. Officials in the UK, France and Germany have all issued warnings detailing how foreign spies have contacted thousands of people through *LinkedIn.* These spies target people through Deepfake-generated composite portraits and fake social media profiles. Once a target person accepts these spy invitations to connect, other users on the site can view the connection as an endorsement, thereby lowering their guard [63]. Even more alarming than false identities is manipulated information. This can be used to shape public opinion or undermine trust in the authenticity of information. As a result, it can disrupt and undermine social order, democratic institutions and national cohesion. The U.S. *Cybersecurity and Infrastructure Security Agency (CISA)* classifies information manipulation into three types, namely Misinformation, Disinformation, and Malinformation [14].

For example, Russian operatives used AI during the 2016 U.S. presidential election to create and disseminate fake news articles and social media posts in order to influence the outcome of the election [41]. These fake stories were designed to spread quickly and widely via the Internet, and many people believed them to be true. The use of artificial intelligence in this context enabled Russian agents to generate large amounts of false content in a short period of time, making it difficult for people to distinguish between truth and falsehood. This type of information manipulation can have serious consequences, as it can erode trust in the media and democracy.

4.2 Human Rights Issues in Artificial Intelligence

As AI technology rapidly evolves, they are also concerns about the ethics of AI. It seems that the generation of new laws is almost always a reactive response to address existing issues. Therefore, they usually lag behind the real-world needs, as do the industry norms and standards. It is important to note that just because an activity does not violate current laws, it does not mean that it is ethical.

The growth of the Internet and IoT has made vast amounts of data readily available; AI and big data technologies have in turn made it easy to infer new insights from a variety of seemingly unrelated data [73]. The aggregation of data often comes with the risk of privacy breaches. For example, the retailer giant *Target* could infer that a high school girl was pregnant based on her purchase history and hand out baby product ads to her, even while the girl's father was still in the dark [38]. As the cost of storing electronic data continues to decrease, more and more users data will be stored for long periods of time. In addition, some deidentified data may be re-identified, and non-sensitive data may become sensitive due to new information generated by data aggregation. This all can threaten the privacy of users. For example, many mobile applications collect users' locations to provide better services and then use this location data for marketing research. This location data, even de-identified, can be used to figure out the privacy of users, where they go, who they meet, and their daily activities, all of which can be easily tracked. Even the most privacy-conscious people, such as former USA President Trump, Secret Service agents or Supreme Court technicians, are not exceptions [67].

Furthermore, the development of Internet of Things (IoT) and wearable devices has resulted in users' physiological data (e.g. breathing, heartbeat, blood pressure, pulse), behavioral habits (e.g. sleep patterns, exercise habits, driving habits), and behavioral preferences (e.g. dietary preferences, shopping preferences, entertainment preferences) being captured precisely. Frighteningly, this data can even analyze information about their state-of-behavior more accurately than the users themselves can. By analyzing a person's posts, likes, ad clicks, and browsing history, it is possible to figure out their personality traits, attitudes, feelings, perceptions, beliefs, and product/service brand preferences. In addition, online users rarely need to disguise themselves to meet social expectations as they do offline, so they will show a more authentic side of themselves. Moreover, this user data collected on an ongoing basis can more accurately reflect the stable characteristics of the users in question. With the help of Content Mining and Natural Language Processing (NLP) techniques, researchers can analyze surveys, web text, online comments, tweets, etc. to generate useful insights, such as for opinion mining and sentiment analysis. This information extracted from text data can often be used for psychographic profiles to more accurately reflect the attitudes, interests, personalities, values, opinions and lifestyles of users, but often without the data subject's knowledge, exposing them to privacy violations [79].

Some data protection laws, such as the EU's *General Data Protection Regulation (GDPR)*, regulate the collection, processing and storage of data. However, there are still many operational gaps and a need for more practical and detailed industry standards to ensure that data is used ethically before, during, and after research. With the rapid development of AI, there is an increasingly urgent need to address the ethical issues associated with it. This requires careful consideration by policymakers, academia, industry, and other stakeholders. Especially, when they need to take effective measures to ensure that the benefits of AI are balanced with the need to protect the rights and the interests of individuals.

Another human rights hazard posed by AI is discrimination and bias. Datasets used to train AI models may have sample bias, or developers may inadvertently introduce bias into the system, which can produce biased results, and the use of these biased results in automated decision-making can result in discriminatory treatment. For example, a study on a health management algorithm that affects millions of people in the United States showed that because the algorithm predicts disease risk in terms of how much health care costs, but inequities in access to care mean that black patients have less health care costs, black patients are in fact much sicker than white patients for a given risk score predicted by the system [53]. As the use of AI becomes more prevalent, discrimination and bias in AI can have widespread negative effects, and without effective action, any individual or group may be treated unfairly because of race, color, gender, age, religion, sexual orientation, and a variety of other reasons.

4.3 Legislative Improvement of Artificial Intelligence

Although various countries have enacted various laws on data protection, the legal protection provided in practice is still inadequate and data privacy violations still occur. For example, in the 2010s, personal data belonging to millions of *Facebook* users was collected without consent by the British consulting firm *Cambridge Analytica* for political advertising. A New York Times report called 'Times Privacy Project Links to an external site' also revealed how cell phone location data can expose an individual's whereabouts. These invasions of privacy in fact put people in danger—even presidents and their security experts, What's more, with the rapid development of artificial intelligence, even some unrelated information may generate new sensitive information under the role of data aggregation. For example, merchants can generate psychographic profiles from users' purchase records to understand their interests, beliefs, values and other personal traits. Therefore, in fact, every piece of personal information we handle is more or less related to personal privacy.

As we discussed earlier, in the era of Big Data, data is the new oil, containing valuable information resources [7] that can bring power and wealth to its owners. The rapid concentration of citizen data to monopolies is a hidden danger to democracy, human rights and even national security. Once such a monopoly is established, it will be difficult to counterweight it. Therefore, the timely establishment of a sound legal system to regulate access to data and the use of artificial intelligence is necessary to ensure long-term healthy economic development and national security. Indeed, to prevent inappropriate uses of technology, data, and automated systems from threatening the rights of the American public, the White House *Office of Science and Technology Policy (OSTP)* has released the *Blueprint for an* AI *Bill of Rights* in October 2022. This bill identifies five principles to guide the design, use, and deployment

of automated systems. These principles include Safe and Effective Systems, Algorithmic Discrimination Protections, Data Privacy, Notice and Explanation, as well as Human Alternatives, Consideration, and Fallback. Of these five principles, data privacy is considered a fundamental though cross-cutting principles are required to achieve all the other goals in this framework [80].

The *Federal Trade Commission (FTC)* is increasingly using algorithmic vandalism as a tool to control technology companies. Algorithmic vandalism requires companies that illegally collect data to "delete the data they illegally obtained, destroy any resulting algorithms, and pay fines for their violations. Algorithmic vandalism may hold organizations accountable not only for the way they collect data, but also for the way they process it" [8, p. 1]. In fact, legislatures around the world are recognizing the need to hold companies that illegally collect data, to develop or train algorithms, accountable. As a result, additional regulations may be introduced to mitigate the problems associated with these practices.

5 Discussion

As we discussed earlier, AI is changing people's lives in every aspect. In doing so, it is also posing some challenges, and we should proactively take effective measures to reduce possible risks and make it better for human beings.

5.1 Reducing Regional Imbalances

Regional imbalances are a common problem between countries and between regions within countries. Various reasons such as historical development and resource distribution cause imbalances, along with differences in development between countries or regions in the fields of economy, culture, healthcare, education, etc. We have seen the impact of this imbalance from the gap between developed and underdeveloped countries, and even between regions in countries with large territories like China, India, and the United States. This imbalance has caused huge differences in people's income levels and even happiness indices. If the opportunity is grasped, the application of AI technology is expected to bring more development opportunities and faster development speed to relatively less developed regions and reduce regional disparities. Less developed regions can even benefit from the development of AI technology more than developed regions, so that people in these regions can enjoy improved medical conditions, equal educational opportunities, and more employment opportunities. For example, the *World Health Organization's World Cancer Report* shows that more than 60% of the global cancer burden occurs in Asia, Africa, and LMICs in Central and South America, and 70% of cancer deaths occur in these regions [82]. Therefore, the development of AI in cancer treatment will benefit people in these regions even more. Smaller regional development gaps will lead to a more

prosperous international market and improve the common welfare of all humanity. But on the other hand, we should also be aware that this gap is likely to continue to widen in the area of cutting-edge technology. A comparative study of eastern, central, and western China shows that industrial intelligence improves inequality in consumer welfare between regions, while having the potential to exacerbate regional inequality in innovation [44]. Less developed regions may face greater technological dependence if they fail to keep and catch up with the ever advancing technological wave.

5.2 Building Accountable AI Systems

To reduce the risks posed by AI, it should first be controlled at the source. By this we mean that it should be transparent, fair, and accountable from the time when it is first designed, developed, and then used. Firstly, accountability should be considered from the earliest stages of development, with a commitment to eliminate bias and/or unfairness. This means incorporating mechanisms to explain and justify decisions made by AI, and ensuring that AI acts in a responsible and ethical manner. Secondly, AI systems should be tested and evaluated to ensure that they operate as intended and do not make biased and/or unfair decisions. Thirdly, there should be a strong legal and regulatory framework to support responsible AI. This framework should define the rights and responsibilities of AI developers, users, and other stakeholders. In addition, it should provide clear guidance on how to ensure that AI systems act in a responsible manner.

To better control the risks of AI systems, many countries and regions are committed to building responsible AI. The European Union proposed the *European Artificial Intelligence (AI) Act* in April 2021, a law that classifies applications of AI into three risk categories [26]. Under this category, applications and systems that pose unacceptable risks would be banned, high-risk applications would need to comply with specific legal requirements, and applications not explicitly banned or classified as high-risk would be largely unregulated [3]. However, some scholars have questioned this conflation of 'trustworthiness' with 'acceptability of risk' and have pointed out that there is still a threat of misalignment between the actual level of trust and the trustworthiness of the applied AI [43].

5.3 Strengthen AI Education and Skills Training

In contrast to the rapid development of AI technology, education is a long, slow process. There is a need to take a forward-looking view to provide the younger generation and those educating them with the knowledge and skills to exist safely in this AI era. For young people, it is important to enhance their AI literacy. This includes giving them a basic understanding of what AI is and how it works. It gives

them access to AI tools and systems, and enables them to learn how to use AI to solve real-world problems. For educators, it should focus on helping students develop the critical thinking and problem-solving skills that are useful in an AI-driven world. This could include teaching them on how to assess and understand the reliability and validity of information, how to solve complex problems, and how to think creatively and innovate. It is also important to educate young people about the ethical and social implications of AI, such as the potential biases and limitations regarding AI algorithms, and how to use AI responsibly and ethically.

For those who may be affected by AI technologies and need to upgrade their skills, they should be provided with vocational skills training tailored to their needs. For example, this can be an AI-driven, personalized system of skills enhancement that helps them learn faster and more effectively to adapt to new job demands. The state should also make more AI research resources available to the public to enrich the AI research ecosystem. More educational resources for AI professionals should be made available to a broader population so that AI professional education is equitably accessible to all, especially those from underrepresented groups, as one measure to avoid or reduce bias. According to an analysis on LinkedIn, only 22% of AI professionals worldwide are women, a huge contrast to the 78% of male professionals [77]. And the AI industry, dominated by men, tends to produce systems and products with gender biases and stereotypes [81]. For example, AI virtual personal assistants are often set up with feminine images and voices that are in fact an extension of the stereotype of the female secretary and reinforce the discrimination of women being in a submissive position [1, 13]. To bridge the resource divide in AI research, the U.S. established the *National AI Research Resources (NAIRR) Task Force* in June 2021 to establish a *National AI Research Resource* that democratizes access to AI R&D for U.S. researchers and students by making computational infrastructure, public- and private- sector data, and testbeds easily accessible [54].

5.4 Improve the Legal and Regulatory System

The data used for AI learning is closely related to people's privacy, and strengthening the protection of data is an important prerequisite for protecting people's privacy. At present, more than 130 countries have enacted laws and regulations related to privacy protection [31], but people's privacy is still not fully protected, and with the development of AI technology, more risks of privacy infringement may emerge. This means that our legal system still needs to be further refined in detail to make it more workable. Traditional privacy protection laws or data protection laws focus on the protection of existing data processing processes such as data collection, storage, and transmission, and fail to provide protection for new information generated in the process of data aggregation, analysis, and inference [74]. In today's increasingly intelligent AI algorithms, the law should provide more detailed and clear regulation of data aggregation and inference, and how to apply the results. Moreover, while many laws provide for data desensitization, there is no clear definition of what standards should

be met [58], which makes it de facto difficult for the laws to be effectively enforced. The authors of this paper argue that when data controllers provide desensitized data to external parties, they should have a reasonable expectation of the possibility of data re-identification and take proactive measures to prevent re-identification.

In response to the increasingly pressing legal issues regarding the use of AI, countries are stepping up their efforts to improve legislation related to AI. In the United States, at least 17 states have introduced general AI bills or resolutions as of August 2022 and enacted them in Colorado, Illinois, Vermont, and Washington [52]. In October 2020, the European Parliament adopted several resolutions related to AI, including on ethics, liability and copyright [22–24], and in 2021, resolutions were adopted on AI in criminal matters and in the fields of education, culture and audio-visual [25, 72]. The European Commission has put forward a proposed regulatory framework on AI that proposes a balanced and proportionate horizontal regulatory approach to AI that is limited to the minimum requirements necessary to address the risks and issues associated with AI without unduly restricting or impeding technological development or otherwise disproportionately increasing the cost of placing AI solutions [27].

6 Conclusion

In the era of Industry 4. 0, artificial intelligence has been involved in all aspects of social development. As we have discussed, this is closely related to the strength of a nation's power. But there needs to be a balance. Indeed, the rapid development of artificial intelligence technology is forcing every country to take on the opportunities it affords but also to face the challenges it creates.

On the one hand, the discipline of artificial intelligence has only been in existence for a few decades, and it has already experienced a stagnation (e.g,. lack of growth). It has really only been developing rapidly in the last two decades. AI's development can be described as being at the starting phase of a race in which every runner has the opportunity to win. Indeed, every country can benefit from this disruptive technology to promote economic development, improve the labor environment, enhance the quality of education, enrich cultural life, and accelerate technological development for the benefit of its people. As we have read, the development of artificial intelligence today cannot be achieved without the contributions of scientists from all disciplines around the world. The prosperity of artificial intelligence depends on having a unified vision of benefiting all of humanity. Artificial intelligence has given people a fairer choice, reduced the gap between the rich and the poor, freed workers from monotonous and heavy manual labor. It has also given people the opportunities to choose job positions that better realize their values. It has made it possible for children in remote areas to enjoy personalized educational experiences. Moreover, it has allowed people suffering from illness to benefit from medical breakthroughs discovered through the technological development.

On the other hand, we should also be soberly aware that the rapid development of AI technology is creating challenges and putting every country under immense pressure. Artificial intelligence is still developing faster than people can think. This challenges and even counters the existing ethical and legal systems which are still dangerously lagging far behind AI. Just as nuclear energy can either generate electricity for the benefit of humanity or be used as a weapon to destroy it, AI is similar. Therefore, we need to push to ensure that the tremendous energy released by AI is used for good purposes, rather than being exploited for the interests of criminal groups. History has shown countless times that when a force is extremely powerful, it often lacks a counterweight. This usually has disastrous consequences, and in many ways, the same can be said for the field of AI. Therefore, the time has come to demand strict governance and control of AI. A nation needs to ensure that their citizens are protected and that their rights are not violated. Especially, when people's data can be in the hands of a few large corporations. There needs to be standards, an approved way of designing, developing and working with AI. We need to make sure that there are checks in place and that these new technologies are not being misused. Finally, we need a society in which everyone benefits from AI, rather than being constrained, manipulated and/or endangered by it. This is our only hope when striving for a responsible national security.

References

1. Adams R (2022) Artificial intelligence has a gender bias problem—just ask siri. The Conversation, Sep. https://theconversation.com/artificial-intelligence-has-agender-bias-problem-just-ask-siri-123937
2. AFP (2022) Dutch police create deepfake video of murdered boy, 13, in hope of new leads. https://www.theguardian.com/world/2022/may/23/dutchpolice-create-deepfake-video-of-murdered-boy-13-in-hope-of-new-leads
3. Artificial Intelligence Act (2022) What is the EU AI act? the artificial intelligence act, Nov. https://artificialintelligenceact.eu/
4. BBC (2020) Scientists discover powerful antibiotic using AI. BBC News, Feb. https://www.bbc.com/news/health-51586010
5. Becker BA, Denny P, Finnie-Ansley J, Luxton-Reilly A, Prather J, Santos EA (2023) Programming is hard–or at least it used to be: educational opportunities and challenges of AI code generation
6. Briscoe E, Fairbanks J (2020) Artificial scientific intelligence and its impact on national security and foreign policy. Orbis 64(4):544–554
7. Buhl HU, Röglinger M, Moser F, Heidemann J (2013) Big data. Bus Inf Syst Eng 5(2):65–69
8. Caballar RD (2022) "Algorithmic destruction" policy defangs dodgy AI new regulatory tactic of deleting ill-gotten algorithms could have bite. https://spectrum.ieee.org/ai-concerns-algorithmic-destruction
9. Callaway E (2022) Alphafold's new rival? meta AI predicts shape of 600 million proteins. Nature News, Nov. https://www.nature.com/articles/d41586-022-03539-1
10. Chen J, Song N, Su Y, Zhao S, Zhang Y (2022) Learning user sentiment orientation in social networks for sentiment analysis. Information Sciences
11. Chen M, Tworek J, Jun H, Yuan Q, Pinto HPdO, Kaplan J, Edwards H, Burda Y, Joseph N, Brockman G et al (2021) Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374

12. Chen X, Liu C, Li B, Lu K, Song D (2017) Targeted backdoor attacks on deeplearning systems using data poisoning. arXiv preprint arXiv:1712.05526
13. Chin C, Robison M (2022) How AI bots and voice assistants reinforce gender bias. Brookings, Mar. https://www.brookings.edu/research/how-ai-botsand-voice-assistants-reinforce-gender-bias/
14. CISA: Homepage: Cisa. Cybersecurity and infrastructure security agency (CISA). https://www.cisa.gov/
15. Cole S (2019) Deepnude: the horrifying app undressing women, Jun. https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudesof-any-woman
16. COMEST: Preliminary study on the ethics of artificial intelligence. Unesdoc.unesco.org. https://unesdoc.unesco.org/ark:/48223/pf0000367823
17. Das D, Nishimura Y, Vivek RP, Takeda N, Fish ST, Ploetz T, Chernova S (2021) Explainable activity recognition for smart home systems. arXiv preprint arXiv:2105.09787
18. Deepmind: alphafold reveals the structure of the protein universe. https://www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-proteinuniverse
19. Department of Justice (2020) Justice department sues monopolist google for violating antitrust laws. The United States Department of Justice, Oct. https://www.justice.gov/opa/pr/justice-department-sues-monopolist-googleviolating-antitrust-laws
20. Ding S, Tian Y, Xu F, Li Q, Zhong S (2019) Trojan attack on deep generative models in autonomous driving. In: International conference on security and privacy in communication systems, pp 299–318, Springer
21. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
22. EU: European parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies (2020/2015(ini)). EUR, https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:52020IP0277amp;from=EN
23. EU: European parliament resolution of 20 October 2020 with recommendations to the commission on a civil liability regime for artificial intelligence (2020/2014(inl)). EUR, https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:52020IP0276amp;from=EN
24. EU: European parliament resolution of 20 October 2020 with recommendations to the commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(inl)). EUR, https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:52020IP0275amp;from=EN
25. EU: European parliament resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters (2021). https://www.europarl.europa.eu/doceo/document/TA-9-20210405_EN.html
26. EU: 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 (Sep 2021). https://artificialintelligenceact.eu/the-act/
27. EU: 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. EUR (2021), https://eur-lex.europa.eu/legalcontent/EN/TXT/HTML/?uri=CELEX:52021PC0206amp;from=EN
28. Gao X (2020) AI is supercharging the creation of maps around the world. Tech at Meta, Jul. https://tech.fb.com/artificial-intelligence/2019/07/ai-is-superchargingthe-creation-of-maps-around-the-world/
29. Geddes L (2022) Deepmind uncovers structure of 200m proteins in scientific leap forward. The Guardian, Jul. https://www.theguardian.com/technology/2022/jul/28/deepmind-uncoversstructure-of-200m-proteins-in-scientific-leap-forward
30. Goldhaber MH (1997) The attention economy and the net. First Monday 2(4). https://doi.org/10.5210/fm.v2i4.519. https://journals.uic.edu/ojs/index.php/fm/article/view/519
31. Greenleaf G (2019) Global data privacy laws 2019: 132 national laws & many bills
32. Guardian (2020) US justice department sues Google over accusation of illegal monopoly. The Guardian, Oct. https://www.theguardian.com/technology/2020/oct/20/us-justice-departmentantitrust-lawsuit-against-google

33. Guardian (2020) What are deepfakes—and how can you spot them? https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakesand-how-can-you-spot-them
34. Hair JF, Ortinau DJ, Harrison DE (2010) Essentials of marketing research, vol 2. McGraw-Hill/Irwin, New York, NY
35. Hawksworth J, Berriman R, Goel S (2018) Will robots really steal our jobs? an international analysis of the potential long term impact of automation
36. Hern A (2021) 'I don't want to upset people': Tom Cruise deepfake creator speaksout. https://www.theguardian.com/technology/2021/mar/05/how-startedtom-cruise-deepfake-tiktok-videos
37. Hill K (2022) Clearview AI, used by police to find criminals, is now in public defenders' hands. https://www.nytimes.com/2022/09/18/technology/facialrecognition-clearview-ai.html
38. Hill K (2022) How target figured out a teen girl was pregnant before her father did. Forbes, Oct. https://www.forbes.com/sites/kashmirhill/2012/02/16/howtarget-figured-out-a-teen-girl-was-pregnant-before-her-fatherdid/?sh=7d59935a6668
39. Innerarity D (2021) Making the black box society transparent. AI Soc 36(3):975–981
40. Karras T (2022) This person does not exist. https://thispersondoesnotexist.com/
41. Kelly M, Samuels E (2019) Analysis|how Russia weaponized social media, got caught and escaped consequences, Nov. https://www.washingtonpost.com/politics/2019/11/18/how-russia-weaponizedsocial-media-got-caught-escaped-consequences/
42. Kemp S (2022) Digital 2022: global overview report—datareportal—global digital insights. Data Reportal, May. https://datareportal.com/reports/digital-2022global-overview-report
43. Laux J, Wachter S, Mittelstadt B (2022) Trustworthy artificial intelligence and the European union AI act: on the conflation of trustworthiness and the acceptability of risk. Available at SSRN 4230294
44. Li S, Hao M (2021) Can artificial intelligence reduce regional inequality? evidence from China
45. Li Y, Choi D, Chung J, Kushman N, Schrittwieser J, Leblond R, Eccles T, Keeling J, Gimeno F, Lago AD et al (2022) Competition-level code generation with alphacode. arXiv preprint arXiv:2203.07814
46. Liu M, Yu J, He H, Wang R, Zhan H (2018) Research on industrial cluster and the siphon effect of talent accumulation. In: 2018 14th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), pp 815–818, IEEE
47. McCarthy J (2007) What is artificial intelligence?
48. McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the dart mouth summer research project on artificial intelligence, August 31, 1955. AI Mag 27(4):12–12
49. Meta (2022) New AI research could drive progress in medicine and clean energy, Nov. https://about.fb.com/news/2022/11/ai-protein-research-could-driveprogress-in-medicine-clean-energy/
50. Meta AI (2022) How maps built with Facebook AI can help with covid-19 vaccine delivery. https://ai.facebook.com/blog/how-maps-built-with-facebook-aican-help-with-covid-19-vaccine-delivery/
51. Microsoft: Machine learning made easy, Lobe. https://www.lobe.ai/
52. NCSL (2022) Legislation related to artificial intelligence, Aug. https://www.ncsl.org/research/telecommunications-and-informationtechnology/2020-legislation-related-to-artificial-intelligence.aspx
53. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science 366(6464):447–453
54. Parker L (2022) Bridging the resource divide for artificial intelligence research, May. https://www.whitehouse.gov/ostp/news-updates/2022/05/25/bridgingthe-resource-divide-for-artificial-intelligence-research/
55. Pearce H, Ahmad B, Tan B, Dolan-Gavitt B, Karri R (2022) Asleep at the keyboard? assessing the security of GitHub Copilot's code contributions. In: 2022 IEEE symposium on security and privacy (SP), pp 754–768, IEEE
56. Person, Francois Aulner, F.Y.C. (2021) Google loses challenge against EU antitrust ruling, fine. Reuters, Nov. https://www.reuters.com/technology/eu-court-upholdseu-antitrust-ruling-against-google-2021-11-10/

57. PwC: Pwc's global artificial intelligence study: sizing the prize. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificialintelligence-study.html
58. Quinn P (2021) The difficulty of defining sensitive data—the concept of sensitive data in the EU data protection framework. German Law J 22(8):1583–1612
59. Roose K (2022) A.I.-generated art is already transforming creative work. https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-dall-e2.html
60. Roose K (2022) An A.I.-generated picture won an art prize. Artists aren't happy. https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligenceartists.html
61. Roose K (2022) A coming-out party for generative A.I., Silicon Valley's new craze. https://www.nytimes.com/2022/10/21/technology/generative-ai.html
62. Russell SJ (2010) Artificial intelligence a modern approach. Pearson Education Inc.
63. Satter R (2019) Experts: spy used AI-generated face to connect with targets. https://apnews.com/article/ap-top-news-artificial-intelligence-socialplatforms-think-tanks-politics-bc2f19097a4c4fffaa00de6770b8a60d
64. Simon HA et al (1971) Designing organizations for an information-rich world. Comput Commun Public Interes 72:37
65. Smith CS (2022) 'No-code' brings the power of A.I. to the masses. The New York Times, Mar. https://www.nytimes.com/2022/03/15/technology/ai-nocode.html?action=clickamp;module=RelatedLinksamp;pgtype=Article
66. Teachable Machine (2022). https://teachablemachine.withgoogle.com/
67. Thompson SA, Warzel C (2019) How to track president trump. The New York Times, Dec. https://www.nytimes.com/interactive/2019/12/20/opinion/locationdata-national-security.html
68. Trafton A (2020) Artificial intelligence yields new antibiotic. MIT News, Massachusetts Institute of Technology. https://news.mit.edu/2020/artificialintelligence-identifies-new-antibiotic-0220
69. Turing AM (2012) Computing machinery and intelligence (1950). In: The essential turning: the ideas that gave birth to the computer age, pp 433–464
70. UNESCO (2021) AI and education: guidance for policy-makers. Unesdoc.unesco.org. https://unesdoc.unesco.org/ark:/48223/pf0000376709
71. United Nations (2022) Will robots and AI cause mass unemployment? not necessarily, but they do bring other threats. https://www.un.org/en/desa/will-robots-and-ai-cause-mass-unemploymentnot-necessarily-they-do-bring-other
72. Verheyen S (2021) Report on artificial intelligence in education, culture and the audiovisual sector: A9-0127/2021: European parliament. REPORT on artificial intelligence in education, culture and the audiovisual sector | A9-0127/2021 | European Parliament. https://www.europarl.europa.eu/doceo/document/A-9-20210127_EN.html
73. Wachter S (2020) Affinity profiling and discrimination by association in online behavioral advertising. Berkeley Tech LJ 35:367
74. Wachter S, Mittelstadt B (2019) A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Colum Bus L Rev, p 494
75. Walsh F (2020) AI 'outperforms' doctors diagnosing breast cancer. BBC News, Jan. https://www.bbc.com/news/health-50857759
76. Wang Y, Chaudhuri K (2018) Data poisoning attacks against online learning. arXivpreprint arXiv:1808.08994
77. Weforum (2018) Global gender gap report 2018. World Economic Forum. https://www.weforum.org/reports/the-global-gender-gap-report-2018
78. Weforum (2020) The future of jobs report 2020. World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2020
79. Wells WD (1975) Psychographics: a critical review. J Mark Res 12(2):196–213
80. Whitehouse: Blueprint for an AI bill of rights. https://www.whitehouse.gov/wpcontent/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf
81. Whittaker M, Crawford K, Dobbe R, Fried G, Kaziunas E, Mathur V, West SM, Richardson R, Schultz J, Schwartz O et al (2018) AI now report 2018. https://ainowinstitute.org/AI_Now_2018_Report.pdf

82. Wild C, Weiderpass E, Stewart BW (2020) World cancer report: cancer research for cancer prevention. IARC Press
83. Xiang C (2022) AI is probably using your images and it's not easy to optout. https://www.vice.com/en/article/3ad58k/ai-is-probably-using-yourimages-and-its-not-easy-to-opt-out
84. Yu S, Carroll F (2022) Insights into the next generation of policing: understanding the impact of technology on the police force in the digital age, pp 169–191

Facial Recognition Technology, Drones, and Digital Policing: Compatible with the Fundamental Right to Privacy?

Océane Dieu

Abstract Drones are the new gadget law enforcement agencies cannot get enough of. These agencies widely deploy drones, amongst others, for search and rescue operations or in response to a natural disaster. The benefits these drones offer are unquestionable. However, these drones are increasingly being deployed for a less self-evident and legitimate purpose: surveillance. The recourse to drones for surveillance operations is highly problematic, given its intrusiveness on citizens' fundamental right to privacy. Furthermore, this intrusiveness becomes even more worrisome when these drones are equipped with facial recognition technology. Consequently, this paper will critically examine law enforcement's recourse to facial recognition technology in drones and the worrying consequences of such deployment on citizens' fundamental right to privacy.

Keywords Artificial intelligence · Drones · Right to privacy · Surveillance · (Live) Facial recognition technology · Law enforcement · Digital policing · European convention on human rights

1 Introduction

The presence of Internet of Things ('IoT') devices in people's daily lives has become a normality. However, the mainstream usage of such devices has desensitised their users regarding the dangers such devices might pose to their privacy. Whilst the use of IoT devices and their consequences on people's private lives is a matter of their own choice, the recourse to these devices is more problematic when the state utilises them in the name of security.

O. Dieu (✉)
Stibbe, Brussels, Belgium
e-mail: oceane.dieu@stibbe.com

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_3

One such problematic recourse by the state is the deployment of facial recognition technology ('FRT') in drones by law enforcement agencies. Academic, governmental and civil society actors have criticised the further development of law enforcement's use of facial recognition technology [1, 12, 25, 32, 53]. This criticism is not limited to the academic sphere. As such, Bragias et al.'s [7] empirical research of 460 public YouTube posts, containing public commentary on the use by police of facial recognition technology, has shown that the majority of the authors of the posts perceive the use of such technology by law enforcement agencies negatively (75.4% against, 15.4% positive perception and 9.2% neutral). Given this negative perception by the people subjected to it, it is questionable why states continue to try to justify the recourse to such technology. Especially in a democracy.

Given its worrisome consequences on citizens' right to privacy, this paper will critically examine the police's recourse to such highly intrusive technology.[1] More specifically, it will analyse the threats to the deployment of facial recognition technology, a type of artificial intelligence ('AI'), in drones by law enforcement agencies for surveillance purposes. To do so, the paper will start by setting the scene (Sect. 2) regarding drones (Sect. 2.1) and facial recognition technology (Sect. 2.2). Afterwards, the paper will address the risks related to the use by law enforcement agencies of drones that rely on facial recognition technology (Sect. 3). Last, whether increased transparency and explainability of facial recognition technology could be the solution to these threats will be discussed (Sect. 4). Consequently, this paper will not examine the military deployment of drones or their commercial use (such as the delivery of goods).[2] Furthermore, given the limited scope of this work, the choice was also made not to address the use of (non-)lethal *armed* drones by law enforcement agencies.[3]

[1] Due to the limited scope of this work, other relevant human rights, such as the right to a fair trial that might be endangered when facial recognition technology wrongly qualifies a person as a criminal (see [22] for more information), the infringement of constant surveillance on a person's right to human dignity (see [16] for more information), or the unlawful interference with a person's right to data protection (see [16] and [32] for more information) will not be covered.

[2] For more information on the military use of drones, see [41]. For more information on the commercial use of drones, see [29] and [41].

[3] For more information on non-lethal and lethal armed drones, see [8, 14, 15, 31, 42, 43, 46, 54].

2 Setting the Scene: Drones, Artificial Intelligence and Facial Recognition Technology

2.1 Drones: Flying Robots with a High Potential for Law Enforcement

Drones, otherwise also known as 'Unmanned Aerial Vehicles' or 'Unmanned Aerial Systems', the former referring to the flying vehicle, the latter additionally comprising the entire system, can be understood as unpiloted, flying robots that can either be controlled remotely or fly autonomously, depending on the degree of technology on which they operate [3, 5, 13]. In the United Kingdom ('UK'), Gallagher's [24] research has shown that out of 48 regional police forces questioned, 33 police forces responded that they directly use or own drones for, amongst others, the use of covert surveillance operations, the monitoring of public protests or the maintenance of public order. The UK is not the only country deploying drones in law enforcement operations. As such, both the UK and the United States of America were heavily criticised for their intrusive and extensive use of drones during the global Black Lives Matter protests of 2020 following George Floyd's death, given the interference that such surveillance techniques constitute with the protesters' fundamental right to freedom of expression and the linked right to freedom of peaceful protest [15, 24, 26, 51, 53].

Law enforcement agencies have recourse to drones for several reasons. Some of these reasons are entirely legitimate when drones are, for example, deployed in search and rescue operations, in pursuit of a suspect, in response to a disaster or ongoing dangerous crime situations, such as shootings [5, 18, 26, 29, 38]. In addition, the wide variety of drones available makes it an attractive tool for law enforcement agencies, given its suitability for different policing operations. As such, drones can vary according to their size (ranging from insect-sized to more noticeable drones), their flying capability, their autonomy (ranging from remote control to autonomous flight and adjustment during flight) or the technology on which they operate. Moreover, drones are cheaper, quieter, more concealable and mobile than expensive helicopters whilst delivering similar results. Furthermore, drones are more flexible than static CCTV cameras and might be considered an appropriate response to the cumbersome task of scanning the extensive amount of CCTV footage [2, 3, 5, 8, 15, 16, 18, 19, 26, 33, 36]. Last, law enforcement can use drones as part of their digital forensic operations when analysing the data the drones contain for evidence purposes [3]. This evidence can then be used to arrest or prosecute suspects [26]. Consequently, drones unquestionably offer law enforcement agencies a range of benefits and opportunities. However, law enforcement's deployment of this type of technology also dangerously contributes to the normalisation of constant surveillance by states [5].

Drones do not fly in a legal vacuum. The UK Civil Aviation Authority issued guidelines 'The Drone and Model Aircraft Code', along with the Air Navigation Order 2016, to which all drone users must comply [10]. As such, operators of drones are required to register themselves and the drones they use and pass an online test [24, 45, 51]. Given their role in the public sector, law enforcement agencies have to comply with a different set of rules. When law enforcement use drones for *covert* surveillance purposes, they are subjected to the rules laid out in Part II of the UK Regulation of Investigatory Powers Act 2000. However, when these agencies mobilise drones for *overt* surveillance purposes, including when facial recognition technology is implemented in the drones, the rules set out in the Surveillance Camera Code of Practice of 2013, as revised and amended in 2022, apply [51]. As such, the Code provides 12 guiding principles to which law enforcement agencies must adhere when utilising drones for overt surveillance practices. Amongst others, the Code provides the principle of proportionality (principle 1), which recently received the attention of the Courts in the R. (Bridges) v. South Wales Police case of 11 August 2020,[4] the requirement to take into account the privacy of individuals subjected to surveillance (principle 2) and publicity requirements (principle 3). Furthermore, the College of Policing published in March 2022 guidelines on the overt deployment of live facial recognition technology to locate people listed on a watch list [11], offering a more practical approach.[5]

2.2 Facial Recognition Technology Implemented in Drones

Facial recognition technology is one of many applications of AI (Sect. 2.2.1) that is increasingly and worryingly being deployed by law enforcement agencies (Sect. 2.2.2).

2.2.1 Artificial Intelligence

In the absence of a universally accepted definition of AI, reference is often made to the distinction between 'general' and 'narrow' AI [4]. General AI can be understood as machine intelligence capable of operating similarly to the human brain in its adaptability, reasoning and agency [4]. Contrary to general AI, which has not yet been achieved and the attainment of its full potential appearing to be decades away [1, 4], narrow AI is widely used. Narrow AI is machine intelligence capable of performing well-defined, single-tasked operations, such as playing chess, translating documents or image recognition [1, 4]. An important form of AI is machine learning, which is a mechanism that utilises AI to learn from its previous decisions and improve its functioning [30, 34, 44].

[4] The content and relevancy of this case will be analysed in Sect. 3.5.

[5] The guidance of the College of Policing will be addressed in more detail in Sect. 2.2.2.

The capability and autonomy of drones depend on the level of sophistication of the technology incorporated in the drone. Accordingly, the more advanced the AI in drones, the more potential the drone has to attain full autonomy and operate without human intervention [9, 27]. As such, highly developed AI, such as biometric tools, thermal scanners or wireless network sniffers, are being implemented in drones used by law enforcement [5, 18]. However, when drones are equipped with biometric tools, such as facial recognition technology, several human rights are endangered [1]. Law enforcement's recourse to facial recognition technology for surveillance purposes is consequently highly worrisome.

2.2.2 Facial Recognition Technology

Facial recognition technology is a form of AI that involves the "biometric processing of captured video images for the purposes of matching to a database and identifying individuals" [22, p. 327]. After such a match is found by the facial recognition algorithm, a human police officer will review the existence of a credible match [22, 23]. This matching process relies on the biometric features of a person, which can be understood as the "physical, physiological or behavioural characteristics of an individual, which allows or confirms the unique identification of that individual, such as facial images or dactyloscopic (fingerprint) data" [32, p. 7]. In S and Marper v. the United Kingdom, the European Court of Human Rights was brought to rule on the state's retention of DNA, a form of biometric data, in light of the right to privacy of the citizens subjected to such retention. The Court ruled that "any State claiming a pioneer role in the development of new technologies bears special responsibility for striking the right balance" regarding the impact such technologies have on citizens' right to privacy (S and Marper v. the United Kingdom, 2008, para. 112). Given that FRT also relies on biometric data, the Court's decision is relevant to the state's retention of biometric data collected by this technology. Hence, when integrating facial recognition technology in drones, the state has a responsibility to adopt a cautious and balanced approach by considering their citizens' right to privacy.

Facial recognition technology can be used for different purposes. This technology can be deployed for verification (one-to-one matching) purposes, such as when verifying a person's identity at the airport based on their passport. Moreover, FRT can be used to identify a person (one-to-many matching). A person's biometric features will be compared against biometric features contained in an external database, such as a watch list, to find a potential match [16].[6] This identification process by FRT is particularly relevant when analysing law enforcement surveillance practices through drones. Given the invasiveness of such technology, the College of Policing has developed guidelines on the use by law enforcement authorities of live FRT in drones when trying to locate certain persons on a watch list [11].

[6] A third possible recourse to facial recognition technology lies in its purpose of categorisation. However, due to the limited relevance of categorisation to the current discussion, this purpose will not further be discussed. For more information on the subject, see [16].

In March 2022, the College of Policing issued guidance on the overt deployment of live FRT to locate people listed on a watch list [11]. This guidance offers a more practical approach and response to the use of such technology by law enforcement authorities. Several highly laudable recommendations have been issued by the College of Policing. As such, they addressed the recourse to FRT when children or people with a disability are listed on the watch list. This explicit attention given to the particular vulnerable situation of these persons and the requirements of heightened scrutiny are praiseworthy elements of the College of Policing's guidance. Equally commendable is the College of Policing's recommendation to delete instantaneously and automatically a person's biometric template, registered by the live FRT when the person enters a room where this technology is deployed, when the biometric template does not lead to a match with a person on the watch list [11].

Moreover, Feldstein's [19] research identified 64 countries out of 176, both authoritarian and democratic regimes, that use FRT for surveillance purposes.

A more specific form of FRT is *live* facial recognition technology. This technology is the real-time "biometric processing of facial images taken in a public place, for the purpose of determining a person's identity and the potential retention of those images" [16, pp. 23–24, 23, 32, 53]. The recourse to live facial recognition technology is, however, significantly more intrusive than 'regular' FRT, given that live FRT operates in real-time. Moreover, live facial recognition technology is more intrusive than CCTV or ANPR cameras. Contrary to CCTV, live FRT relies on a person's biometric data to identify in *real-time* the person. Furthermore, unlike ANPR cameras, the objects of surveillance are not cars and license plates but *human* beings [6].

Similarly to the UK's early adoption of CCTV cameras and the deployment of facial recognition technology therein, its law enforcement body has been an early and active experimenter of live FRT [16, 53]. However, the invasiveness of this live facial recognition technology is highly worrisome, especially when deployed in drones by law enforcement authorities.

3 Risks Related to the Use of Drones Containing FRT by Law Enforcement Agencies

As set out previously, law enforcement agencies can use drones for entirely legitimate reasons, including search and rescue operations. Nevertheless, when drones are equipped with live facial recognition technology, several highly worrisome concerns rise. As such, relying on this technology creates a safety and security risk of the drone being hacked or being deficient (Sect. 3.1), and questions to whom liability of misuse of drones should be attributed arise (Sect. 3.2). Moreover, the deployment of FRT in drones contributes to the never-ending increase and normalisation of surveillance by the state (Sect. 3.3). Furthermore, given the false results, errors and biases in FRT (Sect. 3.4) and the threat such technology constitutes to citizens' right to privacy (Sect. 3.5), it is questionable whether such recourse can ever be justified.

3.1 Safety and Security Risks: Hackability and Deficiency of Drones

Like any other IoT device, drones are subject to hacking and deficiencies [2, 40, 50]. When criminals gain illegal access to drones by exploiting vulnerabilities in the technology, these drones can be misused for purposes initially not foreseen by law enforcement agencies and consequently compromise the drone's security system [35]. Whilst the author is unaware of case law concerning hacked drones of the police, it is conceivable that hackers could illegally take over control of the drone to surveil citizens or steal the data the drones contain. Moreover, hackers could use the drones to perform physical attacks on citizens, which imperils the general safety of citizens. Furthermore, when hackers would obtain access to drones of law enforcement agencies, the malicious usage of these drones could deceit the general population if these drones are labelled with police logos. As such, citizens would be misled in their trust of the drone's legitimate use, given that the hacker would benefit from the trust law enforcement agencies generally enjoy. Last, the general safety of the population could also be threatened by deficient drones when they run out of battery or when their internal systems fail. Consequently, the operator of the drone could lose control and the drone could fall onto people, causing severe physical damage.

3.2 Questionable Attribution of Liability: The Operator, the Programmer or the Drone?

The damage the misuse of drones can create by illegally surveying citizens, stealing the data it contains or physically injuring citizens when falling onto them raises the complex question of the attribution of liability [29]. When law enforcement drones are manipulated maliciously, attribution is complicated, given the difficulty of identifying the criminal operator of such a drone [29]. Furthermore, when drones reach (some level of) autonomy, the question of attribution increases in complexity, given that the drone operator's liability becomes less self-evident. The question then arises whether the drone itself or the programmer who coded the drone with deficient AI should be considered liable. This question of liability grows even more in importance given that machine learning algorithms operate without the programmer knowing the factors that the algorithms took into account to arrive at a particular result. This 'black box'-phenomenon [4] of human's incapacity to scrutinise the algorithm greatly overshadows the attribution of liability.

3.3 *FRT in Drones: The Never-Ending Normalisation of Constant Surveillance*

The recourse to facial recognition technology in drones for surveillance purposes is highly problematic given that it contributes to the already widespread surveillance practices of the state [33]. As such, the United Nations General Assembly recognised and expressed deep concern about the increased recourse and development of technology allowing states to deploy large-scale surveillance measures that endanger their citizens' right to privacy [48].

Surveillance, in general, is highly problematic when it leads to surveillance creep. 'Surveillance creep' refers to the practice where a surveillance mechanism or operation is initially adopted for a well-defined and legitimate purpose but is later expanded to a different purpose, which was initially not foreseen or allowed [2, 5, 13, 20, 28]. For example, when surveillance tools were initially deployed to combat crime, their use got quietly extended after 9/11 to anti-terrorism purposes. Conversely, anti-terrorism surveillance measures have become a widespread tool for traffic management or general surveillance under the pretext of 'increased security purposes' [21, 37]. Similarly, when drones containing facial recognition technology are initially introduced in law enforcement departments for the sole purpose of search and rescue operations, it becomes relatively easy and tempting to utilise the drone to oversee public manifestations under the pretext of public order [5]. This type of mission creep is highly worrisome given the absence of a legal basis for such further deployment and citizens' unawareness of this extensive recourse.

The state's practice of continuously surveying its citizens has a so-called 'chilling effect' on citizens' enjoyment of fundamental rights. A chilling effect is the effect a particular measure has on citizens who will self-censor their behaviour out of awareness of being subjected to that measure. More specifically regarding surveillance by (live) facial recognition technology in drones, when citizens are aware that the state monitors their behaviour through this invasive technology, they will adjust their behaviour. Consequently, they might be less inclined to exercise their fundamental rights to freedom of expression and freedom of assembly (both protected under the Convention for the Protection of Human Rights and Fundamental Freedoms 1950 [European Convention on Human Rights or ECHR], respectively in Articles 10 and 11) out of fear of constantly being surveyed. The consciousness of being watched leads citizens to modify and adapt their behaviour [1, 5, 16, 20, 23, 26, 33, 53]. This restriction of citizens' freedom of expression and assembly consequently undermines the democratic order. The police's recourse to (live) facial recognition technology in drones to survey citizens, therefore, constitutes a threat to democracy.

3.4 Highly Problematic and Untrustworthy: False Results, Errors and Biases in FRT

Although facial recognition technology has evolved rapidly over the last couple of years [16] and that the drones equipped with such technology can allegedly "recognise faces… in photos taken kilometres away" [30, p. 220], the technology is far from perfectly accurate. The inaccuracy of the technology leads to erroneous results that can take the form of false positives and false negatives [16]. Using the example of integrating live FRT in drones to identify, on the basis of an external database, protestors in a peaceful manifestation, false positives would wrongly identify a person as a match with the external database. In contrast, false negatives would erroneously not match a person with the external database. If a decision is made based on these false positive and negative results, the former would constitute a grave violation of the wrongly-qualified person's fundamental rights. The latter would endanger public safety given that the technology failed to recognise a public danger. On a small scale, these false results might appear insignificant. However, they become highly problematic when scaled to a whole population subjected to such technology. Research has shown that facial recognition software highly lacks accuracy when identifying or misidentifying non-Caucasian persons, especially women, but that it reaches a 99% accuracy rate when identifying Caucasian men. During its development and coding process, FRT is widely trained on features of Caucasian men. Consequently, the technology is unable to perfectly identify or reject a non-Caucasian person as a match. This inability leads to discriminatory findings where most false positive results concern non-Caucasian persons [7, 12, 16, 23, 40, 53].

Furthermore, technology is, like humans, embedded with biases. Technology cannot be neutral, given that it is created by (often unconsciously) biased humans. FRT is designed by humans who are biased in their personal beliefs, opinions and views, given their education, background or culture [7, 32, 52]. When creating facial recognition technology, programmers will unconsciously transfer their inherent biases to this technology. This 'human-induced bias' can be accompanied by false outcomes from inaccurate technology (as previously explained) or by historical data that is already biased, creating, as such, a 'data-driven bias' [52]. These biases can further take root when the technique of machine learning is applied to technology containing such biased data. The feedback loops machine learning technology creates, by reinforcing unnoticed errors, biases and false results, only amplify the biases that pre-existed in the programme, creating, as such, a 'machine self-learning bias' [39, 52]. However, if programmers are correctly trained to recognise their own internal biases and biases in technology, this problem could be mitigated. As Babuta et al. rightly point out, people have become aware of institutional and personal biases due to biased outcomes from machine learning technologies [4].

The distrust and scepticism towards facial recognition technology are not limited to academic spheres. For example, research from Urquhart and Miranda [49] showed that interviewed police officers doubted the accuracy and reliability of facial recognition technology. Moreover, the American-based company Axon declared in 2019 that it would refrain from implementing FRT in its devices for police use, given that it "could exacerbate existing inequities in policing, for example by penalising black or LGBTQ communities" [12, 16, 19]. Moreover, the American city of San Francisco has prohibited the use of facial recognition technology, given the potential abuses that could emanate from such use by law enforcement agencies and the intrusiveness of the technology on people's right to privacy [16]. These decisions can only be acclaimed and should be followed by both the private sector that develops such technology for law enforcement agencies and the law enforcement agencies by refraining from adopting such technology.

3.5 FRT, Drones and Privacy: An Incompatible Co-existence?

The recourse to drones undoubtedly endangers citizens' right to privacy [1, 2, 5, 20, 35]. The right to privacy is a fundamental right UK citizens enjoy based on the Human Rights Act 1998, which incorporates the European Convention on Human Rights into UK law.[7] More specifically, Article 8 of the ECHR provides a derogable right to respect someone's private and family life, home and correspondence. Contrary to absolute rights, this derogable right to privacy can be restricted when that restriction is provided by law, necessary in a democratic society and in the interest of one of the legitimate aims enumerated in Article 8(2) of the ECHR, such as national security or for the prevention of disorder or crime. If the interference by the state's recourse to drones with citizens' right to privacy fulfils the previously mentioned requirements, that interference will be justified [16].

This right to privacy covers a wide array of protected aspects of one's private life, including a person's physical and psychological integrity [16, 17] (Von Hannover v. Germany, 2012, para. 95). Both facets are relevant and endangered when law enforcement agencies use facial recognition technology in drones to surveil their citizens. The physical integrity of a person will be threatened by the constant potential visualisation of the person's body characteristics by the police. A person's psychological integrity will equally be endangered, given the sense of being watched continuously and the ensuing distress this can cause to the person. The European Court of Human Rights, which oversees the correct implementation of the ECHR, has ruled that a person's 'reasonable expectation of privacy' in public places has to be taken into

[7] Given that the Human Rights Act 1998 implements the European Convention on Human Rights, only this latter Council of Europe instrument will be discussed.

account, given that it is a significant yet not conclusive factor (Antovic and Mirkovic v. Montenegro, 2017, para. 43; Lopez Ribalda and others v. Spain, 2019, para. 88) [13, 16, 17, 23].

Although the European Court of Human Rights has not yet ruled on the issue of the potential infringement of citizens' right to privacy by law enforcement's recourse to (live) facial recognition technology in drones, similar and relevant case law exists [23]. As such, the Court ruled that while the mere monitoring of people in public spaces does not constitute a violation of the right to privacy, the recording or processing of these images might infringe this right (Peck v. the United Kingdom, 2003, para. 59; Perry v. the United Kingdom, 2003, para. 38). Given that live facial recognition technology in drones processes images of people by recording them and comparing them against external databases, it is difficult to conceive that the Court would conclude a non-violation of the right to privacy.

Contrary to the European Court of Human Rights, the British Courts have been brought to rule on the use of live FRT by the UK's law enforcement agencies in the case R (Bridges) v Chief Constable of South Wales Police [2020] EWCA Civ 1058. In this case, the Court of Appeal of Cardiff overturned the High Court's ruling in finding that, amongst others, the recourse by the South Wales police to live FRT in CCTV cameras did not have a sufficient legal basis (Ground 1 of the judgment), as required by Article 8(2) of the ECHR.[8] This infringement has been remedied by the update of the Surveillance Camera Code of Practice [47]. Disappointingly, the Court held that mass surveillance by the deployment of live FRT in CCTV cameras is justifiable in light of law enforcement's purposes in South Wales. Consequently, the Court found that the interference of such technology with citizens' right to privacy is not disproportionate (Ground 2 of the judgment). The judgment is, however, limited to the recourse to live FRT in cameras by the South Wales police and not the police at the national level. In agreeing with Zalnieriute [53] who considered deplorable that the Court did not analyse the greater danger generalised surveillance constitutes and the chilling effect this occasions on the state's citizens, this finding by the Court is highly regrettable given the extensive intrusion the deployment of live FRT in CCTV cameras for surveillance constitutes on citizens' right to privacy.

Whilst it is deplorable that the Court of Appeal did not find the widespread screening of citizens' faces by live FRT in CCTV *cameras* in the public sphere disproportionate in light of law enforcement's operations, it is questionable whether this reasoning would stand for the recourse to such technology in *drones*. As pointed out earlier, the recourse to drones is considerably more intrusive than the static use of CCTV. Moreover, given that live FRT constantly records and processes the biometric

[8] The other two arguments of the Court of Appeal were that the South Wales police had provided a lacking data protection impact assessment and that the police corps had failed to sufficiently take into account the disproportionate effect such technology would have on minorities and women, hence breaching its public service duty as required by the Equality Act 2010. The former argument pertains to data protection law and consequently lies outside of this work's scope. And although the inherent biases against minorities and women are equally important when discussing the deployment of live FRT in drones and have been discussed in Sect. 3.4, the latter argument relates to the factual analysis the Court makes in specific cases and is thus not relevant to this work.

features of persons in real-time, it is questionable how this permanent and intrusive surveillance could be considered necessary or proportionate in light of the right to privacy. Hence, the recourse by law enforcement to live FRT in drones for surveillance purposes can only be considered an unlawful infringement on citizens' right to privacy that significantly exceeds the interference of the recourse to CCTV cameras. Consequently, the recourse to this technology in drones by the police should be prohibited.

4 A Possible Solution: More Transparency and Explainability?

The algorithmic codes behind (live) facial recognition technology are often considered trade and company secrets that should not be made public. However, and agreeing with Zalnieriute [53] and the civil society organisation Access Now [1], these arguments against the transparency and explainability of AI behind FRT can be counter-argued. Firstly, transparency would allow third-party programmers to find errors, biases and vulnerabilities in the code that the original programmer is unaware of. Consequently, the original programmer could improve the code to eliminate the code's errors and biases and increase the programme's security by patching these vulnerabilities. Hence, the risk that drones are hacked would decrease. Secondly, whilst private companies could rely on the argument of the protection of trade secrets, this argument does not stand for the state, given that citizens do not have a choice but to be subjected to the state's decision to deploy such technology. Transparency is even more required when law enforcement agencies incorporate commercial AI into their drones. Given that law enforcement agents are not specialised in code-writing, it is problematic that they use a tool that they do not understand or even have full access to. The transparency requirement is all the more important when public authorities utilise AI that severely impacts citizens' lives and fundamental rights [1, 53]. Thirdly, if law enforcement would implement AI that is explainable to the citizens subjected to such technology, citizens' trust in the police's recourse to such technology would only increase. Whilst the revised Surveillance Camera Code of Practice provided a recommendation of transparency on behalf of the public authorities deploying such drones towards citizens, the drafters of the Code regrettably did not provide transparency requirements of the technology used in the drones. This insertion could have contributed to an increased trust of the population in this technology and an increased level of understanding by law enforcement agents of the tools they utilise.

Although more transparency and a higher level of explainability of the AI behind (live) facial recognition technology is undoubtedly required when law enforcement agencies use such technology in drones for surveillance purposes, the mere recourse to such technology has to be considered an unnecessary and disproportionate interference with citizens' fundamental right to privacy and should consequently be outlawed.

5 Conclusion

This work analysed several threats the deployment of (live) facial recognition technology in drones by law enforcement agencies for surveillance purposes constitutes. While drones are undoubtedly useful in search and rescue operations or active shooting situations, their increased mainstream use in combination with the invasive live FRT for generalised surveillance practices under the pretext of 'security' is highly worrisome in light of citizens' fundamental rights and, especially, their right to privacy. First, the recourse to such technology creates safety and security risks because the drones and technology the law enforcement authorities use are subject to hacking or default. Consequently, hackers could take over the control of the drones and exploit devices that appear legitimate (with, for example, the logo of law enforcement agencies) for malicious purposes. Moreover, the drones could, due to a defect, fall down onto people and cause severe damages. Second, the attribution of liability of drones and the technology deployed in them is highly questionable. Should the operator, the programmer or the drone be considered liable for incidents engaging drones? Moreover, when these drones are hacked, the question of attribution of incidents that would occur following this hack complicates the question even more given the difficult identification of hackers. Third, the recourse to facial recognition technology in drones worrisomely contributes to the never-ending normalisation of constant surveillance by the state of their citizens, in name of 'security'. This increased surveillance impinges on the citizens' right to privacy, and more generally on a number of human rights. By being aware of the constant surveillance of their movements, citizens will self-censor, which will refrain them from exercising to the fullest their other fundamental rights, such as the right to freedom of expression or free movement. This chilling effect of surveillance is highly worrisome. Last, facial recognition technology is, as any technology, prone to false results, errors and inherent biases, which constitute a real threat to citizens' right to privacy when people are wrongly qualified as criminals or to the population's general safety when criminals are mistakenly not recognised as such.

Consequently, the deployment by law enforcement authorities of facial recognition technology, and, especially *live* FRT, in drones is highly intrusive on the citizens' right to privacy and should therefore be prohibited. Facial recognition technology, drones and the right to privacy are incompatible. The update of the Surveillance Camera Code of Practice could have been the perfect opportunity for the state to grow into a society where citizens' privacy stands central. This Code could have prohibited the highly intrusive recourse by law enforcement to live FRT in drones for surveillance purposes. At a minimum, the Code could have provided transparency requirements of the technology used in the drones in order to increase trust of the population and increase the level of understanding of law enforcement agents of the tools they deploy. However, the government missed this opportunity to evolve into a society that highly regards its citizens' right to privacy under the overly-used pretext of ensuring their citizens' security.

References

1. Access Now (2018) Human rights in the age of artificial intelligence. https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf
2. Ahmad N (2020) Unmanned aircraft system (UAS) and right to privacy: an overview—part 1. Comput Telecommun Law Rev 26(6):153–160
3. Atkinson S, Carr G, Shaw C, Zargari S (2021) Drone Forensics: the impact and challenges. In: Digital forensic investigation of internet of things (IoT) devices. Springer, pp 65–124. https://web.p.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzI3MDQwNzVfX0FO0?sid=c721584f-ab67-40c9-a19e-19ddd6bb43d4@redis&vid=0&lpid=lp__C3_BE_C3_BF65&format=EB
4. Babuta A, Oswald M, Janjeva A (2020) Artificial intelligence and UK national security policy considerations [Occasional Paper]. Royal United Services Institute for Defence and Security Studies. https://static.rusi.org/ai_national_security_final_web_version.pdf
5. Bentley J (2019) Policing the police: balancing the right to privacy against the beneficial use of drone technology. Hastings Law J 70(1):249–296
6. Bradford B, Yesberg JA, Jackson J, Dawson P (2020) Live facial recognition: trust and legitimacy as predictors of public support for police use of new technology. Br J Criminol 60(6):1502–1522. https://doi.org/10.1093/bjc/azaa032
7. Bragias A, Hine K, Fleet R (2021) 'Only in our best interest, right?' public perceptions of police use of facial recognition technology. Police Pract Res 22(6):1637–1654. https://doi.org/10.1080/15614263.2021.1942873
8. Breshears AA (2016) Use of armed drones by domestic law enforcement: presence and the fourth reasonableness factor. Western Michigan University Thomas M. Cooley Law Rev 33(1):183
9. Chatterjee S, Sreenivasulu NS, Hussain Z (2022) Evolution of artificial intelligence and its impact on human rights: from socio legal perspective. Int J Law Manag 64(2):184–205
10. Civil Aviation Authority (2022) The drone and model aircraft code. Civil Aviation Authority. https://register-drones.caa.co.uk/
11. College of Policing (2022) Overt deployment of live facial recognition (LFR) technology to locate persons on a watchlist. College of Policing. https://www.college.police.uk/app/live-facial-recognition
12. Crawford K (2019) Halt the use of facial-recognition technology until it is regulated. Nature 572(7771):565. https://doi.org/10.1038/d41586-019-02514-7
13. Custers B (2016) Drones here, there and everywhere introduction and overview. In: Custers B (ed), The future of drone use, vol 27. T.M.C. Asser Press, pp 3–20. https://doi.org/10.1007/978-94-6265-132-6_1
14. Enemark C (2020) On the responsible use of armed drones: the prospective moral responsibilities of states. Int J Hum Rights 24(6):868–888. https://doi.org/10.1080/13642987.2019.1690464
15. Enemark C (2021) Armed drones and ethical policing: risk, perception, and the tele-present officer. Crim Justice Ethics 40(2):124–144. https://doi.org/10.1080/0731129X.2021.1943844
16. European Union Agency for Fundamental Rights (2020) Facial recognition technology: fundamental rights considerations in the context of law enforcement. Publications Office. https://data.europa.eu/doi/10.2811/524628
17. European Union Agency for Fundamental Rights (2020) Getting the future right: artificial intelligence and fundamental rights. Publications Office. https://data.europa.eu/doi/10.2811/774118
18. Feeney M (2016) Surveillance takes wing privacy in the age of police drones (No. 807). CATO Institute. https://www.cato.org/sites/cato.org/files/pubs/pdf/pa807_1.pdf
19. Feldstein S (2019) The global expansion of AI surveillance [Working Paper]. Carnegie endowment for international peace. https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final1.pdf

20. Finn RL, Wright D (2016) Privacy, data protection and ethics for civil drone practice: a survey of industry, regulators and civil society organisations. Comput Law Secur Rev 32(4):577–586. https://doi.org/10.1016/j.clsr.2016.05.010
21. Fussey P, Coaffee J, Ball K, Haggerty K, Lyon D (2012) Urban spaces of surveillance. In: Routledge handbook of surveillance studies. Routledge, pp 201–208. https://doi.org/10.4324/9780203814949
22. Fussey P, Davies B, Innes M (2021) 'Assisted' facial recognition and the reinvention of suspicion and discretion in digital policing. Br J Criminol 61(2):325–344. https://doi.org/10.1093/bjc/azaa068
23. Fussey P, Murray D (2019) Independent report on the London metropolitan police service's trial of live facial recognition technology (The human rights, big data and technology project). Human Rights Centre University of Essex. http://repository.essex.ac.uk/24946/1/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report-2.pdf
24. Gallagher S (2022) UK: surveillance eye in the sky: drone technology in criminal and regulatory investigations. Mondaq, 21 Feb. https://www.mondaq.com/uk/aviation/1163650/surveillance-eye-in-the-sky-drone-technology-in-criminal-and-regulatory-investigations
25. Ghaffary S (2019) How to avoid a dystopian future of facial recognition in law enforcement. Vox, 10 Dec. https://www.vox.com/recode/2019/12/10/20996085/ai-facial-recognition-police-law-enforcement-regulation
26. Greenwood F (2020) How to regulate police use of drones. Brookings, 24 Sep. https://www.brookings.edu/techstream/how-to-regulate-police-use-of-drones/
27. Guitton MJ (2021) Fighting the locusts: implementing military countermeasures against drones and drone swarms. Scand J Mil Stud 4(1):26–36. https://doi.org/10.31374/sjms.53
28. Haggerty K (2012) Surveillance, crime and the police. In: Routledge handbook of surveillance studies. Routledge, pp 235–243. https://doi.org/10.4324/9780203814949
29. Hartmann K, Giles K (2016) UAV exploitation: a new domain for cyber power. In: 2016 8th international conference on cyber conflict (CyCon), pp 205–221. https://doi.org/10.1109/CYCON.2016.7529436
30. Hayward KJ, Maas MM (2021) Artificial intelligence and crime: a primer for criminologists. Crime Media Cult Int J 17(2):209–233. https://doi.org/10.1177/1741659020917434
31. Heyns C (2016) Human rights and the use of autonomous weapons systems (AWS) during domestic law enforcement. Hum Rights Q 38(2):350–378
32. Information Commissioner's Office (2019) The use of live facial recognition technology by law enforcement in public places (No. 2019/01). https://ico.org.uk/media/about-the-ico/documents/2616184/live-frt-law-enforcement-opinion-20191031.pdf
33. Jones, R. (2017). Visual surveillance technologies. In: McGuire MR, Holt TJ (eds) The Routledge handbook of technology, crime and justice. Routledge, pp 436–450. https://doi.org/10.4324/9781315743981
34. Jørgensen RF (ed) (2019) Human rights in the age of platforms. The MIT Press.
35. Khan MA, Safi EA, Alvi BA, Khan IU (2018) Drones for good in smart cities: a review, pp 1–7. https://www.researchgate.net/profile/Muhammad-Khan-716/publication/316846331_Drones_for_Good_in_Smart_CitiesA_Review/links/5a27c404aca2727dd883c881/Drones-for-Good-in-Smart-CitiesA-Review.pdf
36. Klauser F, Pedrozo S (2015) Power and space in the drone age: a literature review and politico-geographical research agenda. Geographica Helvetica 70(4):285–293. https://doi.org/10.5194/gh-70-285-2015
37. Kroener I, Neyland D (2012) New technologies, security and surveillance. In: Ball K, Haggerty K, Lyon D (eds) Routledge handbook of surveillance studies. Routledge, pp 141–148. https://www.taylorfrancis.com/books/9781136711077
38. Loukinas P (2022) Drones for border surveillance: multipurpose use, uncertainty and challenges at EU borders. Geopolitics 27(1):89–112. https://doi.org/10.1080/14650045.2021.1929182
39. Osoba OA, Welser IV W (2017) The risks of artificial intelligence to security and the future of work. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/perspectives/PE200/PE237/RAND_PE237.pdf

40. Pyzynski M, Balcerzak T (2021) Cybersecurity of the unmanned aircraft system (UAS). J Intell Rob Syst 102(2):35. https://doi.org/10.1007/s10846-021-01399-x
41. Samanth SKVP, Balachandra M (2022) Security in internet of drones: a comprehensive review. Cogent Eng 9(1):2029080. https://doi.org/10.1080/23311916.2022.2029080
42. Stelmack K (2015) Weaponized police drones and their effect on police use of force. Pittsburgh J Technol Law Policy 15(2):276–292. https://doi.org/10.5195/TLP.2015.172
43. Straub J (2014) Unmanned aerial systems: consideration of the use of force for law enforcement applications. Technol Soc 39:100–109. https://doi.org/10.1016/j.techsoc.2013.12.004
44. Surden H (2014) Machine learning and law. Washington Law Rev 89(1):87–115
45. Tarr J-A, Thompson M, Tarr A (2019) Regulation, risk and insurance of drones: an urgent global accountability imperative. J Bus Law 8:559–576
46. Tumbarska A (2018) Remotely controlled non-lethal weapon systems in the context of law enforcement. Secur Futur 3:102–105
47. UK Biometrics and Surveillance Camera Commissioner (2022) Update to surveillance camera code of practice, 3 March 2022. https://www.gov.uk/government/publications/update-to-surveillance-camera-code
48. United Nations General Assembly (2017) Res 34/7 the right to privacy in the digital age (22 March 2017). A/HRC/34/7
49. Urquhart L, Miranda D (2021) Policing faces: the present and future of intelligent facial surveillance. Inf Commun Technol Law, 1–26. https://doi.org/10.1080/13600834.2021.1994220
50. Vattapparamban E, Guvenc I, Yurekli AI, Akkaya K, Uluagac S (2016) Drones for smart cities: issues in cybersecurity, privacy, and public safety. In: 2016 international wireless communications and mobile computing conference (IWCMC) 2016:216–221. https://doi.org/10.1109/IWCMC.2016.7577060
51. Watney M (2022) Ethical and legal aspects pertaining to law enforcement use of drones. Int Conf Cyber Warf Secur 17(1):358–365. https://doi.org/10.34190/iccws.17.1.27
52. Yu S, Carroll F (2021) Implications of AI in national security: understanding the security issues and ethical challenges. In: Montasari R, Jahankhani H (eds), Artificial intelligence in cyber security: impact and implications. Springer International Publishing, pp 157–175. https://doi.org/10.1007/978-3-030-88040-8_6
53. Zalnieriute M (2021) Burning bridges: the automated facial recognition technology and public space surveillance in the modern state. Sci Technol Law Rev 22(2):284–307. https://doi.org/10.52214/stlr.v22i2.8666
54. Zegart A (2020) Cheap fights, credible threats: the future of armed drones and coercion. J Strateg Stud 43(1):6–46. https://doi.org/10.1080/01402390.2018.1439747

The Use of the Internet for Terrorist Purposes: Investigating the Growth of Online Terrorism and Extremism

Zainab Al-Sabahi and Reza Montasari

Abstract The twenty-first century's development of the Internet has led to a significant shift in contemporary communication (Zelin in The state of global jihad online: a qualitative, guantitative, and cross-lingual analysis. New America Foundation, 2013). While the Internet is widely utilised in everyday life to distribute and share information, it has also created an environment in which virtual societies have become a breeding ground for new risks and threats (Hawdon et al. in NORDICOM 3:29–37, 2015). Consequently, terrorist and extremist organisations exploit the accessibility of the Internet to facilitate their violent activities and spread their extremist ideology (Montasari et al. in Privacy, Security and Forensics in the Internet of Things (IoT). Springer International Publishing AG, 2022). This chapter aims to analyse the impact of the Internet on the rise of extremism and terrorism. To this end, the chapter will first investigate the role of the Internet in promoting the online radicalisation process, which leads to participation in terrorist acts. It will then critically examine how the Internet alters the nature of violent extremism, including its fatal consequences in the real world. Finally, the chapter will explore how and for what purposes violent extremists use the Internet, focusing on recruitment through propaganda, training and planning by sharing information, psychological warfare and fundraising.

Keywords Internet · Terrorism · Extremism · Radicalisation · Online terrorism · Warfare propaganda · Online radicalisation · Terrorist groups · Psychological warfare · Violent extremism

Z. Al-Sabahi (✉) · R. Montasari
Department of Criminology, Sociology and Social Policy, School of Social Sciences, Swansea University, Singleton Park, Swansea SA2 8PP, UK
e-mail: kayanalsabahi@gmail.com

R. Montasari
e-mail: Reza.Montasari@Swansea.ac.uk
URL: http://www.swansea.ac.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_4

1 Introduction

The twenty-first century's incredible development of the Internet has led to a significant shift in contemporary communication [1]. While the Internet is widely utilised in everyday life to distribute and share information, it has also created an environment in that virtual societies have become a breeding ground for new risks and threats [2]. Consequently, terrorist and extremist organisations exploit the accessibility of the Internet to facilitate their violent activities and spread their extremist ideology [3]. Extremist ideology refers to a set of texts, including videos and pictures, with solid ideological beliefs identifying who belongs to the in-group or the out-group [4]. It is essential to recognise that these ideologies are hard to prevent entirely,instead, they can be diverted. In addition, the Internet facilitates no violence or terrorism without the ideology of extremism. Violent extremism is defined as people's beliefs and attitudes that promote ideology-driven violence, whereas terrorism is described as the violent act itself to achieve ideological, political and religious intention [4]. Winter et al. [5] describe extremist groups as "early adopters" of the Internet, which has historically been strongly linked to violent and nonviolent online extremism.

It is critical to note that extremism is not an inherently violent phenomenon because not all extremist groups intend to execute violent and harmful acts [6]. Additionally, not all violent activities are motivated by negative extremist beliefs. This makes online extremism a complex and contested situation without a detailed description [3]. It is important to emphasise that online violent extremism is similar to any idea alien to society, constituting anxiety with fear of engaging in it. Thus, there is an ongoing debate questioning the role of the Internet in influencing contemporary violent extremism and terrorism, a prominent issue that needs to be considered (3, 2017) [6].

By critically analysing the existing literature, this chapter aims to analyse the impact of the Internet on the rise of extremism and terrorism. To this end, the chapter will first investigate the role of the Internet in promoting the online radicalisation process, which leads to participation in terrorist acts. It will then critically examine how the Internet alters the nature of violent extremism, including its fatal consequences in the real world. Furthermore, the chapter will explore how and for what purposes violent extremists use the Internet, focusing on recruitment through propaganda, training and planning by sharing information, psychological warfare and fundraising.

The remainder of this chapter is structured as follow: Sect. 2 provides a background into the role of the Internet in promoting the online radicalisation process. Section 3 examines the impact of the Internet on violent extremism while Sect. 4 investigates the use of the Internet by violent extremists. Finally, the chapter is concluded in Sect. 5.

2 Internet and the Online Radicalisation Process

The Internet's system in exhibiting, choosing, connecting, and distributing information was identified as a concern in the context of extremist ideologies based on the user's activity [7]. Thus, the Internet is described by many academics as a tool that gives more alternatives for extremists and terrorist groups to disseminate their activities and engage with people in a way that has never been possible before [8, 9]. To illustrate, the Internet has significantly impacted extremist organisations and has changed the space in which they operate by creating a new subculture to spare their ideology [10]. However, what distinguishes terrorists from other users in using the online environment is online radicalisation. It is a social process through which people become interested in ideologies and adopt extremist beliefs and approaches by utilising the Internet, particularly social media, with substantial real-world consequences [3, 11].

Individuals and organisations utilise a variety of online platforms, including internet communities, websites, news groups, chat rooms, blogs, video games, file archives and web rings, to spread extremist and hate content [2]. Online violent extremist content and hate messages are shaped by societal and religious intolerance, political ideologies and the activities of terrorist groups [11]. The level of the Internet that used to encourage terrorism and radicalisation has been categorised as a global threat, "one of the greatest threats that countries, including the UK face" [12], p. 2). According to a recent study, 61% of terrorists in the UK are actively using the Internet to further their radicalisation or terrorist action [9]. It is significant to determine terminologies that have developed to facilitate the internet impacts of implementing extremist ideologies in online radicalisation to understand the consequences of these manners. For example, [13] pointed out a 'virtual community' concept created by terrorist groups through exploiting capabilities of communication and technological features. Virtual community allows violent beliefs to be shared freely without fear of being labelled anti-extremist to identify those who might be readily affected and lured into extremist activities [13]. Virtual communities also can generate a sense of belonging that attracts vulnerable individuals or those who lack strong social ties within their community [3]. However, it is essential to note that everybody is vulnerable to being a victim of online extremism because of the manipulative nature of the content [14].

During which people excessively spend time in virtual communities, the Internet will act as an echo chamber to reinforce radical views and extremist identities to isolate individuals socially [7, 13, 14]. Echo chambers are an online environment where individuals congregate with those who share similar views, encourage each other's arguments and amplify thoughts of violent voices [15]. As a result, people develop a distorted perception of reality in which involvement in radical beliefs and violence is no longer seen as harmful but acceptable, normalised and desirable [7]. Additionally, radicalisation facilitated by echo chambers encourages copycat attacks and the merger of a different kind of ideology. A notable example of this is the case of Andrew 'Isa Ibrahim'a 20-year young vulnerable individual who was marginalised

in his society. In 2008, Ibrahim was sentenced to 10 years in custody after being found guilty of planning a suicide bombing at a Bristol Shopping Centre using an explosive devices vest [16]. It was believed that Ibrahim was radicalised online after accessing extreme online content and spending time playing the role of a terrorist in video games [16]. Eventually, it was clear that Ibrahim had been influenced online by violent games, contributing to his offending act.

Moreover, while people have always gravitated towards thoughts and individuals who coincide with their own beliefs to the detriment of others, there has been much focus on the phenomenon which [17] termed a 'filter bubble' [11]. A filter bubble is an Internet process in which consumers are continually subjected to the same type of content based on previous Internet participation [17]. As an explanation, the Internet tools filter restricts what users can access, creating a personalized bubble of content based on their interests [3]. However, the use of filtering systems itself might be crucial in increasing the radicalisation of individuals to violent extremism [11]. For instance, Internet platforms act as a conduit for polarisation and radicalisation because their proposal systems advance information associated with extremely harmful attitudes, particularly violent extremist content [18]. Although there is still a gap in understanding the impact of such extensive processes on radicalisation, their significance and potential for widespread generalisation are undeniable [11].

Nonetheless, some contemporary and historical literature has argued that the Internet's influence on radicalisation is insufficient, minimal and therefore exaggerated [19, 20]. Conway [21] emphasises that extensive research has not yet confirmed that consuming and interacting with violent extremist internet content has any consistent links to online radicalisation, participation in violent extremism, or support for terrorism. This is because most of the research has focused on extremist content online rather than how interaction with such content affects the process of radicalisation [22]. Furthermore, considering the high use of the Internet by terrorists, several academics are still debating whether the Internet can replace physical interactions and whether online networks exert a similar impact on people as real-world social networks [9, 15]. Despite the evidence demonstrating that the Internet is playing an increasingly important role in radicalisation processes, it cannot be argued that the online setting is simply replacing the offline setting, considering offline effects were present to some level for the majority of sentenced extremists in the data set [23]. For example, [24] found in his investigation that the majority of terrorist actors were more likely to impact by social exchange and external offline factors such as friends and family. Similarly, those who only largely radicalised online were less likely to be socially linked to risk in the context of terrorist acts and related offending compared to those who have offline radicalised [23, 25].

However, [3] state that although the Internet is a crucial tool and cause of violent radicalisation, it is unusual for someone to become radicalised only online. Rather, online and offline influence factors shape the process of becoming a violent extremist and they tend to complement each other [26]. In the same way that the Internet has transformed people's lives by shifting communication with other individuals, it seems to be providing a more significant role in radicalising individuals, leading to increased extremist and terrorist offending.

3 The Impact of the Internet on Violent Extremism

Extremist and terrorist groups have increasingly shifted their focus from traditional media such as newspapers, television and radio to the Internet. Simi and Futtrell [27] state that communities interlinked by the Internet improve how people can connect to groups by lowering barriers of time and geography, allowing for a large volume of information to be transferred and promoting user solidarity. The Islamic terrorist group Al-Qaeda is identified as the first extremism movement that transfer from physical space to an online environment [10]. Therefore, Islamic groups have been strongly linked to online extremism and terrorism in much literature [4]. Contrary to popular belief, extremism involves not only Islam religion and not all Muslims have extremist goals [28]. Al-Qaeda's transformation is considered a significant development in the history of terrorism [29]. "With laptops and DVDs, in secret hideouts and at neighbourhood Internet cafes, young code-writing jihadists have sought to replicate the training, communication, planning and preaching facilities they lost in Afghanistan with countless new locations on the Internet" [10], p. 7]. The reason behind this significant transformation is the nature of the online environment that offers certain unique advantages for terrorists [29]. Tsfati and Weimann [30] claim that the Internet has enabled isolated extremist groups to establish a global network, enabling them to become transnational rather than limited to particular countries. To illustrate, the Internet has an extensive global reach and audience, which enables terrorist organisations to convey their messages of hate in a more effective, accessible, faster and cheaper way as compared to traditional media [31].

As well, anonymity is a crucial feature of the Internet which can enable individuals to access things they might not have access to in the real world and feel less restrained [32]. Christopherson [33] claims that individuals and groups on social media are protected online when they are anonymous. This, in turn, gives space to the spread of online violent extremism and makes it more challenging for authorities to monitor their activities [32]. Furthermore, terrorist groups have complete control over a new type of media [14]. They have the ability to set up their own websites and social media pages, giving them control over the posted and shared content with minimum risk [19]. An Internet search conducted in 2006 discovered there were over 4,800 violent extremist websites supporting terrorism [34]. In that regard, it appears that extremist and terrorist websites affected actual terrorist operations that were both direct and more widespread [29].

4 The Use of the Internet by Violent Extremists

Not only has the number of terrorist websites expanded, but terrorists' Internet usage has also increased [35]. There are several ways in which modern terrorists and extremists use the Internet to encourage and support terrorist acts that illustrate the impact of

the digital world space. Weimann [34] identifies eight different, sometimes overlapping ways that contemporary terrorist groups use the Internet to further their political and ideological aims online, including propaganda to recruitment, sharing information, planning attacks, psychological warfare and fundraising. Holt et al [36] note that extremist groups have quickly created marketing strategies online to expand their engagement in offending behaviours. Extremist organisations use the Internet to actively recruit new members by disseminating propaganda to support and join their movements [14]. Terrorist propaganda usually occurs through technological means that provide ideological or practical guidance, justifications, explanations, or advocacy of terrorist activities, which shape public perceptions and mobilise support [37]. Such propaganda may include images, blogs, websites, email and messaging services, audio files, videos, online publications (such as magazines and newsletters) and video games created by terrorist and extremist groups [37].

It is important to note that online terrorist propaganda is everywhere because online extremism is vast and not limited to one specific type of thing or platform. The propaganda strategies utilised by extremist and terrorist organisations differ from one group to another, but there are some common propaganda strategies [14]. One of the most powerful strategies used by terrorists is exploiting feelings to convince potential recruits. They create content such as language, texts, images and videos that appeal to the emotions of vulnerable individuals [37]. This content normalises and creates positive messages about their activities. According to a study byGaudette et al. [8], emotive language and imagery are used by extremist groups to promote a sense of community and belonging. Many extremist groups employ violent imagery, statements, and pictures to generate a strong emotional response from their audience. Thomas [38] claims this tactic can turn "fence sitters" into promoters for the effects. This refers to intentionally using misleading and biased information, including a mix of true and false statements and emotional appeals targeting the audience's hidden prejudices [37]. Also, Terrorists' online emotional materials promote a sense of victimisation among individuals and present violence as a justifiable means of fighting against aggression [30]. According toTodorovic and Trifunovic [37], advocating violence is common in terrorist propaganda. Similarly, Weimann [34] states that terrorist groups justified their acts by saying that violence was the last solution when all other options had been explored. By generating an emotional response, recruitment improves lead people to join extremist organisations and influences them to engage in violence without fully understanding the consequences of their choice [39].

Once individuals show interest in emotional materials, they can interact with particular individuals through online communities such as chat rooms. This makes members feel as if they are part of and belong to that community without being aware that they are being used to further the extremist group's objectives and coordinate their activities [31]. Most experts agree that terrorist groups such as Al-Qaeda realised the chatrooms' potential early due to their ability to attract and reach prominent members globally [40]. Electronic jihad is a large part of what al-Qaeda's jihadist movement tries to do, using chat rooms as clandestine recruitment to incite users from around the world and attempt to recruit them to join the jihad [29]. Chatrooms enabled terrorist

groups to interact quickly and privately without needing face-to-face meetings or sending real mail [40]. This participation enables groups to identify new followers and change the website's pitch to improve its attractiveness [40]. A further feature of chatrooms for terrorist groups like Al-Qaeda was the possibility to communicate using aliases and code words, making it difficult to monitor who was communicating [38]. This made it less challenging for Al-Qaeda to operate in hidden spaces and avoid being identified by law enforcement agencies [29].

Another notable example of a jihadist movement is the Islamic State of Iraq and Syria (ISIS), which has made significant investments in social media (such as Facebook, YouTube and Twitter) and propaganda camping to attract sympathisers and create a new generation of cyber jihadists [14, 31]. Increased use of social media platforms means more users are vulnerable to recruitment by extremist and terrorist groups [31]. ISIS's recruiting platforms provide affordable and accessible opportunities for individuals, especially young people, to engage in nefarious terrorist activities, seen as an appealing environment for them [2, 41]. This environment is a place to learn about terrorist groups, support them, and participate in direct activities to achieve their goals [34]. However, Rogan (2006, as cited in [31] argues that the Internet cannot replace in-person recruitment but rather as a tool for communication among individuals to express their concerns and reduce involvement in radicalisation. In contrast, what drove this new type of recruit to involve in extremist and terrorist causes was not poverty or religion but rather a sense of marginalisation and a search for meaning in society [42].

Powell et al. [43] argue that marginalised individuals are more likely to use the Internet for terrorist activities. Therefore, the Internet has evolved into a haven for potential extremists to 'groom' individuals with vulnerabilities [14]. An example of a high-profit case involved recruiting three young girls, including Shamima Begum, from London in 2015. These young girls were convicted of travelling to Syria, joining ISIS militants, and being brides of ISIS fighters [44]. It is believed that ISIS took advantage of Begum and her friends' teenage immaturity and recruited them through social media [44]. This case demonstrates ISIS's use of the Internet to attract vulnerable persons, mainly young individuals looking for a sense of significance and belonging in society. Thus, the Internet has become a virtual space for bringing together and mobilising these marginalised individuals [28]. Subsequently, the need for members to engage in criminal activities makes recruitment crucial for terrorists and extremists.

Moreover, the Internet has opened new avenues for terrorist and extremist groups to share information and exploit new opportunities [35]. Even though this information was available before the Internet, it is now more efficient and accessible. Weimann [34] claims that the Internet has become an online learning space for terrorists, classified as an online terrorism university. In other words, the Internet has facilitated terrorist and extremist groups' ability to share and access sensitive or dangerous information that can be used as a training field for terrorist attacks [37]. A study by [9] suggests that 14% of offenders choose to engage in violence after viewing things online. Through the Internet, individuals can access detailed instructions from terrorist manuals, which provide data on bombing making, weaponry, suicide vests,

specific methods, and information about potential targets to carry out an attack [35]. This was shown in the 2007 terrorist attack on Glasgow airport, perpetrated by individuals who had downloaded bomb-making manuals online [9]. Thus, online terrorist and extremist training are designed to prepare recruits for roles as activists and operatives to commit actions of terrorism [29].

In addition to online learning about how to make bombs, contemporary terrorist organisations utilise the Internet as a critical resource for planning and coordinating specific attacks [21]. These organisations use publicly available information and applications (such as Google Earth) to gather data about security measures, find potential techniques and purchase materials on the Internet, enabling them to congregate attacks more effectively globally [37]. For example, the 9/11 attack was confidentially orchestrated and planned on the Internet by Al-Qaeda [35]. Planning these acts allows terrorists to gain the acceptance they require to achieve their ideological goals [21]. Despite law enforcement agencies' efforts to monitor terrorists' planning process, recently, terrorist and extremist groups have moved into encrypted communication platforms such as WhatsApp and Telegram to plan their operations [45]. Encrypted platforms allow terrorists to communicate with each other in a way that enhances security and privacy between them [37]. For example, ISIS uses Telegram's private chat to create promotional materials and plan their attack, allowing them to send messages which immediately disappear as opened [31]. As a result, law enforcement agencies cannot see and understand how terrorists and extremists engage in these platforms [45]. This also makes protecting individuals by intercepting and interpreting their messages more challenging. Hence, it is crucial to acknowledge that extremism and terrorism are progressing alongside modern internet technologies.

Another way that the Internet has often prompted extremism and terrorism is by enabling groups to engage in psychological warfare. This refers to exploiting terrorist propaganda to make threatening tactics that aim to create anxiety, fear or panic, intimidation, hate and division to meet their objectives [37]. To achieve this, terrorists and extremists post and share horrific materials among users on the Internet, including videos of beheadings, suicide bombings, and mass killings among their opponents and audiences [35]. Lachow and Richardson [46] claim that the Internet facilitates psychological warfare by allowing more terrorist organisations to appear online and spread fear globally, contrasted to traditional modes of communication. This is demonstrated by the most recent action, including the awful killing of American journalist Daniel Pearl. Terrorist groups in Pakistan kidnapped him, and his execution by beheading was recorded and uploaded to thousands of terrorist websites [35]. These actions affect individuals in society, they feel they are under constant threat of attack [47]. Although the Internet is an unregulated media that disseminates news, images, threats, or messages regardless of their veracity or potential consequence, It is effectively suited for allowing even a small group to amplify its message and raise its significance and threat [35]. As mentioned earlier, propaganda plays a significant role in psychological warfare tactics, given its capacity to manipulate and influence a broader target audience. This audience provides extremist groups with feedback, which in turn helps them frame issues in a way that serves their interests [31]. This is because the primary goal of these groups is to influence public

opinion by raising suspicion and confusion to undermine the security and legitimacy of governments [46]. As a result, decision-makers give in to extremists' demands, which provides complete control to achieve their goals.

The Internet also can be used by extremist and terrorist organisations as a means of fundraising, an effort to fund their acts of terrorism [5]. Money is a way of continuity of terrorism, it is "the engine of the armed struggle" ([48], p. 1). Terrorist organisations use different websites or social media to solicit donations from sympathisers using illegal means [21]. For instance, Al Qaeda has relied significantly on donations and its global fundraising network by setting up fraudulent businesses that look like charities to collect money online [35]. Since funding for terrorism is illegal in many places, terrorist groups have developed their money-laundering skills [29]. These groups utilise anonymous digital payment methods such as crypto currency (including Bitcoins) and crowd funding to receive fund transactions from supporters [37]. They also use the dark web to sell illegal services and goods, including weapons and drugs, to create revenue [47]. What helps to facilitate the movement of these funds is the anonymity offered by the Internet which allows these groups to hide their identities and receive payments from anywhere in the world [35]. Digital payment methods also allow terrorist to hide their transaction and avoid detection, making it difficult for law enforcement agencies to follow their financial transactions [21].

5 Conclusion

To conclude, this paper raises critical issues concerning the Internet's impact on extremism and terrorism. The advent of the Internet has made communication, information gathering, and publishing more convenient and faster in recent times. Although the Internet has provided numerous benefits, it has also become a breeding ground for illegal activities such as violent extremism, which presents a significant challenge for law enforcement agencies. This refers to the social process of radicalisation when people become interested in extremist ideologies through the Internet, which leads to involvement in terrorist acts. Extremists amplify thoughts of violent voices in echo chambers. The filter bubble makes this possible by repeatedly exposing individuals to the same violent content previously shared online. Moreover, the Internet has provided terrorists with certain advantages, including global reach, anonymity, and complete control over new media. As a result, terrorist and extremist groups find the Internet a facilitating tool as it helps them achieve their political and ideological objectives. It is clear that terrorist groups have found different ways to use the Internet to carry out their operations with minimal risk, as evidenced by the explored cases. For example, terrorist propaganda and disinformation have attracted many international recruits to extremist virtual communities. Also, the ability to share information has made the Internet a training and planning space for terrorist operations. The Internet has also enabled terrorist groups to engage in psychological warfare to spread fear among their audience and enemies. In addition, terrorist groups need funding to carry out their operations, and they often use the Internet

to acquire financial support for their activities. To sum up, the Internet's impact on extremist purposes that lead to terrorism is a phenomenon alien to society and requires everyone to beware of it.

References

1. Zelin AY (2013) The state of global jihad online: a qualitative, guantitative, and cross-lingual analysis, 1st edn. New America Foundation
2. Hawdon J, Oksanen A, Räsänen P (2015) Online extremism and online hate. Exposure among adolescents and young adults in four nations. NORDICOM 3(4):29–37
3. Montasari R, Carroll F, Mitchell I, Hara S, Bolton-King R (2022) Privacy, security and forensics in the internet of things (IoT). Springer, International Publishing AG, Berlin
4. Berger JM (2018) Extremism 1st edn. Mit Press
5. Winter C, Neumann P, Meleagrou-Hitchens A, Ranstorp M, Vidino L, Fürst J (2020) Online extremism: research trends in internet activism, radicalization, and counter-strategies. Int J Confl Violence 14(2):1–20
6. Sullivan A, Montasari R (2022) The Use of the Internet and the internet of things in modern terrorism and violent extremism. In: Montasari R, Carroll F, Mitchell I, Hara S, Bolton-King R (eds) Privacy, security and forensics in the internet of things (IoT), 1st edn. Springer, Cham, pp 151–165. https://doi.org/10.1007/978-3-030-91218-5_7
7. Neumann PR (2013) Options and strategies for countering online radicalization in the United States. Stud Conflict Terrorism 36(6):431–459
8. Gaudette T, Scrivens R, Venkatesh V (2022) The role of the internet in facilitating violent extremism: insights from former right-wing extremists. Terrorism Polit Violence 34(7):1339–1356
9. Gill P, Corner E, Conway M, Thornton A, Bloom M, Horgan J (2017) Terrorist use of the Internet by the numbers: quantifying behaviors, patterns, and processes. Criminol Public Policy 16(1):99–117
10. Coll S, Glasser SG (2005) Terrorist turn in the web as base of operations [Electronic Version]. Washington Post. Retrieved June 18, 2007 from http://msl1.mit.edu/furdlog/docs/washpost/2005-08-07_washpost_www_weapon_01.pdf
11. Binder JF, Kenyon J (2022) Terrorism and the internet: how dangerous is online radicalization? Front Psychol 6639
12. UK House of Commons Home Affairs Committee (2017) Radicalisation: the counter narrative and identifying the tipping point, Eighth report of session 2016–17 (HC 135). https://www.parliament.uk/documents/commons-committees/homeaffairs/Correspondence-17-19/Radicalisation-the-counter-narrative-and-identifying-thetipping-point-government-response-Eighth-Report-26-17-Cm-9555.pdf
13. Bowman-Grieve L (2013) A psychological perspective on virtual communities supporting terrorist and extremist ideologies as a tool for recruitment. Secur Inf 2(1):1–5. https://doi.org/10.1186/2190-8532-2-9
14. Awan I (2017) Cyber-extremism: Isis and the power of social media. Society 54(2):138–149. https://doi.org/10.1007/s12115-017-0114-0
15. Von Behr I, Reding A, Edwards C, Gribbon L (2013) Radicalization in the digital era: the use of the internet in 15 cases of terrorism and extremism. RAND corp 47. https://doi.org/10.7249/RR453
16. BBC News (2018) Bristol suicide-vest terrorist Isa Ibrahim denied parole. BBC News. https://www.bbc.co.uk/news/uk-england-bristol-46404028
17. Pariser E (2011) The filter bubble: what the internet is hiding from you, 1st edn.Penguin, London

18. Whittaker J, Looney S, Reed A, Votta F (2021) Recommender systems and the amplification of extremist content. Int Policy Rev 10(2):1–29
19. Conway M (2017) Determining the role of the internet in violent extremism and terrorism: six suggestions for progressing research. Stud Conflict Terrorism 40(1):77–98
20. Laqueur, W. (1999). The new terrorism: Fanaticism and the arms of mass destruction, 1st edn. Oxford University Press
21. Conway M (2006) Terrorism and the internet: new media—new threat? Parliam Aff 59(2):283–298
22. Mølmen GN, Ravndal JA (2021) Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists. Behav Sci Terrorism Polit Aggression 1–25
23. Kenyon J, Binder J, Baker-Beall C (2021) Exploring the role of the Internet in radicalization and offending of convicted extremists. Ministry of Justice Anal Ser
24. Whittaker J (2020) Online echo chambers and violent extremism. In: the digital age, cyber space, and social media: the challenges of security and radicalization, 1st edn. Institute for Policy, Advocacy, and Governance, pp 129–150
25. Hamid N, Ariza C (2022) Offline versus online radicalisation: which is the bigger threat? Glob Net Extremism Technol
26. Valentini D, Lorusso AM, Stephan A (2020) Onlife extremism: dynamic integration of digital and physical spaces in radicalization. Front Psychol 11(1):524. https://doi.org/10.3389/fpsyg.2020.00524
27. Simi P, Futrell R (2006) White power cyberculture: building a movement. Public Eye 20(1):7–12
28. Torok R (2013) Developing an explanatory model for the process of online radicalisation and terrorism. Secur Inf 2(1):1–10
29. Rudner M (2017) Electronic Jihad: The internet as Al Qaeda's catalyst for global terror. Stud Conflict Terrorism 40(1):10–23
30. Tsfati Y, Weimann G (2002) : Terror on the internet. Stud Conflict Terrorism 25(5):317–332. www.terrorism.com
31. Piazza JA, Guler A (2021) The online caliphate: Internet usage and ISIS support in the Arab world. Terrorism polit violence 33(6):1256–1275. https://doi.org/10.1080/09546553.2019.1606801
32. Pasculli L (2020) The global causes of cybercrime and state responsibilities. Towards an integrated interdisciplinary theory. J Ethics Leg Technol 2(1):48–74
33. Christopherson KM (2007) The positive and negative implications of anonymity in internet social interactions: on the internet, nobody knows you're a dog. Comput Hum Behav 23(6):3038–3056
34. Weimann G (2006) Terror on the internet: the new arena, the new challenges, 1st edn. US Institute of Peace Press
35. Weimann, G (2004)www.terror.net: how modern terrorism uses the Internet 1st edn. United States Institute of Peace. www.terror.net
36. Holt T, Freilich JD, Chermak S, McCauley C (2015) Political radicalization on the internet: extremist content, government control, and the power of victim and jihad videos. Dyn Asymmetric Conflict 8(2):107–120
37. Todorovic B, Trifunovic D (2020). Prevention of (Ab-) Use of the internet for terrorist plotting and related purposes 1st edn. International centre for counter-terrorism (ICCT)
38. Thomas TL (2003) Al Qaeda and the internet: the danger of cyberplanning. Foreign military studies office (ARMY) Fort Leavenworth Ks 33(1):113–123
39. Freiburger T, Crane JS (2008) A systematic examination of terrorist use of the internet. Int J Cyber Criminol 2(1):309–319
40. Weimann G (2010) Terror on Facebook, Twitter, and Youtube. Brown J World Affairs 16(2):45–54
41. Zanini M, Edwards SJ (2001) The networking of terror in the information age. In: Arquilla J, Ronfeldt D (eds) Networks and netwars: the future of terror, crime, and militancy, 1st edn. RAND. pp 29–60

42. Sageman M (2008) The next generation of terror. Foreign policy 16 (5):37
43. Powell B, Carsen J, Crumley B, Walt V, Gibson H, Gerlin A (2005) Generation Jihad. Time 16(6):56–59
44. Baker J, Lee J (2023) Shamima Begum accepts she joined a terror group. BBC News. https://www.bbc.co.uk/news/uk-64222463
45. Fernandez M, Alani H (2021) Artificial intelligence and online extremism: Challenges and opportunities. In: McDaniel JL, Pease K (eds) Predictive policing and artificial intelligence, 1st edn. Routledge, London, pp 132–162. https://doi.org/10.4324/9780429265365-7
46. Lachow I, Richardson C (2007) Terrorist use of the internet: the real story. Jt Force Q 4(5):100–103. National Defense University, Washington DC
47. Butler G, Montasari R (2023) The impact of the internet on terrorism and violent extremism. In: Jahankhani H (eds) Cybersecurity in the age of smart societies, 1st edn. Springer, Cham. https://doi.org/10.1007/978-3-031-20160-8_24
48. Napoleoni L (2004) Money and terrorism. Strateg Insights 3(4):47–50

Cyber-Security and the Changing Landscape of Critical National Infrastructure: State and Non-state Cyber-Attacks on Organisations, Systems and Services

Joseph Rees and Christopher J. Rees

Abstract The main aim of this chapter is to identify and explore key issues relating to cyber-attacks on critical national infrastructure. The chapter commences by clarifying the term critical national infrastructure. It then proceeds to highlight the rise in international incidents of cyber-attacks on critical national infrastructure. Vignette case studies, drawn from countries such Australia, USA, Ukraine and the UK are integrated into the analysis for illustrative purposes. The chapter emphasises the need for more attention to be placed on the vulnerabilities of critical national infrastructure in the light of trends such as the convergence of Information Technology and Operational Technology systems and the increasing use of Internet of Things (IoT) devices as a means of bringing systems online. Further, the chapter draws attention to the relatively low entry cost of engaging in cyber-attacks using malware, in contrast to the relatively high cost and logistical complexity of mounting physical attacks on well protected critical national infrastructure sites. One of the main conclusions drawn from the analysis is the extent to which addressing vulnerabilities in critical national infrastructure cyber-systems is likely to involve a wide range of actors, such as State-level emergency planners, manufacturers of IoT devices, and white hat hackers.

Keywords Critical national infrastructure · Cyber-attacks · State sponsored · Proxies · Operational Technology · Information technology · Internet of Things · Industrial Internet of Things · Ukraine · United Kingdom · United States · Ransomware · Prevent · IT · OT · CNI · IoT · IIoT · Cyber warfare

J. Rees (✉)
Swansea, UK
e-mail: joedrees@protonmail.com

C. J. Rees
Global Development Institute, University of Manchester, Manchester, UK
e-mail: chris.rees@manchester.ac.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_5

1 Introduction

Over recent decades, the security of critical infrastructure has been frequently highlighted as an issue of key international concern [18, 22, 121] with governments specifically identifying the critical nature of infrastructure security. In the 1990s, the fear of a 'Digital Pearl Harbor' [110] concerned many states and became a talking point in the field of security. Following 9/11 and the war on terror, governments across the world increased their security expenditure to bolster the security of critical national infrastructure [5]. Indeed, Ellis [34] notes that around one third of all Federal Homeland Security funding relates to infrastructure protection within the United States of America (USA). Yet, it could be argued that, for a period, the discussion surrounding the importance of the security of critical national infrastructure had been neglected. In recent years, however, an increase in the number of cyber-attacks, including several high-profile attacks on critical infrastructure, has led to renewed interest in this subject [68] and [89]. This chapter commences by clarifying the term critical national infrastructure. It then explores the continuing rise in incidents of cyber-attacks including attacks on critical national infrastructure and highlighting the potential outcomes of insecure critical infrastructure. Vignette case studies, drawn from countries such Australia, USA, Ukraine and the UK are integrated into the discussion for illustrative purposes. Finally, a discussion is presented which identifies elements of current protection systems including frameworks, policies and legislation.

2 Definitions of Critical National Infrastructure

In order to understand the importance of the protection of critical national infrastructure first the term itself must be examined. Organisations such as the Organisation for Economic Co-operation and Development [86] define the 'critical' aspect of national infrastructure as being the element of infrastructure that: "… *provides life sustaining and essential services required for the economic and social well-being of citizens, national and public security and key government functions.*" (Fjäder [41], p. 125). At this preliminary stage of the discussion, the phrase 'life sustaining' is notable as it provides an early sign that critical infrastructure includes systems and facilities, such as health, energy and water, which, when disrupted, will endanger the day to day lives and possibly very existence of relatively large groups of people. From a legislative perspective the USA Patriot Act of 2001 [21] (42 U. S. C. 5195c (e)) defines critical infrastructure as "*systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters.*". Again, a notable aspect of this definition is that, in addition to humanitarian considerations, it describes infrastructure with reference to 'economic security'. In a similar vein, following 9/11, during the formation of the Department of Homeland Security in the USA, it was

decided that the department would be responsible for co-ordinating policy relating to the protection of critical national infrastructure. The Department of Homeland Security defines critical national infrastructure as consisting of: "*the physical and cyber systems and assets that are so vital to the United States that their incapacity or destruction would have a debilitating impact on our physical or economic security or public health or safety. The nation's critical infrastructure provides the essential services that underpin American society.*" [27].

These definitions are extremely helpful in providing overviews of the nature of critical national infrastructure. In addition, former USA President Obama provided a detailed list of the services which are encompassed within the term 'critical national infrastructure'. The Presidential Policy Directive 21 (PPD-21) ([122], p. 1) defines 16 critical infrastructure sectors namely: "*chemical, commercial facilities, communications, critical manufacturing, dams, defense industrial base, emergency services, energy, financial services, food and agriculture, government facilities, health and public health, information technology, nuclear reactors, materials and waste, transportation systems, water and wastewater systems*". Academics have concurred with the inclusion of this wide array of sectors by stating that many supply chains of products such as food are included within the term [102]. Importantly, it is clear from the list of sectors, that not all systems are physical/tangible assets, namely 'financial sectors'.

In relation to the deliberate disruption of these sectors, IBM states that cyber-attacks are an "*unwelcome attempts to steal, expose, alter, disable or destroy information through unauthorized access to computer systems.*" [58]. Thus, when utilising the term cyber-attack, it becomes evident that, in light of the preceding discussion, a wide range of sectors may be possible targets and further, that there is an expanse of other possible variables such as the duration of attack (or downtime of systems) and varying classifications of severity [61]. In the case of severity, the picture becomes nuanced further as the significance of cyber-attacks does not necessarily lie solely in the severity of the attack but, also on other considerations such as, for example, the attack type or method which may signify emerging threats in the landscape [85]. This perspective has been evidenced with the suggestion for various models of taxonomy for cyber-attacks [107].

3 Attacks on Critical National Infrastructure

The security of critical infrastructure encompasses both online and offline systems. It is within the scope of this chapter to consider cyber elements of infrastructure as opposed to physical aspects such as target hardening. Bronk and Conklin [14] note that the rise in attacks on critical national infrastructure in recent years has changed from a "hypothetical concern" to a real one, especially in countries such as the USA. Australia's annual report from the CISC ([17], p. 6) outlines a variety of

ways in which critical national infrastructure can be impacted, threats "*emanate from inside or outside an organisation and range from hostile or criminal activity, foreign interference, terrorism and natural disasters through to poor physical, personnel and cyber security practices*". Interestingly, also out of Australia, the annual Cyber Threat Report from the [1] noted that around one quarter of incidents reported to the organisation were related to critical infrastructure or essential services. Similarly, reports such as the 2022 Microsoft Digital Defence Report [77] describe a dramatic increase in attacks on infrastructure. In particular the report suggests that national states, including Russia, China and Iran, are backing cyber-attacks in other national contexts. It should be noted that although Microsoft's report specifically refers to attacks on infrastructure, it has been suggested that the actual number of attacks may much higher than the report indicates. This is due to the difficulty of detection, [125, 129] particularly in developing countries or those with fewer cyber resources and also in cases where detailed information about attacks is not placed in the public domain as a result of national security considerations. For example, in late 2020 revelations of the SolarWinds hack came to light. This hack of a USA-based software company enabled hackers to access the systems of Solarwinds' clients including USA government agencies. When discussing the breach before a select committee hearing on intelligence, Microsoft's President stated before the USA Senate that the attack was "*the largest and most sophisticated sort of operation we have seen*" [90]. What is highly pertinent to this discussion is that, although the attack penetrated and gained unauthorised access to sensitive information of several federal agencies and over a hundred private businesses, the breach remained undetected for months [56]. Notably, Microsoft has referred to this attack as 'NOBELIUM' and described it as: "the most successful nation-state attack in history" [76].

4 The Importance of Protecting Infrastructure

A multitude of organisations and academics have analysed cyber-attacks at a macro level in order to understand further the scale of the problem and attack trends [43]. In 2018, the Center for Strategic and International Studies (CSIS) partnered with McAfee and reported that the annual economic loss due to cyber-crime is estimated to be in the region of $600 billion [23]. Calculating exact figures for cyber-crime is incredibly difficult, as not all crimes are reported, and crimes such as intellectual property theft are very difficult to evaluate [43]. These difficulties in measurement often lead to varying monetary estimations of cyber-crime. From a statistical perspective the Office for National Statistics in the UK have calculated that Computer Misuse grew 89% in 2022 compared to 2020 [87]. It is data such as these that confirm the need for academic research into the prevalence of cyber-crime and ways to improve defences against cyber-attacks [105].

Cyber-attacks across society have a multitude of implications outside of monetary, such as identity theft [116] and data breaches [7]. As noted above, however, cyber-attacks on critical national infrastructure have, by definition, the potential to

be catastrophic and life threatening—and security is therefore paramount [80, 81]. Indeed, the recent conflict in Ukraine has highlighted the extent to which cyber-attacks on critical infrastructure can damage a nation [16]. Even aside from the direct impact of such a cyber-attack, it is noted here that the public display of a crippling cyber-attack on critical national infrastructure can also have serious ramifications for social unrest by stimulating behaviours such as panic buying and stockpiling [67]. To mitigate risks and prepare for potential problems, models and calculations have been utilised to predict possible outcomes of cyber-attacks on critical national infrastructure systems [19]. Indeed, even 'pinprick' cyber-attacks on critical national infrastructure can be detrimental [65].

Although cyber security is a rapidly evolving subject and the threat landscape frequently changes and evolves, reviewing trends over time is a useful method to assess current security and potentially prevent future attacks on critical national infrastructure [78]. Indeed, the sophistication of cyber-attacks has advanced over time. For example, while [66] may have been accurate when writing that: "*cyber weapons seem to be of limited value in attacking national power or intimidating citizens*", it would be difficult to come to the same conclusion today. Similarly, Rid [98] wrote of the contention surrounding the concept of cyber war and even questioned the likelihood that it will ever occur. Yet, more recently, Bērziņš [10] suggests that recent events in Ukraine and Syria appear to display the use of 'hybrid warfare' in which cyber elements are used to strike economic targets. Unlike other cyber-attacks, disruption to critical national infrastructure has the potential to cause a "high consequence event" [30]. Therefore, it could be argued that a 'cyber war' is not required in order to have potentially catastrophic effects, as even an isolated event could lead to profound consequences at various societal levels. In 2021 the Texas Power Crisis, caused by extreme weather, resulted in over four million people losing water and power for nearly a week and ultimately led to loss of life [63]. This event, although caused by extreme weather, rather than a cyber-attack, highlights the importance of the preservation and protection of critical national infrastructure using, among other resources, cyber security.

To add further difficulty to the threat landscape of critical national infrastructure, the ubiquitous nature of cyber space, as a global realm inherently not bound by physical location, enables attacks to originate from anywhere on earth and beyond [71]. In relation to changing landscapes, emerging globalisation has become associated with a convergence of digital systems and structures as well as mindsets across societies [73]. This convergence has profound implications for cyber-security due to the emergence of global systems, actors, and threats. In addition, as there is no proximal requirement for attackers, the ability to pivot between targets is further enhanced. Attackers can move between targets and even nations with relative ease and at a moment's notice. Bad actors attacking critical national infrastructure range from lone individuals and hacktivist groups to an advanced persistent threats such as a nation state [70]. Currently, there are difficulties dealing with the diplomacy and potential retaliation of a cyber-attack by a nation state. In 2022, the Speaker of the House of the United States, Nancy Pelosi, travelled to Taiwan on a sensitive diplomatic visit. The trip reportedly enraged Beijing and resulted in a number of threats

of attacks and reportedly cyber-attacks towards Taiwan. This incident highlights the potential for cyber-attacks to be used as a tool of diplomatic expression of ability, in this case in combination with the military drills which also took place at the time [60]. This specific example also highlights the problems of jurisdiction of cyber-attacks and the difficulties that are likely to be encountered when attempting to apprehend cyber-attackers especially in cases where there is direct or indirect involvement of nation States [117].

In a discussion of this nature, it is also important to emphasise the extent to which the actual design of cyber-systems may lead to vulnerabilities to cyber-attacks. For example, in 2007, an early experiment which analysed how cyber could impact electric grid components such as diesel generators was conducted by the Idaho National Laboratory (INL). The team managed to adjust the timing of several circuit breakers within the generator, using a computer, which caused irreparable damage to the generator. This became known as the Aurora Vulnerability [120]. Although the results of the experiment remain contested, the experiment is considered to be an early demonstration of how cyber capabilities could cause irreparable damage, that is, have a kinetic effect in the physical realm with a specific focus on industrial control systems. Indeed, industrial control systems consisting of Supervisory Control and Data Acquisition (SCADA) and hardware elements such as Programmable Logic Controllers (PLCs) which were originally built for internal monitoring, reliability, improved productivity, and safety, now face a new challenge, namely security [126].

The Convergence of Operational Technology (OT) and Information Technology (IT)

One specific systems issue which has been raised in relation to cyber-attacks on critical national infrastructure is the convergence between OT and IT systems. While acknowledging the debates surrounding the definitions of OT and IT, and even classification of the tasks each complete, IT Systems refer to the "*any use of computers, storage, networking and other physical devices, processes to create, process, store, secure and exchange all forms of electronic data*" ([45], p201). In contrast OT systems refer to the "*associated specific functions like manufacturing and industrial environments, which include industrial control systems such as supervisory control and data acquisition (SCADA)"* ([45], p.201). Although some scholars have differentiated PLCs and SCADA, it is assumed here that PLCs are part of the SCADA system.

Traditionally, many organisations have maintained relatively discrete OT and IT systems. Industrial control systems utilise architecture models such as the Perdue Model to further protect OT systems by using network segmentation and network boundary controls within these layers. In the past, organisations such as the National Institute of Standards and Technology [82] were advising against the use of 'security through obscurity' alone. More recently, when highlighting factors relating to the inherent insecurity of OT, one recent discussion pointed out that: "*So far, OT-targeted cyber-attacks are considered so complex that only sophisticated teams with significant technical and organisational resources are likely to succeed*" [6]. This issue is highly relevant to the security of critical infrastructure as, traditionally, there was an expectation that the very use of proprietary bespoke software formed a layer of security against attacks. Nevertheless, it must be noted that, in contrast to models

such as security by design, there are existent arguments for combining of 'security through obscurity' with other security models as a means of cyber deception for cyber defence [4].

Connection of Critical Infrastructure to the Internet

As noted above, IT and OT have traditionally been 'siloed' from one another, though frequently there are still essential 'trusted' communications between the IT and OT layers [2]. In the past IT evolution was considered separate from the evolution of OT. Yet recent discussions have noted the convergence of IT and OT [79] particularly with the increase in the remote management of systems and demand-smart analytics [127]. These features have been made feasible by the widespread implementation of ethernet, WI-FI and TCP/IP as a means to access IT and OT systems [103]. The use of WI-FI technology in particular has increased in the sector due to it being a fast and cost-effective way of getting systems online for remote access and maintenance [83]. This is an interesting and significant trend in infrastructure systems because, as some academics have noted, the IT systems in infrastructure were frequently updated yet OT systems including components such as PLC's had lifespans of over 10 years [93]. This longevity has been attributed to cost and the need for components and systems to remain online with little or no possibility to be offline [69], even though this online status poses risks to security.

The widespread implementation of networks into industrial control systems is clearly evident in the use of Internet of Things (IoT) devices. Notably, the evolution of the use of IoT devices can be seen in both IT and OT spaces, with a rapid rate of implementation [46]. Although this implementation has met many demands for modern industrial control systems the large-scale implementation of IoT devices has security implications [28]. These security implications range from their potential use in attacks in domestic settings [96] to their potential use in attacks on critical infrastructure [75]. The Industrial Internet of Things (IIoT) has been revolutionary in OT spaces. However, IIoT now poses unique security challenges. These challenges have been described as different to security challenges traditionally associated with IT systems [104], particularly as many SCADA systems are dated and do not necessarily have security as a core design feature. Recent research has argued that, in order to ensure security, utilising hardware and software beyond its supported lifespan needs to cease [69].

The permanent connection to the internet and/or the cloud, of systems that were previously commonly isolated secure environments has security implications. Even the implementation of remote cyber security services requires the establishment of an internet connection for third party vendors to install software patches. This was highlighted in the 2020 SolarWinds hack [91]. As noted above, the SolarWinds breach penetrated a multitude of organisations including elements of the United Kingdom and United States Governments, NATO and Microsoft [125]. After gaining access to the SolarWinds infrastructure the hackers utilised this to send remote access tools and controls within the software updates that were being sent to customers. Methods such as this potentially provide easier means of accessing critical infrastructure as

the security of software vendors, for example, may be less difficult to breach than accessing critical infrastructure directly.

5 Further Examples of Cyber-Attacks on Critical Infrastructure

Reviewing past incidents, both recent and historic, of cyber-attacks on critical infrastructure can be a useful way of furthering understanding the nature and impact of these attacks and, in addition, to provide insights as to what could be done in future to prevent such events reoccurring. Hence, at this juncture of the chapter, we introduce vignettes which provide additional examples of cyber-attacks on critical infrastructure. Two of the examples presented below are of Ransomware attacks which occurred, in short succession, within the USA in 2021. Both of these attacks involved an actor gaining unauthorised access to a system, then proceeding to encrypt all of the system files and demanding a 'ransom' in order to decrypt the files, thus using encryption as a "weapon" ([47], p. 22). If the victim fails to comply the files, including those sensitive in nature, are either deleted or in some instances leaked onto online forums [112, 113, 129]. In recent years these attack types have become prolific in the cyber-crime community. Once in a system, a perpetrator can unleash Ransomware with relative ease [53]. Access to systems can be accomplished in a variety of ways but in many instances result from spear phishing campaigns [112, 113]. Perpetrators use this method to gain credentials in order to then move vertically throughout the system utilising privilege escalation in order to gain administrative rights. These types of attacks have grown in notoriety and popularity due to their ease of implementation once within a system, the financial gain of these attacks, and their widespread availability as a result of the commoditisation of malware [129].

Vignette One: Ukraine's Power Grid

In 2015, Ukraine was bombarded with cyber-attacks during the initial conflict in the Donbass region. In 2015 Ukraine's power grid suffered a substantial cyber-attack reportedly instigated by the group Sandworm. Sources have repeatedly stated that this came from within Russia, and instigated by the group Sandworm. The synchronised attack caused major power outages across multiple regions leaving hundreds of thousands of Ukrainians without power. Arguably, this critical infrastructure attack on a Powergrid was the first of its kind [64]. The software utilised in the attack specifically targeted SCADA proprietary software. This aspect of the attack differentiates it from many other cyber-attacks which primarily target IT systems. It has been reported that the mode of attack involved phishing of power plant employees emails. There has been speculation that, due to the fact that much of the dilapidated Soviet infrastructure had been updated with Russian systems, attackers had prior knowledge of both the system and the proprietary software, further attesting to the need to refrain from relying on obscurity as a means of defense. Whatever the precise mechanisms

involved, Sullivan and Kamensky ([111], p. 31) referred to the attack as "brilliantly executed" and contend that aspects of the attack including the "attack methodology, tactics, techniques and procedures" could potentially be deployed elsewhere around the globe.

Vignette Two: Petroleum Factory in Saudi Arabia

Although not widely publicised at the time, accounts have emerged of a cyber-attack which reportedly targeted a petroleum factory in Saudi Arabia in 2017. The attack is said to have targeted the factory's Triconex technology [106]. The malware used, known as Triton, has been described by [29] as the first of its kind, primarily because it was a cyber-attack which specifically focused on the Safety Instrumented System (SIS) of an industrial plant. By all accounts, this was a significant hack involved accessing the physical controls of the industrial control systems.

Vignette Three: Iran Nuclear Facility

Although not believed to have been caused by the connection of critical infrastructure systems to the internet, reports have emerged of the use of Stuxnet, a computer worm, to irreparably damaged centrifuges in an Iranian nuclear facility [20]. This case provides a stark example of the potential damage a cyber-attack can have on SCADA systems [100]. Indeed, the attack caused significant concern in the cyber community. Although the worm had targeted the Iranian nuclear facility, the case has raised concerns that the malware could be adapted to physically degrade SCADA systems in other countries, such as the UK and USA [49, 57]. The damage to the nuclear facility was reported to be significant particularly as the worm was able to "wreak physical damage" ([26], p. 673), thus confirming fears set out by experiments such as the Aurora Vulnerability. As in the case of Vignette One, the attack has been attributed to the actions of several nation states [48]. The worm also fuelled debates surrounding cyber war as Iran's nuclear program was impeded by the worm at the cost of less than one military jet, and without the loss of life [39].

Vignette Four: Colonial Pipeline

On 7th May 2021, the Colonial Pipeline which transports refined fuel from the Gulf Coast to the East Coast of the USA was attacked [109]. The group DarkSide group are said to have claimed responsibility for the attack [74]. Reports of the attack indicate that the actual controls of the pipeline were not directly affected. Rather, it has been reported that Ransomware was uploaded to the IT system which impacted accessibility to systems such as financial accounts software. As a result, the company was unable to bill customers accurately and therefore shut down operations [54]. Although the security of the pipeline Operational Technology (OT) system does not appear to have been attacked, the case underlines the interconnectedness of other aspects of the system, and how an attack on one aspect of the operation impacted upon the operation as a whole. The pipeline was offline for a total of five days. It has been reported that Colonial Pipeline paid a ransom to the hackers in Bitcoin. More recently, the US Government has claimed that the majority of the ransom paid has been recovered [95]. As this pipeline was a major distributor of refined fuel, the effects of this shutdown were felt throughout the USA, particularly on the East Coast.

Interestingly, this case study also highlights the psychological effects of a critical infrastructure attack. Aside from the fact that fuel supply to the east coast had been limited during the shutdown, causing an increase in fuel prices [114], individuals also began hording fuel, thus heightening the problem of a fuel shortage [54]. It has been claimed that the attack occurred even after "years of warnings" about the cyber security vulnerabilities of the USA power grid [108]. (Reeder and Hall [95], p. 16) describe this case as "a Pearl Harbor moment for cybersecurity" as it represented a significant attack on the Eastern Seaboards economy and emphasises the need for further "*protection and prevention, resilience and recovery, and deterrence*" of critical infrastructure.

Vignette Five: JBS and Supply Chain Security

In May 2021, JBS SA, a Brazilian based meat processing company suffered a cyber-attack. Although based in Brazil, JBS S.A is reportedly the largest meat supplier in the world, operating in the USA, Canada and Australia. The cyber-attack was reportedly a Ransomware attack, in which JBS paid $11 million dollars in Bitcoin as ransom [51]. Unlike the Colonial Pipeline case, the ransom does not appear to have been recovered. In a similar vein to the Colonial Pipeline attack, a foreign hacking group, in this case the Russian group 'Ransomware Evil' (REvil) was reportedly responsible for the attack [31]. Following the incident, President Biden held a press conference stating that he expects the Russian government to act if given intelligence pointing towards the illegal cyber-attack having been instigated in Russia. Shortly after the incident and a conversation between President Biden and Putin, the REvil servers went offline and have not resurfaced since. Further, the FSB published a video of a raid supposedly relating to REvil arrests. In a White House Press Conference following the release of the video, a White House correspondent stated that one of the individuals arrested in the raid was also responsible for the Colonial Pipeline attack. Europol arrest warrants from other group members, mainly residing in Eastern Europe, were later unsealed by the FBI. The significance of the interruption of a critical infrastructure supply chain must be noted [31]. In this particular case this was attributed to the consolidation of certain aspects of the meat supply chain, particularly case meat packing which brings into question wider aspects of security relating to global food security and monopolising industries.

Vignette Six: Cyber-Attack on Royal Mail, UK

In January 2023, media reports emerged of a cyber-incident on the UK Royal Mail organisation. This incident was confirmed by the Royal Mail and the UK's Cyber Security Centre which indicated that it was working alongside the National Crime Agency to investigate what had happened. The incident resulted in customers unable to make payments or print of postage labels at home. At the time of writing in January 2023, media reports were categorising the incident as a cyber-attack which they were attributing to a cyber-gang called Lockbit [24]. Their Ransomware, 'Lockbit black', encrypts software with payment for lifting of the encryption being demanded in cryptocurrency. It was reported that the attack led Royal Mail to suspend its international delivery service, thus affecting over 250,000 parcels and letters. Returning

to the preceding discussion on definitions of critical national infrastructure, the case is particularly helpful in emphasising the wide-range of sectors, including national postal services, which provide critical national infrastructure services. In the years preceding this attack, the Royal Mail's national network had been included in the UK Government's: "*Critical National Infrastructure framework, as a sub-sector of the Communications sector*" ([55], p. 2).

6 Current Government Policies and Protective Systems

In the light of the above cases and the impact of these cyber-attacks at national level, it is apparent that governments around the world have become increasingly aware of the potential for cyber-attacks to disrupt societies and threaten life. Thus, when discussing the security of critical national infrastructure, protective systems and legislation must be discussed. It is not within the scope of this chapter to seek to offer a wide-ranging and in-depth analysis of legislative measures put into place by various governments. Rather, at a general level, we note that governments have introduced legislation to cover cyber-attacks even though the likelihood of cyber-crimes reaching prosecution is relatively low [11]. Furthermore, these legislative measures, introduced via policy and law, frequently appear to have a regulatory approach as a method of vulnerability reduction [44, 65].

There is a multitude of legislative measures across the globe that aim to improve cyber-security and reduce the likelihood of cyber-attacks. For example, in November 2022 the European Union accepted the NIS2 Directive [37] (EU) 2022/2555. The legislation is likely to come into effect across the EU's member states by 2024. This directive is intended to fill the gaps left from the initial NIS directive [36] ((EU) 2016/1148) and builds upon aspects of the European Programme for Critical Infrastructure Protection (EPCIP). Although the directive covers a broad range of topics regarding cyber security, the directives proposal specifically covers seven key sectors: "energy, transport, banking, financial market infrastructures, healthcare, drinking water supply and distribution, and digital infrastructures" ([38], p. 2). Interestingly this policy applies to both private and public companies which are operating on the defined 'essential services' (OES). Each must conduct a cyber assessment and abide by certain security measures. This perhaps works towards addressing problems identified by [119] who have discussed the difficulty in the accountability of cyber security, stating that, although the security of critical infrastructure is a public reasonability, it often becomes a private industry task. Søby (2020) concurs, arguing that the private sector organisations are integral to critical infrastructure and therefore the security implications.

Other legislative measures appear to target critical national infrastructure protection in more specific ways. In the USA, President Biden's recent Executive Order (14028) [124] appears to signify a push from the Federal level to move towards the use of Zero Trust Architecture. The Executive Order states that measures such as guaranteed encryption between devices, as part of a Zero Trust Architecture Model,

will help strengthen national security by means of improving system security. These measures are mandated to companies wishing to sell to the Federal government. Perhaps an expansion of this Zero Trust Architecture into guidelines already in place, such as the North American Electric Reliability Corporation Critical Infrastructure Protection (NERC CIP) standards will be seen in time.

Another example of legislative intervention emanates from Europa which has recently proposed the Cyber Resilience Act ([35] p. 1). The proposed Act is designed to increase the security of hardware and software products by introducing "*a regulation on cybersecurity requirements for products with digital elements*". This is an interesting development surrounding policy as it targets the regulation of products themselves and holds manufacturers responsible for the security of their products, in an attempt to protect users. It could be argued that there may be limitations surrounding this initiative as credential phishing is frequently the entry point for cyber-attacks. In January 2022, the White House specifically listed phishing as problematic and stated that the use of multi-factor authentication would be required in certain aspects of IT infrastructure [128]. Further, scholars have noted the importance of a continued effort towards better cyber-hygiene and education in the workplace as part of a security process [84]. Ensuring security by design alongside security measures and upkeep would provide a holistic approach in the hope for more secure critical national infrastructure.

Arguably, while the application of the currently proposed Cyber Resilience Act to critical national infrastructure may be limited, similarities can be drawn with now revoked Presidential Executive Order 13920 [123]. The Executive Order emerged after concerns were raised surrounding the potential existence of hardware backdoors in components such as transformers used in critical national infrastructure. This was evidenced in 2019 when a large industrial transformer destined for use in critical national infrastructure was seized by USA Department of Energy and the Department of Homeland Security for security reasons. This concern has also been seen elsewhere in the world, for example, the cessation of the Huawei's 5G network deal in the UK [42].

In recognising that ransomware payments are being made by organisations to groups potentially associated with terrorism, questions do emerge as to whether the victims of cyber-attacks may be exposing themselves to accusations of criminal activity by making such payments. That is, one must question the legality of paying a ransom to an organisation which may be linked directly or indirectly to terrorism. For example, in the UK, the payment of a ransom may leave organisations and individuals potentially open to prosecution under legislation such as s15(3) and s17 of the Terrorism Act 2000 ("TACT") [115] which prohibits the funding of listed groups. Attacks such as the JBS and Colonial Pipeline involved substantial payments made to criminal enterprises. It is reasonable to speculate that, in time, paying Ransomware fees may be prohibited under law, particularly in Western nations which explicitly state they do not negotiate with terrorists; rather, that an increasing onus may be placed on organisations to protect themselves from cyber-attacks. With terrorism an

ever-increasing threat, legislation such as this will likely continue to be discussed [97].

Legislation and the Detection of Cyber-Attacks

The detection of cyber-attacks is an essential aspect of cyber system security [127] though it is a truism that no system is impenetrable. Arguably, the implementation of aspects of the legislative measures discussed above will help to facilitate detection of cyber-attacks. For example, mandating the reporting of data leaks and attacks enables information to be communicated rapidly in order to reduce the likelihood of the attack being deployed elsewhere. Nevertheless, the detection of cyber-attacks is likely to be increasingly important in the defence of critical national infrastructure due to the changing threat landscape. In addition to ensuring that organisations maintain certain levels of security, legislation is also a useful means to ensure communication between the private sector and government agencies responsible for security. For example, the Security of Critical Infrastructure Act 2018 in Australia provides prevention regulations but also a post-incident framework [17]. Aspects of this Act are designed to ensure that regional and perhaps global efforts to patch vulnerabilities and create solutions to cyber-attacks can be conducted in a timely manner. For example, information reported under the auspices of the Act can be disseminated to software providers or other IT professionals in order to ensure the same problems do not continue to arise in other contexts. Similarly, bipartisan legislation such as the recent Infrastructure Investment and Jobs Act (2021) is playing a role in the continuation of funding for current projects and the creation of new funding for cyber security projects specifically regarding infrastructure. Thus, this funding for digital and cyber-security projects is likely to promote research, skill training and employment opportunities specifically relating to cyber-security and hopefully lead to, for example, the development of models designed to aid in the detection of cyber-attacks [62, 101].

Jurisdiction

The issue of jurisdiction arises when considering cyber-attacks on critical national infrastructure [13]. As many cyber-attacks take place across national borders, there have been calls for national law enforcement agencies to work together to decide under whose jurisdiction the perpetrators will be prosecuted [3]. If a State wishes to apprehend an alleged instigator of a cyber-attack who is residing outside of its jurisdiction extraterritorial jurisdiction can be sought. However, situations can arise where different States are seeking to prosecute the same individual leading to an impasse. Similarly, a State may not have an extradition treaty with the country in which the wanted person is residing. The problems surrounding these types of cases become all the more vexatious if an individual suspected of a cyber-attack was in fact working as a proxy for the State in which they are currently residing. Clearly, if an individual or group is potentially acting on the State's behalf or with its agreement to carry out a cyber-attack in another country, it is not in the interest of the State which instigated the attack to aid law enforcement, particularly if no extradition agreement is in place with the country which was attacked. Notably, law enforcement agencies

such as Europol have gone as far as waiting until individuals have left a nation in order to make arrests within an allied nation [92].

Nevertheless, it should be noted that international cooperation does occur even between nations of contention. As previously discussed, the arrest of members of the ReEvil group, who were reportedly responsible for cyber-attacks, involved cooperation between Russia and the USA. International efforts to curb cyber-crime are certainly in place as seen by the Convention on Cybercrime (also known as the Budapest Convention) which came into effect 2004. In situations where cooperation between States is not taking place, Braw and Brown ([12], p. 48) propose a '*personalised deterrence*' strategy, that is, a strategy involving: "… governments of targeted countries communicating directly to individual cyber attackers their intent to hold them personally responsible through denial of benefits and use of criminal law".

7 State Sponsored Attacks

While initiatives, such as the legislation discussed above, provide evidence that governments have sought to minimise the likelihood and prevalence of cyber-attacks on their countries' critical national infrastructure, there is also emerging evidence that countries themselves sponsor cyber-attacks as a means to promote their own interests [88]. For example, the Council on Foreign Relations (CFR) Cyber Operations Tracker [25] estimate that over 500 state sponsored cyber-attacks have occurred since 2005. Of these 77% were sponsored by either Russia, China, Iran or North Korea. It is highlighted that other States, while instigating cyber-attacks are, nevertheless, developing their capabilities to engage in this activity. Hermann [52] notes that, although States may not necessarily be engaging in cyber-attacks, the development of their capabilities to launch such attacks ensures, if need be, that they have the ability to spy on or directly damage potentially adversarial nation state systems at a relatively low financial cost [59]. For example, [50] contend that: "*While the Stuxnet attack against Iran was quite sophisticated, it does not necessarily require a strong industrial base or a well-financed operation to find ICS vulnerabilities*". Further, it appears that State involvement with cyber-attacks is only likely to intensify; nation States across the globe are heavily investing in their cyber military departments in what appears to be a 'cyber race' relating to activity such as espionage. In recent years, cyber-espionage has become increasingly prevalent. Cyber-espionage includes Computer Network Exploitation attacks, in which there is no intent to cause damage to a system but rather gather information from it. Data gathered by the Council on Foreign Relations (CFR) Cyber Operations Tracker [25] found that, when reviewing incidents of State sponsored cyber incidents in 2019, most were espionage.

Nation states are drawn to considering the use cyber-attacks for a variety of reasons, from espionage to modern day tactics of hybrid forms of war [57] in which States attack, both physically and in cyber-space, political adversaries and entities which are geographically distant [19]. For example, when exploring State-sponsored

cyber-attacks, the issue of the relative anonymity offered by the internet and therefore the potential for plausible deniability, is a key consideration. This is particularly evident in cases where State-influenced proxies are involved. That is: "*By outsourcing to proxies, this logic goes, a host government can plausibly deny its involvement in operations that advance its military and foreign policy aims*" ([15], p. 1). Thus, while media reports of State-involvement in cyber-attacks (including some of the examples discussed earlier in this chapter) are ubiquitous, it is extremely rare for a State to claim responsibility for a cyber-attack.

When examining the involvement of States in cyber-attacks, it is important to raise two issues relating to resources and ownership. First, the resources available to many nation States to research, plan and implement cyber-attacks are likely to far exceed the resources available to cyber-criminal organisations. As such, the increasing involvement of State actors in cyber-attack research and development, if unchecked, is likely to lead to the creation of capabilities which, if harnessed by non-State actors, will have the potential to create more prolific and devastating cyber-attacks. For example, the WannaCry Ransomware, derived from the Petya/NotPetya malware, is reported to have been based on hacker tools developed by the National Security Agency (NSA) in the USA [121].

Second, nation states are not motivated by traditional cost–benefit analyses associated with other actors such as Ransomware-for-hire organisations. The corollary of this observation is that, in countries operating under democratic principles, the allocation of resources to developing cyber-attack capabilities requires explanation and accountability. It leads to the need for an overarching State and international narrative to explain why resources are being devoted by States to developing this type of capability. Under the assumption that nation states will continue to conduct cyber-attacks, [110] discusses the formation of a position on the international stage that deters cyber-attacks due to fear of response and retaliation. Straub ([110], p. 12) proceeds to identify the impact of national policies as a means to avoid the creation of "adversarial relationships" (p. 12) and therefore minimise the risk of State actors instigating cyber-attacks. Arguably, the recent high profile use of cyber-attacks on critical national infrastructure by States, in conflicts such as those in Syria and Ukraine, is likely to lead to proliferation of their use [64].

Third, having considered the involvement of the State in cyber-security, it is important to highlight the issue of ownership of infrastructure in relation to protecting infrastructure against cyber-attacks. Referring specifically to the USA, Warfield ([118], p. 133) points out that: "*The combination of private sectors' voluntary status for information sharing, their ownership of the infrastructures on which governments depend, and the fact that, at least in the United States, the private sectors' CIN* [critical infrastructures] *is not under government jurisdiction and control represents the daunting problem facing information security of critical infrastructures*". Thus, while the protection of critical national infrastructure is necessary for the very existence of a State and the well-being of its population, private sector ownership of infrastructure is a key concern relating to cyber-attacks. This has led writers such as [99] to argue that argue that the scope of corporate social reasonability (CSR) should be extended to issues such as national security and the security of critical

national infrastructure. Yet the issue of resources cannot be overlooked as there are difficulties surrounding the onus on private businesses to ensure a high standard of cyber-security including the disclosure of data breaches and cyber-attacks [8, 94]. It appears that an observation made by Eckert ([32], p. 1) remains topical, that is: "*Because little new money has been provided to state and local authorities for infrastructure protection, let alone to the private sector, questions concerning who pays for security will remain problematic*".

It is evident that some private sector organisations are striving to develop advanced cyber-security methods. For example, industry associations such as the Fast Identity Online (FIDO) Alliance [40] is specifically attempting to reduce the use of passwords, arguing that the use of authentication technologies such as biometrics, USB security tokens and trusted platform modules (TPM) serves to minimise the use of IoT devices with generic factory passwords. Sector-led initiatives such as this may well prove to be effective ways of promoting the standardisation of improved security without the need for draconian legislative measures. A further motivating factor for private businesses to improve their cyber-security is the rising costs of cyber-insurance premiums, particularly in relation to the use of ransomware. Arguably, cyber insurance companies are forcing the hand of private businesses by increasing cyber-security requirements in order for them to obtain and maintain insurance cover [33]. In the light of earlier discussion in this chapter, a further interesting development for cyber insurance is that several providers have explicitly stated that they will no longer cover State sponsored attacks [9]. It is also highlighted that, at the level of the individual, some organisations in both the private and public sectors are, through the use of bounty programmes, encouraging both employees and certain non-employees (sometimes referred to as white hat hackers) to search for vulnerabilities within cyber systems [72]. In effect, freelance security workers can search, find, and report system vulnerabilities to companies for substantial financial rewards. This type of initiative also provides highly skilled individuals with challenges and financing outside of criminal activity.

8 Conclusion

This chapter has explored a range of issues relating to cyber-attacks on critical national infrastructure. Existent research indicates that the most dangerous attacks on critical infrastructure tend to involve malware which is able to access the physical controls of the industrial control systems. Our discussion of the nature of critical national infrastructure and vignettes of past cyber-attacks in various countries and sectors emphasise the wide-ranging and potentially catastrophic effects of these types of attacks. Despite the development of cyber-security systems, the fact remains that: "*The ability to cause serious and long-lasting damage to an enemy's industrial complex and war machine from anywhere on the planet, using intangible software to cause tangible effects, is incredibly powerful*" [54] and remains a very real threat. Notably, factors such as the convergence of IT and OT systems, and the increasing use

of the internet for IoT devices has led to the need for more focus on the vulnerabilities of critical national infrastructure within this changing landscape. The discussion has also drawn attention to the relatively low entry cost of engaging in cyber-attacks using malware, in contrast to the relatively high cost and logistical complexity of mounting physical attacks on well protected critical national infrastructure sites particularly in other countries. On this theme, one of the main conclusions which can be drawn from the discussion is the extent to which addressing vulnerabilities in critical national infrastructure cyber-systems can, potentially, involve a wide range of actors, such as State-level emergency planners, manufacturers of IoT devices, and white hat hackers.

The chapter has also identified various legal initiatives that have been taken by States to enhance cyber-security in both general and specific terms. On a positive note, there is some evidence of cooperation between States in areas such as the apprehension of those suspected of involvement in cyber-attacks on critical national infrastructure. Nevertheless, it is also evident that State-sponsored cyber-attacks, especially involving proxies, are a major confounding factor which is hindering the development of internationally enforceable protocols and laws designed to reduce the likelihood of cyber-attacks. In drawing attention to State-sponsored cyber-attacks, the discussion has specifically highlighted the danger of resource-rich States developing cyber-attack technologies which could end up being used against them. Having considered various legal national and cross-national initiatives linked to cyber-security, it appears that, from an international perspective, many aspects of legislation and policy are likely to remain more regulatory than statutory for the foreseeable future. That is, further and more significant developments to international cyber laws are needed to widen legal jurisdiction and enable the prosecution of individuals who commit cyber-attacks from distant locations. Legislation such as the push for Zero Trust Architecture in the USA appears to be a step in the right direction as a means of securing both IT and OT environments.

In conclusion, current discourse and events indicate that cyber-attacks on critical national infrastructure by State and non-State entities are likely to increase over coming years. Further, it is clear that any entity, State or otherwise, which engages in cyber-attacks on critical national infrastructure is likely to face both covert and overt reprisals especially if an attack has chemical, biological, or nuclear implications. As such, the potential consequences of these types of cyber-attacks have global ramifications and provide impetus to enhance and develop existing security infrastructure cyber-systems as well as presenting an overwhelming motivation to establish clear and legally enforceable accountability mechanisms.

References

1. ACSC (2021) ACSC Annual cyber threat report, July 2020 to June 2021. Published 15 September 2021. Retrieved from https://www.cyber.gov.au/sites/default/files/2021-09/ACSC%20Annual%20Cyber%20Threat%20Report%20-%202020-2021.pdf. Accessed on 30 Nov 2022

2. Alcaraz C (2019) Secure interconnection of IT-OT networks in industry 4.0. In: critical infrastructure security and resilience, Springer, Cham pp 201–217
3. Al Hait AAS (2014) Jurisdiction in Cybercrimes: a comparative study. J Law Policy Glob 22:75–84
4. Almeshekah MH, Spafford EH (2016) Cyber security deception. In: cyber deception, Springer, Cham pp 23–50
5. Anderson R, Fuloria S (2010) Security economics and critical national infrastructure. In: economics of information security and privacy, Springer, Boston, pp 55–66
6. Assenza G, Faramondi L, Oliva G, Setola R (2020) Cyber threats for operational technologies. Int J Syst Syst Eng 10(2):128–142
7. Baballe MA, Hussaini A, Bello MI, Musa US (2022) Online attacks, types of data breach and cyber-attack prevention methods. Curr Trends Inf Technol 12(2):21–26
8. Badhwar R (2021) CISOs need liability protection. In: The CISO's transformation, Springer, Cham pp 161–165
9. Baker T, Shortland A (2022) Insurance and enterprise: cyber insurance for ransomware. Geneva Pap Risk Insur-Issues Pract 1–25
10. Bērziņš J (2020) The theory and practice of new generation warfare: The case of Ukraine and Syria. J Sl Mil Stud 33(3):355–380
11. Boes S, Leukfeldt ER (2017) Fighting cybercrime: a joint effort. In: Cyber-physical security, Springer, Cham pp 185–203
12. Braw E, Brown G (2020) Personalised deterrence of cyber aggression. RUSI J 165(2):48–54
13. Broadhurst R (2006). Developments in the global law enforcement of cyber-crime. Policing An Int J 29(3):408–433
14. Bronk C, Conklin WA (2022) Who's in charge and how does it work? US cybersecurity of critical infrastructure. J Cyber Policy 7(2):155–174
15. Canfil JK (2022) The illogic of plausible deniability: why proxy conflict in cyberspace may no longer pay. J Cybersecur 1–16. https://doi.org/10.1093/cybsec/tyac007
16. Case DU (2016) Analysis of the cyber attack on the Ukrainian power grid. Electr Inf Sharing Anal Cent (E-ISAC) 388(1–29):3
17. CISC (2022) Cyber and infrastructure security centre. Protecting Australia together. Retrieved from: https://www.cisc.gov.au/critical-infrastructure-centre-subsite/Files/protecting-australia-together.pdf. Accessed on 30 Nov 2022
18. Clinton B (1998) A national security strategy for a new century. White House
19. Colarik A, Janczewski L (2015) Establishing cyber warfare doctrine. J Strateg Secur. Palgrave Macmillan, London 5(1):37–50
20. Collins S, McCombie S (2012) Stuxnet: the emergence of a new cyber weapon and its implications. J Polic Intell Counter Terrorism 7(1):80–91
21. Congress (2001) United States Patriot Act (2001). Retrieved from: https://www.congress.gov/107/plaws/publ56/PLAW-107publ56.pdf. Accessed Jan 2023
22. Conrad SH, LeClaire RJ, O'Reilly GP, Uzunalioglu H (2006) Critical national infrastructure reliability modeling and analysis. Bell Labs Tech J 11(3):57–71
23. Center for Strategic and International Studies (CSIS) (2018) Economic Impact of Cyber Crim–No Slowing Down. Retrieved from: http://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/economic-impact-cybercrime.pdf. Accessed 30 Nov 2022
24. Corfield G (2023) Russia–linked hackers behind Royal Mail cyber-attack. Daily telegraph, 12th January 2023. Retrieved from https://www.telegraph.co.uk/business/2023/01/12/russia-linked-hackers-behind-royal-mail-cyber-attack/. Accessed 12th Jan 2023
25. Council on Foreign Relations (2022) Cyber operations tracker. Retrieved from https://www.cfr.org/cyber-operations/#Glossary. Accessed 12th Dec 2022
26. Denning DE (2012) Stuxnet: What has changed? Future Int 4(3):672–687
27. Department of Homeland Security (2022) Critical infrastructure security and resilience research (CISRR) Fact Sheet. Retrieved from: https://www.dhs.gov/science-and-technology/publication/critical-infrastructure-security-resilience-research-fact-sheet#:~:text=Critical%20Infrastructure%20(CRITICAL%20INFRASTRUCTURE)%20consists%20of,or%20public%20health%20or%20safety. Accessed 30 Nov 2022

28. Dhatrak A, Sarkar A, Gore A, Paygude M, Waghmare M, Sahane H (2020) Cyber security threats and vulnerabilities in IoT. Int Res J Eng Technol 7(03)
29. Di Pinto A, Dragoni Y, Carcano A (2018) Triton: The first ICS cyber attack on safety instrument systems. In: Proc Black Hat USA Vol 2018, pp 1–26
30. Donnelly P, Abuhmida M, Tubb C (2022) The drift of industrial control systems to pseudo security. Int J Crit Infrastruct Prot 100535
31. Duncan S, Carneiro R, Braley J, Hersh M, Ramsey F, Murch R (2022) Cybersecurity: Beyond ransomware: securing the digital food chain. Food Aust 74(1):36–40
32. Eckert S (2005) Protecting critical infrastructure: the role of the private sector. Guns Butter Political Econ Int Secur 1
33. Eling M, Elvedi M, Falco G (2022) The economic impact of extreme cyber risk scenarios. North Am Actuarial J 1–15
34. Ellis R (2020) Letters, power lines, and other dangerous things: the politics of infrastructure security, MIT Press
35. Europa (2022) Cyber resilience act. Europa. Retrieved from: https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act. Accessed on 30th Nov 2022
36. Europa (2022) EU Directive 2016/ 1148. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L1148&from=EN. Accessed on 2nd Jan 2023
37. Europa (2022) EU Directive 2022/2555. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2555&from=EN. Accessed on 2nd Jan 2023
38. European Union (2020) Directive of the European parliament and of the council on measures for a high common level of cybersecurity across the Union, repealing
39. Farwell JP, Rohozinski R (2011) Stuxnet and the future of cyber war. Survival 53(1):23–40
40. Fast Identity Online (FIDO) Alliance (2022) The internet of things IoT. Retrieved from https://fidoalliance.org/internet-of-things. Accessed 15th Dec 2022
41. Fjäder C (2014) The nation-state, national security, and resilience in the age of globalisation. Resilience 2(2):114–129
42. Friis K, Lysne O (2021) Huawei, 5G and security: technological limitations and political responses. Dev Chang 52(5):1174–1195
43. Furnell S, Heyburn H, Whitehead A, Shah JN (2020) Understanding the full cost of cyber security breaches. Comput Fraud Secur 2020(12):6
44. Fuster GG, Jasmontaite L (2020) Cybersecurity regulation in the European Union: the digital, the critical and fundamental rights. In: The ethics of cybersecurity, Springer, Cham pp 97–115
45. Garimella PK (2018) IT-OT integration challenges in utilities. In: 2018 IEEE 3rd international conference on computing, communication and security (ICCCS) IEEE, pp199–204
46. Giannelli C, Picone M (2022) Editorial "Industrial IoT as IT and OT Convergence: Challenges and Opportunities." IoT 3(1):259–261
47. Glassberg J (2016) Defending against the ransom ware threat. Powergrid Int 21(8):22–24
48. Hagerott M (2014) Stuxnet and the vital role of critical infrastructure operators and engineers. Int J Crit Infrastruct Prot 7(4):244–246
49. Harrop W, Matteson A (2014) Cyber resilience: a review of critical national infrastructure and cyber security protection measures applied in the UK and USA. J Bus Contin Emer Plan 7(2):149–162
50. Hathaway M, Klimburg A (2012) Preliminary considerations: on national cyber security. Nat Cyber Secur Framework Manual. NATO Coop Cyber Defence Centre of Excellence Tallinn
51. Hayes K (2021) Ransomware: a growing geopolitical threat. Net Secur 2021(8):11–13
52. Herrmann D (2019) Cyber espionage and cyber defence. In: information technology for peace and security, Springer Vieweg, Wiesbaden pp 83–106
53. Hernandez-Castro J, Cartwright A, Cartwright E (2020) An economic analysis of ransomware and its welfare consequences. Roy Soc open sci 7(3):190023
54. Hobbs A (2021) The colonial pipeline hack: exposing vulnerabilities in us cybersecurity. In: SAGE Business Cases. SAGE Publications: SAGE business cases originals

55. House of Commons (2017) Post sector report for the house of commons committee on exiting the European Union. Retrieved from: https://www.parliament.uk/globalassets/documents/commons-committees/Exiting-the-European-Union/17-19/Sectoral-Analyses/27-Post-Report.pdf Accessed on16th January 2023
56. Huddleston J, Ji P, Bhunia S, Cogan J (2021) How vmware exploits contributed to solarwinds supply-chain attack. In: 2021 international conference on computational science and computational intelligence (CSCI) pp 760–765 IEEE
57. Hunter LY, Albert CD, Garrett E, Rutland J (2022) Democracy and cyberconflict: how regime type affects state-sponsored cyberattacks. J Cyber Policy 7(1):72–94
58. IBM (2022) Cyber-attacks. Retrieved from: https://www.ibm.com/uk-en/topics/cyber-attack. Accessed on 18th Dec 2022
59. Izycki E, Vianna EW (2021) Critical infrastructure: A battlefield for cyber warfare?. In: ICCWS 2021 16th international conference on cyber warfare and security, Academic Conferences Limited, p 454
60. Jacob JT (2022) A potential conflict over Taiwan: a view from India. Wash Q 45(3):147–162
61. Jones KS, Lodinger NR, Widlus BP, Namin AS, Maw E, Armstrong M (2022) Grouping and determining perceived severity of cyber-Attack consequences: gaining information needed to sonify cyber-attacks. J Multimodal User Interfaces 16(4):399–412
62. Kalech M (2019) Cyber-attack detection in SCADA systems using temporal pattern recognition techniques. Comput Secur 84:225–238
63. Kemabonta T (2021) Grid Resilience analysis and planning of electric power systems: The case of the 2021 Texas electricity crises caused by winter storm Uri. Electr J 34(10):107044
64. Kostyuk N, Kostyuk N, Zhukov YM (2019) Invisible digital front: can cyber attacks shape battlefield events? J Conflict Resolut 63(2):317–347
65. Lemay A, Fernandeza JM, Knight S (2010) Pinprick attacks, a lesser included case. In: conference on cyber conflict proceedings, Tallinn, Estonia: CCD COE, pp 183–194
66. Lewis JA (2002) Assessing the risks of cyber terrorism, cyber war and other cyber threats. Center for Strategic and International Studies, Washington, DC, p 12
67. Limba T, Plėta T, Agafonov K, Damkus M (2019) Cyber security management model for critical infrastructure
68. Lukasik SJ, Goodman SE, Longhurst DW (2020) Protecting critical infrastructures against cyber-attack. Routledge
69. Maglaras LA, Kim KH, Janicke H, Ferrag MA, Rallis S, Fragkou P, Cruz TJ (2018) Cyber security of critical infrastructures. Ict Express 4(1):42–45
70. Maglaras L, Ferrag MA, Derhab A, Mukherjee M, Janicke H, Rallis S (2019) Threats, protection and attribution of cyber attacks on critical infrastructures. arXiv preprint arXiv:1901.03899
71. Maillart JB (2019) The limits of subjective territorial jurisdiction in the context of cybercrime. In: Era Forum 19(3):375–390, Springer Berlin Heidelberg
72. Maillart T, Zhao M, Grossklags J, Chuang J (2017) Given enough eyeballs, all bugs are shallow? Revisiting eric raymond with bug bounty programs. J Cybersecur 3(2):81–90
73. Mamman A, Kamoche K, Rees C (2021) Attitudes to Globalization in the Public, Private and NGO Sectors. In: Baba Abugre J, Osabutey ELC, Sigué SP (eds) Business in Africa in the era of digital technology. Springer, London, pp 157–174
74. Martinelli F, Mercaldo F, Santone A (2022) A method for intrusion detection in smart grid. Procedia Comput Sci 207:327–334
75. Mcginthy JM, Michaels AJ (2019) Secure industrial internet of things critical infrastructure node design. IEEE Int Things J 6(5):8021–8037
76. Microsoft (2022) The hunt for NOBELIUM, the most sophisticated nation-state attack in history. Retrieved from: https://www.microsoft.com/en-us/security/blog/2021/11/10/the-hunt-for-nobelium-the-most-sophisticated-nation-state-attack-in-history/. Accessed on 18 Nov 2022
77. Microsoft (2022) Microsoft digital defense report 2022. Retrieved from: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5bUvv?culture=en-usandcountry=us Accessed on 19 Nov 2022

78. Miller T, Staves A, Maesschalck S, Sturdee M, Green B (2021) Looking back to look forward: lessons learnt from cyber-attacks on industrial control systems. Int J Crit Infrastruct Prot 35:100464
79. Murray G, Johnstone MN, Valli C (2017) The convergence of IT and OT in critical infrastructure. In: The Proceedings of 15th Australian information security management conference, Edith Cowan University, Perth, Western Australia. pp 149–155
80. Miller B, Rowe D (2012) A survey SCADA of and critical infrastructure incidents. In: Proceedings of the 1st annual conference on research in information technology, pp 51–56
81. Milone M (2003) Hacktivism: securing the national infrastructure. Knowl Technol Policy 16(1):75–103
82. National Institute of Standards and Technology (NIST). (2008). Guide to General Server Security. Retrieved from: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-123.pdf Accessed on 12th Dec 2022
83. Nazir S, Patel S, Patel D (2021) Autoencoder based anomaly detection for SCADA networks. Int J Artifi Intell Mach Learn (IJAIML) 11(2):83–99
84. Neigel AR, Claypoole VL, Waldfogle GE, Acharya S, Hancock GM (2020) Holistic cyber hygiene education: accounting for the human factors. Comput Secur 92:101731
85. Nguyen T, Wang S, Alhazmi M, Nazemi M, Estebsari A, Dehghanian P (2020) Electric power grid resilience to cyber adversaries: state of the art. IEEE Access 8:87592–87608
86. OECD (2008) Protection of 'Critical Infrastructure' and the role of investment policies relating to national security. Retrieved from http://www.oecd.org/daf/inv/investment-policy/40700392.pdf Accessed on 27 Nov 2022
87. Office for National Statistics. (2022). Nature of fraud and computer misuse in England and Wales: year ending March 2022. Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/natureoffraudandcomputermisuseinenglandandwales/yearendingmarch2022#:~:text=An%20estimated%2061%25%20of%20fraud,England%20and%20Wales%20(CSEW). Accessed 3 Jan 2023
88. Osawa J (2017) The escalation of state sponsored cyberattack and national cyber security affairs: is strategic cyber deterrence the key to solving the problem? Asia-Pac Rev 24(2):113–131
89. Osei-Kyei R, Tam V, Ma M, Mashiri F (2021) Critical review of the threats affecting the building of critical infrastructure resilience. Int J Disaster Risk Reduction 60:102316
90. Paul K (2021) Solar Winds hack was work of 'at least 1000 engineers'. The guardian. Retrieved from: http://www.theguardian.com/technology/2021/feb/23/solarwinds-hack-senate-hearing-microsoft. Accessed Dec 2022
91. Peisert, Sean, Bruce Schneier, Hamed Okhravi, Fabio Massacci, Terry Benzel, Carl Landwehr, Mohammad Mannan, Jelena Mirkovic, Atul Prakash, James Bret Michael. Perspectives on the solar winds incident. IEEE Secur Privacy 19(2):7–13
92. Peters A, Jordan A (2019) Countering the cyber enforcement gap: Strengthening global capacity on cybercrime. J Nat Secur Law Policy 10:487–495
93. Pérez-Martínez MM, Carrillo C, Rodeiro-Iglesias J, Soto B (2021) Life cycle assessment of repurposed waste electric and electronic equipment in comparison with original equipment. Sustain Prod Consumption 27:1637–1649
94. Radvanovsky R, McDougall A (2018) Critical infrastructure: homeland security and emergency preparedness. CRC Press
95. Reeder JR, Hall T (2021) Cybersecurity's Pearl Harbor moment. Cyber Defense Rev 6(3):15–40
96. Rees J (2022) The internet of things and terrorism: a cause for concern. In: privacy, security and forensics in the internet of things (IoT). Springer, Cham, pp 197–202
97. Rees J, Montasari R (2022) The Impact of the Internet and cyberspace on the rise in terrorist attacks across the US and Europe. In: disruption, ideation and innovation for defence and security. Springer, Cham, pp 135–148
98. Rid T (2012) Cyber war will not take place. J Strateg Stud 35(1):5–32

99. Ridley G (2011) National security as a corporate social responsibility: critical infrastructure resilience. J Bus Ethics 103(1):111–125
100. Sembiring Z (2020) Stuxnet threat analysis in SCADA (supervisory control and data acquisition) and PLC (Programmable logic controller) systems. J Comput Sci Inf Technol Telecomm Eng 1(2):96–103
101. Semwal P, Handa A (2022) Cyber-attack detection in cyber-physical systems using supervised machine learning. In: handbook of big data analytics and forensics, Springer, Cham pp 131–140
102. Serra KLO, Sanchez-Jauregui M (2021) Food supply chain resilience model for critical infrastructure collapses due to natural disasters. Bri Food J
103. Shahzad A, Lee M, Xiong NN, Jeong G, Lee YK, Choi JY, Ahmad I (2016) A secure, intelligent, and smart-sensing approach for industrial system automation and transmission over unsecured wireless networks. Sensors 16(3):322
104. Serpanos D, Wolf M (2018) Industrial internet of things. In: internet-of-things (IoT) Systems, Springer, Cham pp 37–54
105. Sharif MHU, Mohammed MA (2022) A literature review of financial losses statistics for cyber security and future trend. World J Adv Res Rev 15(1):138–156156
106. Silverman D, Hu YH, Hoppa M (2020) A study on vulnerabilities and threats to SCADA devices. J Colloquium Inf Syst Secur Edu 7(1):8
107. Simmons C, Ellis C, Shiva S, Dasgupta D, Wu Q (2009) AVOIDIT: a cyber attack taxonomy. University of Memphis. Technical report CS-09-003
108. Smith DC (2021) Cybersecurity in the energy sector: are we really prepared? J Energy Nat Res Law 39(3):265–270
109. Smith S (2022) Out of gas: a deep dive into the colonial pipeline cyberattack. In: SAGE Business Cases SAGE Publications, Ltd. Retrieved from https://doi.org/10.4135/9781529605679. Accessed on 16 Jan 2023
110. Straub J (2021) Defining, evaluating, preparing for and responding to a cyber Pearl Harbor. Technol Soc 65:101599
111. Sullivan JE, Kamensky D (2017) How cyber-attacks in Ukraine show the vulnerability of the US power grid. Electr J 30(3):30–35
112. Thomas J (2018) Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. Thomas JE (2018). Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. Int J Bus Manag 12(3):1–23
113. Thomas K, Li F, Zand A, Barrett J, Ranieri J, Invernizzi L, Bursztein E (2017) Data breaches, phishing, or malware? Understanding the risks of stolen credentials. In: proceedings of the 2017 ACM SIGSAC conference on computer and communications security, pp 1421–1434
114. Tsvetanov T, Slaria S (2021) The effect of the colonial pipeline shutdown on gasoline prices. Econ Lett 209:110122
115. United Kingdom Government. Terrorism Act 2000. Retrieved from: https://www.legislation.gov.uk/ukpga/2000/11/part/III/crossheading/offences. Accessed on 16th Jan 2023
116. Van de Weijer SG, Leukfeldt R, Bernasco W (2019) Determinants of reporting cybercrime: a comparison between identity theft, consumer fraud, and hacking. Eur J Criminol 16(4):486–508
117. Van der Meer S (2020) How states could respond to non-state cyber-attackers. Clingendael Policy Brief. Retrieved from: https://www.clingendael.org/sites/default/files/2020-06/Policy_Brief_Cyber_non-state_June_2020.pdf. Accessed on 16th Jan 2023
118. Warfield D (2012) Critical infrastructures: IT security and threats from private sector ownership. Inf Secur J Glob Perspect 21:127–136
119. Weiss M, Biermann F (2021) Cyberspace and the protection of critical national infrastructure. J Econ Policy Reform 1–18
120. Weiss J (2016) Aurora generator test. Handbook of SCADA/Control Systems Security 107
121. Watson FC, CISM C, ECSA A (2017). Petya/NotPetya: why it is nastier than wannacry and why we should care.ISACA 6:1-6

122. White House Archives (2013) Presidential policy directive PPD21. Presidential policy directive: Critical infrastructure security and resilience. Retrieved from: https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil. Accessed on 30th Nov 2022
123. White House Archives (2020) Executive order on securing the United States bulk-power system EO 13920. Retrieved from: https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-securing-united-states-bulk-power-system/. Accessed on 16th Jan 2023
124. White House Archives (2021) Executive Order on improving the nation's cybersecurity EO 14028. Retrieved from: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/. Accessed on 16th Jan 2023
125. Wolff ED, Growley KM, Gruden MG (2021) Navigating the solarwinds supply chain attack. Procurement Lawyer 56(2):3–11
126. Yadav G, Paul K (2019) Assessment of SCADA system vulnerabilities. In: 2019 24th IEEE international conference on emerging technologies and factory automation (ETFA), pp 1737–1744 IEEE
127. Yılmaz EN, Gönen S (2018) Attack detection/prevention system against cyber attack in industrial control systems. Comput Secur 77:94–105
128. Young S (2022) Moving the U.S. government toward zero trust cybersecurity principles Retrieved from: https://www.whitehouse.gov/wp-content/uploads/2022/01/M-22-09.pdf. Accessed on 30 Dec 2022
129. Yuste J, Pastrana S (2021) Avaddon ransomware: an in-depth analysis and decryption of infected systems. Comput Secur 109:102388

Police and Cybercrime: Evaluating Law Enforcement's Cyber Capacity and Capability

Nina Kelly and Reza Montasari

Abstract The COVID-19 pandemic has thrown the international community into disarray, resulting in a significant impact both on the rate at which digital technologies are incorporated into organisations' processes and on the cyber threat landscape. The pandemic led society and institutions to a global accelerated digitalisation, and in doing so reshaped the landscape in which cybercrime and cybersecurity operate (Horgan et al. in J Crim Psychol 11:222–239, 2020). Furthermore, Russia's recent invasion of Ukraine has had significant impact on the cyber threat landscape. Therefore, threats posed by cybercrime should be considered as the top priority for research attention, which is foundational to broader research in other fields such as Cyber Security, National and International Security, and Foreign Policy. To this end, this chapter aims to provide a critical analysis of the challenges that police and the wider law enforcement community encounter when responding to cybercrime. In view of this, the chapter also aims to assess law enforcement's cyber capacity and capability concerning their fight against cybercrime.

Keywords Cybercrime · Police · Law enforcement · Cybercrime · National security · International security · Foreign policy · Cyber security · Cyber criminals

1 Introduction

In the post-COVID-19 era, international consensus is in agreement of the threat of cybercrime being a top-level concern [1]. The pandemic led society and institutions to a global accelerated digitalisation, and in doing so reshaped the landscape in

N. Kelly (✉) · R. Montasari
Department of Criminology, Sociology and Social Policy, School of Social Sciences, Swansea University, Singleton Park, Swansea SA2 8PP, UK
e-mail: nina.devi.kelly@gmail.com

R. Montasari
e-mail: Reza.Montasari@Swansea.ac.uk
URL: http://www.swansea.ac.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_6

which cybercrime and cybersecurity operate [2]. With domestic statistics suggesting an 89% increase in computer misuse offences and a 25% increase in fraud offences in the year ending 2022 compared to the year ending 2020, it is apparent that the issue of cybercrime is becoming ever more persistent [3]. Additionally, Russia's invasion of Ukraine in 2022 further influenced the cyber threat landscape, with estimates suggesting that Russian-based phishing attacks increased eightfold against EU and US businesses compared to the time before the invasion [4].

To approach and mitigate this growing threat, the importance of essential collaboration across law enforcement, government, and the private sector in both a domestic and international context has been recognised [5, 6]. However, the high frequency and borderless nature of cybercrime severely complicates the global harmonisation that is needed from these sectors [7, 8]. This chapter will critically discuss some of these roadblocks and challenges faced by police and the wider law enforcement community in order to assess whether the battle on cybercrime is being lost in both a domestic and international context.

The remainder of this chapter is structured as follows. Section 2 discusses cybercrime in a domestic context focusing on the inconsistencies of recording cybercrime data and the challenges this poses when assessing the efficacy of efforts from law enforcement. Section 3 examines the issue of tackling cybercrime in an international context and the manner in which frameworks such as the Council of Europe's Convention on Cybercrime can be used to alleviate some of the roadblocks in investigating cybercrime. Section 4 explores how current limitations can be overcome in relation to deterrence-based control strategies and legislative restrictions through the Computer Misuse Act. Finally, the chapter is concluded in Sect. 5.

2 Cybercrime in the Domestic Landscape

As a starting point to assess the domestic landscape of cybercrime offending, official statistics such as the Crime Survey for England and Wales (CSEW) and police recorded data can provide a general indication of crime trends and patterns over a longitudinal period [9]. The Office for National Statistics (ONS) first introduced fraud and computer misuse into the CSEW in 2016, which resulted in the yearly crime rate almost doubling to 11.1 million with 3.4 million counts of fraud and 1.8 million computer misuse offences [10]. Since this inclusion, fraud and computer misuse remained at a constant at a count of approximately 6 million until the period March 2020 to April 2021 where the COVID-19 lockdowns increased computer misuse by 85% to 1.7 million offences [11, 12]. However, the most recent release for the year ending June 2022 shows a 27% decrease in computer misuse to pre-pandemic levels [13]. This trend is echoed similarly for fraud, which is estimated to have returned to the level it was in the year ending March 2020 [13]. When considering this data at face-value, it could be considered an indication that the battle on cybercrime is

turning in the police's and wider law enforcement's favour. However, closer analysis of these statistics and situating the data in the wider context of domestic cybercrime challenges this sentiment.

When looking further into the methodology of the CSEW, limitations become apparent. A first issue emerges as the data for the year ending March 2021 is not used in official crime trend estimates [13]. This is due to the CSEW data collection method changing to operate via telephone to accommodate COVID-19 restrictions, and ONS cautioning against the use of this data for comparison with previous years [12]. As previously mentioned, this was the time period for which computer misuse offences increased by 85% compared to the year previous [13, 12]. The cautious omission of this data from the wider set presents an issue when attempting to understand the scale of cybercrime to accurately observe trends, as the data must be consistently observed over a longer time period to increase accuracy [9]. A further limit of the CSEW in building a picture of domestic cybercrime offending is that it does not cover organisations or businesses in the data set, which further muddies the landscape of domestic cybercrime offending. This is particularly pervasive considering that small businesses have been consistently identified as one of the most vulnerable industries to cyberattacks, as evidenced by the 2019 Verizon data breach report finding 43% of cyber-attacks being directed to small businesses [14].

Police recorded data has the potential to address the gap in organisational cybercrime left by the CSEW, as it encompasses data from a range of industry sources through Action Fraud, Cifas, and UK Finance [15]. However, caution must be taken in considering the reliability of the data as there is no mechanism which cross-references the cases from each of these sources, making it probable that double and triple counting has occurred within the data set [15]. Further criticisms of the data can also be made considering the inconsistent reporting mechanism. It was noted that cases from Cifas are only recorded once they have been investigated, which could lead to an underestimate of cases [15]. However, the comparison of this data to other sources highlights a much larger issue in relation to this. For the period of June 2021 to June 2022 computer misuse cases totalled 29,637, which just about reflects 5% of the number of offences reported in the CSEW for the same period [13, 15]. This number is consistent with the reports from previous years, and highlights under-reporting as one of the main issues faced when attempting to understand the scale of the battle against cybercrime faced by wider law enforcement [15].

The issue in the under-reporting or non-reporting of cybercrime has constantly been identified in relation to policing [16–18]. Law enforcement agencies have been known to acknowledge that more effort needs to be made to firstly encourage victims to report cybercrime and secondly, demonstrate a level of response that makes reporting worthwhile [19]. Firstly, the issue of encouraging victims to report cybercrime has been addressed as part of a national effort to increase cyber awareness and recognise the seriousness of cybercrime in the UK [6]. Evidence of this can be seen through the Cyber Aware campaign, national Cyber Resilience Centres, and funded Essential Digital Skills qualifications as in the National Cyber Strategy 2022 [6].

The government's investment in these cyber security resources can also be seen as a positive, as in the previous National Cyber Security Strategy Report for 2016–2021, an investment of £1.9 billion was given over the 5 years [20].

The 2022 National Cyber Strategy Report details an increased investment of £2.6 billion over the next three years, with part of this investment going into a National Cyber Advisory Board aiming to directly reduce cyber harms and to support citizens and businesses with knowledge to protect themselves [6]. However, when it is considered that the National Fraud Intelligence Bureau (NFIB) have reported the loss of £4.3 billion over the past 13 months, and a past report suggesting that cybercrime has a much higher annual cost of £27 billion on the UK economy, it puts the impact of the £2.6 billion into perspective [21, 22]. The issue of ensuring that the police response to reports of cybercrime is adequate enough to encourage the reporting of cybercrime is an additional issue which threatens to undermine the aforementioned efforts if not done effectively [19]. A key report which reviewed police response in relation to cyber-dependent crime came from Her Majesty's Inspectorate of Constabulary Crime, Fire and Rescue Services [23]. Although the development of a strong national network was identified as a positive response, there was also an array of challenges and limitations [23].

An overarching issue was identified as inconsistent responses from local forces, from both an organisational perspective and in their responses to victims [23]. With only 8 out of 27 forces being able to provide data on cyber-dependent crime (all with low confidence in the accuracy of the data), no specialised analysts, and forces rarely seeing cyber-dependent crime as a priority, the response given to victims was noted as confusing and misleading [23]. These findings also echo the literature on police perceptions of cybercrime such as in Forouzan et al. [24], where the majority of Metropolitan police officers felt as they were not educated enough or adequately trained to respond to cybercrime. The range of frontline officers' perceptions of cybercrime were also gathered in Holt et al. [25], where a major deviation was observed in the perception of constables versus reality. These challenges arise in a climate where previous governmental reports assert that victims of cybercrime fail to report cybercrimes immediately due to perceiving the police to be ill-equipped for it, and therefore feeds into the cycle which prevents the police from learning how to respond effectively [26].

The issues which face the police in relation to understanding the commission of technology in cybercrime and investigating the case effectively is also impactful in the wider criminal justice system [27]. The borderless nature of cybercrime and the use of digital evidence in prosecution cases requires prosecutors, judges, and jurors to have a suitable understanding on the intricacies of technology aided crime [27]. However, prosecution statistics for cybercrime are notoriously low with sentences under the Computer Misuse Act 1990 totalling 216 for the period of 2018–2022 [28]. These low statistics could be an indication as to why there are no sentencing guidelines for computer misuse offences, as there are not enough cases to meet a consistent approach [29, 30].

Considering this lack of specific sentencing guidelines and the inconsistency in police responses to cybercrime offences as detailed in the [23] report, it suggests that domestically, the UK is struggling in the battle against cybercrime. The failure to put consistent procedures and responses into place has a large impact on the victims of cybercrime, with victimology studies highlighting the extent to which cybercrime offences impact individuals in a way in which is comparable to physical crime [29, 31]. In one survey of 252 victims of computer misuse, the majority of participants believed that the severity of computer misuse crimes was equivalent to or more serious than burglary, a physical trespass [29]. Although this study is biased toward victims who chose to report their crime, it does highlight the serious nature of cybercrime in a system which fails to prosecute offenders. However, an estimated 99% of overall cybercrime going unpunished also signals to the wider issue of jurisdiction as a vast amount of cybercrime is committed abroad [32]. Therefore, the international response to cybercrime has potential to guide domestic frameworks and is also of much importance [32].

3 Cybercrime in an International Context

The need for international cooperation in the context of cybercrime has been recognised as imperative in succeeding in the battle of cybercrime [33, 34]. The aim behind the creation of international guidelines is to establish procedures which enable digital investigations to cross international borders while also providing legislative framework to eliminate the areas in which criminals can operate beyond the reach of national legislation [33].

Historically, cyber criminals have been seen to take advantage of double criminality- a concept of international law which requires that the criminal act must be recognised as criminal in both the country of the offender and the country impacted by the crime for extradition to occur [35]. This was observed in the case of the 2000 'Love Bug' virus, where the creator could not be charged by the US National Bureau of Investigation due to there being no computer hacking laws in the creator's residing country of the Philippines [36]. Further examples of the complexity of extradition laws have been seen in the case of Gary McKinnon—a UK citizen charged on accounts of hacking into sensitive US military computers in 2002 [37]. Attempts to extradite him to the US resulted in a ten-year legal battle, ultimately ending in the Home Secretary blocking the extradition and the Crown Prosecution Service taking no further action [37]. These cases demonstrate the aspects of moral, political, and legislative differences that can arise between nations when cooperation is needed [34].

As a way to alleviate some of these barriers in cybercrime investigation, the Council of Europe designed a set of provisions to create a common criminal policy for cybercrime [34]. The result of this was the Budapest Convention on Cybercrime, which was passed in 2001 and currently ratified by 68 global states [38]. However, attempting to cast a framework over a large collective of states also opens up the

Convention for criticism. One area of concern targeted by critics is the level of investigative powers in relation to internet service providers (ISPs) and the search and seizure of customer's data [39]. These criticisms refer to Articles 16–21 which require legislation to be adopted to ensure the preservation of specific computer and network traffic data for up to a renewable 90 days, as well as access to real time traffic and ISP customer data [40].

The ability of the police to break beyond the traditional veil of private communication and utilising private-sector ISPs similarly to the role of law enforcement agents in obtaining data has been viewed as part of the wider issue of a disproportionate increase in the surveillance of private citizens [39]. These increased permissions for law enforcement come with an absence of clear standards to limit the use of the power and ensure it is not abused [41]. Concerns have also been made regarding Article 19 Clause 4:

> Each Party shall adopt such legislative and other measures as may be necessary to empower its competent authorities to order any person who has knowledge about the functioning of the computer system or measures applied to protect the computer data therein to provide, as is reasonable, the necessary information, to enable the undertaking of the measures referred to in paragraphs 1 and 2. ([40], p. 10)

The above clause of Article 19 (for which the legislation has been implemented in the UK [42]) has been raised as an issue in relation to self-incrimination, as turning over a decryption key could be considered as an action of self-incrimination [41]. Cryptographer Bruce Schneier also echoed concerns on this, suggesting that guilty individuals may benefit more by facing the maximum sentence of two years or five years (for national security or child indecency cases) for refusing to cooperate if turning over the key would result in a longer penalty in comparison [43]. In spite of these concerns, the Court of Appeal rejected the defence of self-incrimination when raised in an appeal, ruling that an encryption key exists as an independent fact as to whether the content of the data is incriminating or not [44]. These concerns illustrate the issues experienced when attempting to balance international law enforcement powers against global privately owned infrastructure and data which are integral parts of cybercrime investigations.

Despite these criticisms, the framework and influence of the Convention on Cybercrime has attempted to harmonise international and domestic cybercrime legislation, which has been seen as extremely relevant for international investigations [45]. The widespread impact of the Convention has been detailed in a publication where it was reported that 79% of United Nation members had used the Convention as a guidance for legislation reform and 55% of States adopted specific domestic provisions which correspond to the criminal law articles in the Convention [46].

Also included in this report is an array of case studies from member states which illustrate the impact of the treaty. From France utilising the 24/7 network of the Convention to obtain digital evidence at the time of the Charlie Hebdo terrorist attack, to Georgia dismantling an international child-trafficking ring, the impact of the framework has direct benefits in cybercrime investigations which cannot be denied [46]. The mechanism of creating additional protocols in which parties can

present amendments to the Committee of Ministers has also been utilised as an effective response, with a first amendment being used to create the Additional Protocol on Acts of a Racist and Xenophobic Nature [47]. In extension to this, a Second Additional Protocol is also currently in progress to provide Parties with additional tools for further collaboration in relation to digital evidence [48, 49]. This comes at a time where ransomware has been identified as a top threat, therefore amending the Convention to address this seems to be a step in the right direction to tackle the issue [50].

4 Future Challenges in Cybersecurity and the Wider Enforcement Community

Against the backdrop of cybercrime both internationally and domestically, a recurring theme which emphasises the need for balance between government, law enforcement agencies, and privately-owned internet service providers has emerged. However, central to this also lies cyber security professionals—key figures acting in the public interest to detect and prevent cyber threats. Finding a balance in giving cyber security professionals adequate legal boundaries to effectively tackle cyber threats within the constraints of legislation historically been identified as difficult under current domestic legislation [51]. The Computer Misuse Act (CMA) prohibits all unauthorised access to computer material regardless of intention, and in doing so criminalises a large majority of threat analysis and vulnerability testing for cybersecurity professionals [52].

It has been recognised that the threat of prosecution under computer misuse legislation can lead to security researchers feeling discouraged to find vulnerabilities, with cases of security researchers being treated like criminal hackers and being convicted under the CMA contributing to this sentiment [53, 54]. A more recent report found that four out of five professionals worry about breaking the law when defending cyber-attacks, and a near unanimous agreement that the CMA is not a piece of legislation fit for this century [55]. This issue particularly pertains to the independent group of vulnerability researchers who operate outside the authorisation which is granted to employees by IT vendors and the grey area in which intelligence agencies are able to operate [53]. The importance of needing the skills from these researchers in finding vulnerabilities have been highlighted in the 2022 National Cyber Strategy report [6]. It appears that policy makers are taking heed of the limitations arising from the CMA, with a series of debates considering the reform of the act and the inclusion of a form of statutory defence for cybersecurity professionals acting in the public interest [32]. This planned reform looks to be a positive step forward in the area of domestic cybercrime legislation.

Although proactive steps are being made on the defence front, in a policing context there are still challenges in creating a suitable crime control strategy to effectively tackle cybercrime. Notable differences in the commission of cybercrime weakens

the hold of an already developed strategy of deterrence, which is a key pillar of the crime, punishment, and the rule of law as developed for traditional crimes [56]. With the threat of criminal sanctions being drowned out by the lucrative profit from providing cybercrime as a service and a suggested 1% chance of conviction, it is clear that deterrence calculus tips in favour of criminality as a high reward, low risk calculation [19, 32]. As a result of this, strategies which utilise different methods have been seen to emerge. An example of this is [57]'s strategy of responsibilisation, which approaches crime control through a distributed policing strategy which individualises risk to reduce vulnerability. By doing this, the individual has a duty to make efforts to adequately protect themselves [57].

To some, this could be seen as a way of giving individuals more freedom on their choices, actions, and responsibility [58]. However, an approach which places large amounts of responsibility on an individual has the potential to lead way into victim blaming, which has already been seen in the media in the 2017 Wannacry ransomware attack [59]. In this case, journalists were quick to assign someone to blame for the attack, initially blaming a user for falling for a phishing scam, then moving on to blame Microsoft for not patching Windows XP and reaching further and further to assign blame—ultimately landing on the North Korean government [59].

Under a responsibilisation strategy such as Brenner's [57], this could have harsh implications, especially considering there is a limit to how much protection can be put in place by an individual, as it is not uncommon for several failures from multiple circumstances to create a situation that is vulnerable to cyber-attack [59]. However, current domestic strategies do take use of the elements of responsibilisation, particularly in the use of cyber-awareness and digital skills campaigns which informs individuals on the steps they can make to become more cyber secure [6]. Proactive strategies to disrupt the infrastructure in Cybercrime as a Service operations have also been suggested for centralised law agencies under infrastructural policing [60]. The strategy of infrastructural policing of cybercrime has been the result of a policy transfer from strategies used by the FBI in historic organised crime, offering a more practical and developed strategy [60]. The impact of infrastructural policing has been seen to be effective through case studies as illustrated in Collier et al. [60].

In 2016, part of a popular cybercrime forum Hackforums had announced their removal of the 'Server Stress Testing' part of their site which sold booter services for Denial of Service (DoS) attacks following warnings from the FBI [61]. The impact of this led to a drop in DoS attacks of approximately 28%, and a 16% decrease in Russian targeted attacks, a drop of 35% attacks targeting the US, and a decrease in 45% DoS attacks targeting the UK [60]. Influence policing was also observed as a strategy, which was originally developed as part of the UK counterterrorism PREVENT duty [60]. This was used in a 'Cyber Choices' campaign which adopted a behaviourist approach to target young people and young adults who are deemed at risk of becoming involved in cybercrime [62]. The case documented in Collier et al. [60] focused on digital advertising targeted to 16 to 24-year-olds who were searching for booter related terms on the Google search engine between December 2017 and June 2018. The results from this form of influence policing changed the

existing upward trend in attacks (a positive gradient of 2.9) to appear as flattened (a positive gradient of 0.1) for the duration of the campaign in the UK, whereas the rest of the worldwide attacks continued to rise [60].

Although further empirical support for the effectiveness of influential and infrastructural policing of cybercrime as a service is not entirely present due to the recency of the study, the Policing Vision 2025 report recognises the need to develop new law enforcement capabilities which tackle terrorism, cybercrime, and organised crime [63]. As infrastructural policing drew upon strategies in organised crime, and influence from PREVENT was present in the behavioural approach of influence policing, combining developed approaches to suit the nuances of cybercrime therefore has the potential to allow law enforcement to increasingly reclaim ownership of cybercrime issues [60].

5 Conclusion

Overall, cybercrime can be considered as an overwhelming opponent in the battle faced by international and domestic law enforcement. With domestic statistics underestimating the number of offences and suggesting a fractional percentage of offenders are bought into justice, encouraging victims to report cybercrime to the authorities can be challenging. This issue is further exacerbated by the fact that the internal procedures and understanding of cybercrime in the police varies with inconsistency from each local force. Challenges are also faced when the training and resources needed by the police to improve their local response is reliant upon on the amount of funding received by the government, which in the past has been disproportionate to the cost of cybercrime—although now funding is increasing slightly. In a domestic context, it can be considered that the battle to curb cybercrime is slow with plenty of struggle. However, actions such as planning to reform the CMA to assist cybersecurity professionals, and new impactful strategies being developed in the ineffectiveness of deterrence also suggest that there is a domestic effort to equip the nation with more useful tools, although only time will tell if this makes an impactful change.

In an international context, the Convention on Cybercrime has totemic value in providing a framework to harmonise domestic and international laws on cybercrime, and in doing so, reducing the areas in which cyber criminals can hide from the law. Although criticised for undermining data privacy and encryption mechanisms, it has created a global network which works to investigate and prevent cyber related harms. The global nature of cybercrime is one of, if not the biggest advantage for offenders in the battle against law enforcement, and the creation of a mechanism which mitigates this advantage is a strong effort. With future amendments to the treaty aiming to strengthen and further these mechanisms in the context of the increased threat from ransomware, it can be seen as a confident move from international law enforcement in tackling cybercrime.

References

1. INTERPOL (2022) 2022 INTERPOL global crime trend summary report. https://www.interpol.int/en/content/download/18350/file/Global%20Crime%20Trend%20Summary%20Report%20EN.pdf
2. Horgan S, Collier B, Jones R, Shepherd L (2020) Re-territorialising the policing of cybercrime in the post COVID-19 era: towards a new vision of local democratic cyber policing. J Crim Psychol 11(3):222–239. https://doi.org/10.1108/JCP-08-2020-0034
3. ONS (2022) Nature of fraud and computer misuse in England and Wales: year ending March 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/natureoffraudandcomputermisuseinenglandandwales/yearendingmarch2022
4. Griffiths C (2023) The latest 2023 cyber crime statistics (updated January 2023). AAG. https://aag-it.com/the-latest-cyber-crime-statistics/
5. Europol (2021) Internet organised crime threat assessment (IOCTA) 2021. Publications Office of the European Union. https://www.europol.europa.eu/cms/sites/default/files/documents/internet_organised_crime_threat_assessment_iocta_2021.pdf
6. HM Government (2022) National cyber strategy 2022. Cabinet Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1053023/national-cyber-strategy-amend.pdf
7. Curtis J, Oxburgh G (2022) Understanding cybercrime in 'real world' policing and law enforcement. Police J 1–20. https://doi.org/10.1177/0032258X221107584
8. Khan S, Saleh T, Dorasamy M, Khan N, Leng OTS, Vergara RG (2022) A systematic literature review on cybercrime legislation. F1000Research 11:971. https://doi.org/10.12688/f1000research.123098.1
9. Newburn T (2017) Criminology, 3rd edn. Routledge
10. ONS (2017) Crime in England and Wales: year ending Mar 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmar2017
11. ONS (2020) Crime in England and Wales: year ending September 2019. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingseptember2019#fraud
12. ONS (2021) Crime in England and Wales: year ending March 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingmarch2021#computer-misuse
13. ONS (2022) Crime in England and Wales: year ending June 2022. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingjune2022
14. Verizon (2019) 2019 Data breach investigations report. https://www.verizon.com/business/resources/reports/2019-data-breach-investigations-report.pdf
15. ONS (2022) Crime in England and Wales: appendix tables. Year ending June 2022 edition of this dataset. https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/crimeandjustice/datasets/crimeinenglandandwalesappendixtables/yearendingjune2022/appendixtablesjune22.xlsx
16. Correia SG (2022) Making the most of cybercrime and fraud crime report data: a case study of UK Action Fraud. Int J Popul Data Sci 7(1):1–17. https://doi.org/10.23889/ijpds.v7i1.1721
17. McGuire M, Dowling S (2013) Cyber crime: a review of the evidence research report 75—chapter 4: improving the cyber crime evidence base. Home Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/246756/horr75-chap4.pdf
18. Wall DS (2008) Cybercrime, media and insecurity: the shaping of public perceptions of cybercrime. Int Rev Law Comput Technol 22(1–2):45–63. https://doi.org/10.1080/13600860801924907
19. Saunders J (2017) Tackling cybercrime—the UK response. J Cyber Policy 2(1):4–15. https://doi.org/10.1080/23738871.2017.1293117

20. HM Government (2016) National cyber security strategy 2016 to 2021. Cabinet Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf
21. Detica (2011) The cost of cyber crime. Cabinet Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60943/the-cost-of-cyber-crime-full-report.pdf
22. NFIB (2023) NFIB fraud and cyber crime dashboard—13 months of data. Retrieved from https://colp.maps.arcgis.com/apps/dashboards/0334150e430449cf8ac917e347897d46. Accessed on 12 Jan 2023
23. HMICFRS (2019) Cyber: keep the light on—an inspection of the police response to cyber-dependent crime. Justice Inspectorates. https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/cyber-keep-the-light-on-an-inspection-of-the-police-response-to-cyber-dependent-crime.pdf
24. Forouzan H, Jahankhani H, McCarthy J (2018) An examination into the level of training, education and awareness among frontline police officers in tackling cybercrime within the Metropolitan Police Service. In: Jahankhani H (ed) Cyber criminology. Springer, pp 307–323
25. Holt TJ, Burruss GW, Bossler AM (2019) An examination of English and Welsh constables' perceptions of the seriousness and frequency of online incidents. Polic Soc 29(8):906–921. https://doi.org/10.1080/10439463.2018.1450409
26. HMICFRS (2015) Real lives, real crimes: a study of digital crime and policing. Justice Inspectorates. https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/real-lives-real-crimes-a-study-of-digital-crime-and-policing.pdf
27. Brown CS (2015) Investigating and prosecuting cyber crime: forensic dependencies and barriers to justice. Int J Cyber Criminol 9(1):55–119. https://doi.org/10.5281/zenodo.22387
28. ONS (2022) Criminal justice system statistics quarterly: June 2022—outcomes by Offence data tool: June 2022. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1126556/outcomes-by-offence-june-2022-revised.xlsx
29. Button M, Shepherd D, Blackbourn D, Sugiura L, Kapend R, Wang V (2022) Assessing the seriousness of cybercrime: the case of computer misuse crime in the United Kingdom and the victims' perspective. Criminol Crim Justice. https://doi.org/10.1177/17488958221128128
30. CPS (n.d.) Computer Misuse Act. Retrieved from https://www.cps.gov.uk/legal-guidance/computer-misuse-act. Accessed on 12 Jan 2023
31. Button M, Blackbourn D, Sugiura L, Shepherd D, Kapend R, Wang V (2021) From feeling like rape to a minor inconvenience: victims' accounts of the impact of computer misuse crime in the United Kingdom. Telematics Inform 64:1–11. https://doi.org/10.1016/j.tele.2021.101675
32. House of Commons (2022) Computer Misuse Act 1990: Volume 712: debated on Tuesday 19 April 2022. https://hansard.parliament.uk/commons/2022-04-19/debates/AE9413F3-D4F2-44EC-890E-75B0250328C4/ComputerMisuseAct1990
33. Cerezo AI, Lopez J, Patel A (2007) International cooperation to fight transnational cybercrime. In: Second international workshop on digital forensics and incident analysis, pp 13–27. https://doi.org/10.1109/WDFIA.2007.4299369
34. Marion NE (2010) The Council of Europe's cyber crime treaty: an exercise in symbolic legislation. Int J Cyber Criminol 4(1&2):699–712. https://www.cybercrimejournal.com/pdf/marion2010ijcc.pdf
35. UNODC (2019) Formal international cooperation mechanisms. https://www.unodc.org/e4j/en/cybercrime/module-7/key-issues/formal-international-cooperation-mechanisms.html
36. Arnold W (2000) TECHNOLOGY; Philippines to drop charges on e-mail virus. The New York Times. https://www.nytimes.com/2000/08/22/business/technology-philippines-to-drop-charges-on-e-mail-virus.html
37. BBC (2012) Gary McKinnon: timeline. https://www.bbc.co.uk/news/19959726
38. Council of Europe (2023) Chart of signatures and ratifications of Treaty 185. Retrieved from https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=185. Accessed on 13 Jan 2023

39. Huey L, Rosenberg RS (2004) Watching the web: thoughts on expanding police surveillance opportunities under the cyber-crime convention. Can J Criminol Crim Justice 46(5):597–606. https://doi.org/10.3138/cjccj.46.5.597
40. Council of Europe (2001) Convention on cybercrime (European treaty series—No. 185). Council of Europe. https://rm.coe.int/1680081561
41. Taylor G (2001) The Council of Europe cybercrime convention: a civil liberties perspective. Priv Law Policy Reporter 8(4). http://classic.austlii.edu.au/au/journals/PrivLawPRpr/2001/35.html
42. Regulation of Investigatory Powers Act 2000 c.23, p.III. Retrieved from https://www.legislation.gov.uk/ukpga/2000/23/part/III
43. Schneier B (2007) UK police can now demand encryption keys. Schneier on Security. https://www.schneier.com/blog/archives/2007/10/uk_police_can_n.html
44. R v. S, Anor (2008) Court of appeal, criminal division, case 2177. Bailii. Retrieved from http://www.bailii.org/ew/cases/EWCA/Crim/2008/2177.html
45. Clough J (2014) A world of difference: The Budapest convention on cybercrime and the challenges of Harmonisation. Monash Univ Law Rev 40(3):698–736. https://www.monash.edu/__data/assets/pdf_file/0019/232525/clough.pdf
46. Cybercrime Convention Committee (2020) The Budapest convention on cybercrime: benefits and impact in practice (T-CY (2020)16). Council of Europe. https://rm.coe.int/t-cy-2020-16-bc-benefits-rep-provisional/16809ef6ac
47. Council of Europe (2003) Additional protocol to the convention on cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems (European treaty series: No 189). Council of Europe. https://rm.coe.int/168008160f
48. Council of Europe (2022) Second additional protocol to the convention on cybercrime on enhanced cooperation and disclosure of electronic evidence (Council of Europe treaty series—No. 224). Council of Europe. https://rm.coe.int/1680a49dab
49. Cybercrime Convention Committee (2022) T-CY Guidance Note #12 Aspects of ransomware covered by the Budapest Convention (T-CY(2022)14). Council of Europe. https://rm.coe.int/t-cy-2022-14-guidancenote-ransomware-v4adopted/1680a9355e
50. ENISA (2022) ENISA threat landscape 2022. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022/@@download/fullReport
51. ENISA (2015) Good practice guide on vulnerability disclosure from challenges to recommendations. European Union Agency for Network and Information Security. https://www.enisa.europa.eu/publications/vulnerability-disclosure/@@download/fullReport
52. House of Commons Library (2022) Westminster Hall debate on the Computer Misuse Act 1990. UK Parliament. https://commonslibrary.parliament.uk/research-briefings/cdp-2022-0082/
53. Guinchard A (2017) The Computer Misuse Act 1990 to support vulnerability research? Proposal for a defence for hacking as a strategy in the fight against cybercrime. J Inf Rights Policy Pract 2(2):1–35. https://doi.org/10.2139/ssrn.2946763
54. Oates J (2005) Tsunami hacker convicted. The Register. https://www.theregister.com/2005/10/06/tsunami_hacker_convicted/
55. CyberUp (2020) Time for reform? Understanding the UK cyber security industry's views of the Computer Misuse Act. Tech UK. https://static1.squarespace.com/static/5e258d570aee2d7e8a7bcad9/t/5fb628ff3955d5421c935807/1605773584121/CyberUp-techUK_Time_for_reform.pdf
56. Maimon D (2020) Deterrence in cyberspace: an interdisciplinary review of the empirical literature. In: Holt T, Bossler A (eds) The Palgrave handbook of international cybercrime and cyberdeviance. Palgrave Macmillan, pp 1–19
57. Brenner S (2007) Rethinking crime control strategies. In: Jewkes Y (ed) Crime online. Willan Publishing, pp 12–29
58. Biebricher T (2011) (Ir-)Responsibilization, genetics and neuroscience. Eur J Soc Theory 14(4):469–488. https://doi.org/10.1177/1368431011417933

59. Renaud K, Flowerday S, Warkentin M, Cockshott P, Oregon C (2018) Is the responsibilization of the cyber security risk reasonable and judicious? Comput Secur 78:198–211. https://doi.org/10.1016/j.cose.2018.06.006
60. Collier B, Thomas DR, Clayton R, Hutchings A, Chua YT (2022) Influence, infrastructure, and recentering cybercrime policing: evaluating emerging approaches to online law enforcement through a market for cybercrime services. Polic Soc 32(1):103–124. https://doi.org/10.1080/10439463.2021.1883608
61. Kan M (2016) Hacking forum cuts section allegedly linked to DDoS attacks. Networkworld. https://www.networkworld.com/article/3136727/hacking-forum-cuts-section-allegedly-linked-to-ddos-attacks.html
62. Cyber Choices (n.d.) Helping you choose the right and legal path. National Crime Agency. https://nationalcrimeagency.gov.uk/cyber-choices?highlight=WyJtaXNzaW5nIiwibWlzc2VkIiwibWlzcyIsIidtaXNzaW5nJywiLCJwZXJzb24iLCJwZXJzb25zIiwicGVyc29uYWwiLCJwZXJzb25hbGl0eSIsInBlcnNvbmFsbHkiLCJwZXJzb24ncyIsInBlcnNvbnMnIiwiZGF0YSIsIm1pc3NpbmcgcGVyc29uIl0
63. NPCC (2016) Policing vision 2025. https://www.npcc.police.uk/documents/Policing%20Vision.pdf

Law Enforcement and Digital Policing of the Dark Web: An Assessment of the Technical, Ethical and Legal Issues

Charlotte Warner

Abstract This chapter aims to investigate the challenges that law enforcement agencies (LEAs) face in policing the dark web and explore ways in which these challenges impact the effectiveness of law enforcement responses to crime. Following a comprehensive review of the literature, the main trend discovered is that these challenges arise especially due to the anonymous profile of the dark web and the ethics involved in detecting criminal activity. The issues examined have led to recommendations for reducing the negative impact of the Dark Web on policing activities. Conclusions drawn from the analysis support the recommendations in the need for increased cyber threat intelligence, the need for new regulations and a deeper concern for ethics. The findings also reveal the importance of ensuring a right balance between policing the Dark Web and respecting individuals' civil liberties. In order to achieve a common objective, cross-jurisdictional law enforcement co-ordination and international co-operation are essential. In addition, the results suggest that in order for digital policing to be effective, the techniques used must also be unpredictable to criminals.

Keywords Dark web · digital policing · Cybercrime · Law enforcement · Internet · Digital forensics · Cyber security · Digital investigation · ethics · Civil liberties · Tor browser · Privacy

1 Introduction

The internet is a vast place, accessible for everyone to access endless information [1]. However, there are darker aspects of the internet, such as the dark web, which are a hub for illicit activities to take place, due to it allowing total anonymity to its servers [2]. This chapter focuses on assessing the weaknesses of digital policing methods and problems faced with techniques used to detect criminal activity. Following this

C. Warner (✉)
Department of Criminology, Sociology and Social Policy, School of Social Sciences, Swansea University, Singleton Park, Swansea SA2 8PP, UK
e-mail: charlottelwarner@gmail.com

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_7

assessment, suggestions will be made on how to overcome or mitigate these issues. This will be achieved by carrying out an extensive review of previous literature. This review focuses on analysing recent cybercrime research and investigating different activities used by LEAs and their effectiveness and ethicality. It is apparent that the vast and secret nature of the dark web means it is difficult to fully understand its inner workings [3–6]. Furthermore, it is important to delve into the issues that the use of each of these methods poses to the general public such as infringing on human rights and individual privacy, as well as to legal systems across different jurisdictions and how to prevent future crime [7]. The chapter proceeds to outline recommendations in order to improve the effectiveness and ethical aspects of digital policing methods. These suggestions are discussed, and conclusions are reached on the importance of a balance between civil liberties and crime investigation concerning the dark web.

2 Background

Because it is virtually limitless and allows for instantaneous communication, cyberspace, which is made up of the surface web, deep web, and dark web [3–6], is particularly vulnerable to exploitation and can be used to carry out a wide range of illegal and malicious activities in very efficient ways. In recent years, academic writing on crime and deviance in cyberspace has surged [8]. The searchable portion of cyberspace that is made up of search engines such as Amazon, where users are recognised and their activities are monitored, is known as the surface web (clear web). Approximately 1–4% of the internet is made up of the surface web. Anything on the internet that is not indexed by a search engine is referred to as the deep web, and the material includes anything protected by sign-in forms or paywalls. Compared to the surface web, this portion of the internet is enormous. The deep web, which can only be accessed with a Tor browser, constitutes the bulk of the dark web [9]. Its exact size is unknown, however, estimates place it at 5% of the total Internet and according to research, 57% of the dark web hosts illicit material as of 2015 [10].

In the late 1990s, The Onion Router project (Tor network) was created to protect private communications from US espionage [11]. Every communicatison on the Dark Web is encrypted and routed through a network of proxy servers, even the smallest transaction necessitates the use of a Pretty Good Privacy (PGP) key [12]. Its own search engines are available and are an additional way to use link lists to browse the dark web. Tor is the most well-known low-latency anonymity network to date. Users can remain anonymous through Tor, which also facilitates the rollout of hidden services or anonymous services [13].

The dark web's privacy and freedom from authoritarian governments' surveillance have positive aspects. Whistle-blowers are shielded from retaliation, and anyone concerned about their data can benefit from privacy [2]. There are numerous organisations that have a dark web page. The same seclusion and anonymity of the Dark Web, however, make it a haven for criminal activity. It facilitates criminals in participating in illegal operations such as hiring hit men, hacking, and technological crime

services, as well as trafficking weapons, people, and drugs [14]. Thus, the dark web swiftly turned into a tool for terrorists to disseminate their message while also serving as an accessory to cybercrime and other illegal activities [15]. As a result, governments, and LEAs were compelled to develop strategies to end its anonymity and find techniques to catch internet-savvy criminals [16].

3 Challenges

Despite its importance, digital policing the dark web presents LEAs with several technical, legal and ethical challenges. For instance, digital policing employs surveillance software to directly access and manage criminals' devices while malware exerts total control over foreign systems in order to investigate crime on the dark web [17]. However, there are legal obstacles to hacking enforcement. There is a lack of uniformity in international law, and when a crime occurs across borders, the responsibility is shared, which can make it challenging to administer justice fairly [18]. This creates jurisdictional challenges, for instance, China, which is adamantly opposed to the use of TOR in any situation, has policies and laws regarding its users that are quite different from those in the United States (US and inconsistent with their goals [19]. The Federal Rule of Criminal Procedure 41 rules that there needs to be probable cause to issue a warrant in order to hack into a person's device. As of 2016, if the location of the device to be searched had been "concealed through technological means", a remote access warrant might be obtained in a U.S. federal court even if the court was not aware of the location of the device [20], therefore, allowing the hacking of devices that may be located in other jurisdictions.

National sovereignty may be threatened if criminals are pursued across borders [21] or if there are unauthorised investigations of foreign governments. It is considered an incursion on another state's sovereignty to carry out law enforcement functions within another state without that state's consent. When considering who can legally sanction these operations, where they can be deployed, and who they can be used against, overseas cyber-operations create challenging considerations. Because they permit these decisions to be made despite potentially disruptive foreign relations ramifications, the laws of criminal procedure fall short of regulating law enforcement hacking [21]. A cross-border cyber operation may be categorised as an internationally wrongful act, allowing a state to respond with countermeasures under customary international law. This obliges the harmed states to use otherwise banned force in self-defence depending on the scope and seriousness of the harm caused by the operation [22]. For instance, Edward Snowden disclosed to the South China Morning Post that The NSA was in charge of more than 61,000 cyber actions worldwide [23], many of which were carried out in Hong Kong and mainland China [24].

In order to safeguard the public's civil freedoms, the fourth amendment shields people from arbitrary search and seizure [25]. Respecting one's family and private life while not feeling that everything is being watched over by the government is a fundamental human right [26]. Thus, hacking devices may be seen to undermine

this freedom for many. It is argued that governments take this authority too far, as shown by documents leaked by Edward Snowden in which the US National Security Agency (NSA) was collecting the telephone records of tens of millions of Americans [27].

Although authoritarian governments continue to try and prevent access to the dark web and promote de-anonymisation, there are challenges with the idea of shutting it down completely [28]. Allowing anonymity to the public is vital because many people such as whistleblowers, journalists, human rights activists, and spies engaging in specialised forms of communication were concerned about hiding their identities as Internet users [18]. The New Yorker's Strongbox, for instance, is accessible through Tor and allows individuals to communicate and share documents anonymously with the publication [29]. Many believe that the Tor network should remain uncensored because it allows for people such as Edward Snowden, who used Tails, an "operating system built for anonymity" that launches Tor automatically, to communicate with reporters and disclose sensitive information on American mass surveillance programmes [27]. A top-secret presentation discussing NSA plans to take advantage of the Tor browser and de-anonymise users was one of the documents exposed by Snowden [30]. It is important to allow civilians a safe way to expose immoral acts by the government. However, allowing this freedom enables more criminal activities to proceed undetected, highlighting the importance of finding a balance between both authority and freedom.

There are also many ethical problems in relation to the digital policing methods and techniques. For example, the use of honeypot traps is controversial. Honeypot traps involve an actor posing as a person in an effort to start a connection online or persuade the victim to visit a website that is infected with malware in order to lure them into performing an act, typically sexual [31]. The searchers of this content may come across sites that are legal and feature adults posing as youngsters, which are "sting" operations, or vigilante sites [32].

However, there are moral problems with utilising the honeypot trap method, such as increased harm to children. This can be seen in the Child's Play operation, which was one of the biggest child abuse websites on the dark web. While the website was under police control, by Taskforce Argos, other users continued to submit and view photographs as undercover officers also shared and posted offensive content on Childs' Play to keep viewers from realising the site was compromised [33]. Furthermore, this method of policing involves entrapment, and a violation of civil rights. Falsehoods are unethical as individuals essentially manipulate others. Deception may cause people who are usually perceived or act as normal citizens to behave abnormally, which is shown in Milgram's study of obedience in which everyday people inflicted fatal electric shocks on another participant, who was unbeknownst to them, not real [34].

There is currently inadequate evidence that shows honeypot traps to be the most efficient approach. Due to the ubiquity of undercover law enforcement agents on the dark web, there are issues with the efficiency of operations employed to arrest paedophiles there as well [35]. For instance, password-protected chat rooms can be used to exchange child pornography and find future victims. Experienced child

pornographers avoid open chat rooms because they are frequently infiltrated by undercover police, and the effectiveness of these operations is limited [32].

Notwithstanding the note above, this method of digital policing has been successful to some degree. For instance, investigation of child's play resulted in the arrest of the two individuals running the site. Furthermore, the continued undercover work allowed police to investigate up to 4000 active users and led to significant rescues of children globally [36]. As it contributes to the arrest and removal of paedophiles from society, it may be seen as a lesser evil and useful for the greater good of public safety. Thus, it can be an effective method for arresting paedophiles.

Digital policing also faces technical challenges due to blockchain technology. For instance, the use of bitcoin has increased privacy because the wallet and private key are not visible in the public ledger. However, this has allowed both legitimate and illegitimate users to utilise it [37]. The wallet stores a user's private key, which functions as a password-like secret code that enables that user to spend bitcoins from the associated wallet. Transactions are verified using the address for the transaction and a cryptographic signature. It has become a major factor in the growth of the dark web and is widely used in money laundering activities. The underlying strength of cryptocurrency is its encryption [38]. In 2014, Amir Taaki and Cody Wilson founded Dark Wallet, which is an open-source bitcoin platform that aims to make users anonymous and obscure bitcoin transactions. Coin mixing is one of its main applications. Coin mixing combines a user's transaction with those of other arbitrary users who just so happen to be executing different transactions across the system concurrently [39]. The user is not made anonymous when they use Bitcoin. Instead, they are given a pseudonym. Each wallet is an address made up of 26–35 alphanumeric characters in the blockchain. The public ledger for bitcoin is not always associated with an individual, but each transaction is tracked independently and may be linked to the user's available bitcoin address. Instead of identifying individuals, coin mixing software makes them anonymous by blending their bitcoins with those of other users. This provides technical challenges as it is increasingly challenging to identify criminals acting on the dark web due to the increased anonymity.

In response, using chat rooms to combat illicit activities concerning bitcoin has displayed tangible results such as the successful capture of "Mr Hotsauce", a dealer who openly offered to sell various illegal narcotics to customers around the world [40]. However, despite being a highly helpful strategy, it is no longer as effective a solution to the problem as it once was, and criminals are becoming suspicious [41]. As a result, it is unlikely to be a lasting solution. Therefore, the challenge lies in the abilities of LEAs and the tools they possess to determine how to identify the dark web servers.

4 Recommendations

International collaboration must be considered in response to jurisdictional challenges that the dark web poses to digital policing. Improving information exchange between LEAs, financial institutions, and regulators worldwide, will make digital policing much more effective, enabling the arrest of criminals across jurisdictions [19]. The international operation to seize AlphaBay's infrastructure was coordinated by the US and involved assistance from LEAs in Thailand, the Netherlands, Lithuania, Canada, the United Kingdom and France as well as the European LEA, Europol [42]. LEAs rely on government diplomatic relations and treaties with other countries, seeking permission from the host state before deploying personnel and requesting assistance from local authorities to collect foreign-located evidence when possible [43]. This demonstrates the importance of having strong foreign relations. For instance, the Drug Enforcement Administration has recently confirmed that it has used hacking tools on seventeen devices in foreign countries, pursuant to a foreign court order and with the cooperation of foreign officials [44]. This means that national sovereignty is not challenged and highlights the significance of jurisdictions working in collaboration to combat cybercrime on the dark web.

To address challenges associated with the use of cryptocurrencies in money laundering through the dark web, it is vital for new regulations to be introduced. Regulating cryptocurrencies that fuel dark web marketplaces. The U.K.'s National Crime Agency (NCA) has called for regulations that would force coin mixers to comply with anti-money laundering laws, including carrying out KYC (know your customers) checks on customers and conducting audit trails [45]. These goals may be attained through multi-partner projects that combine the knowledge and abilities of the aforementioned groups and their members. An example of a good initiative by the Financial Action Task Force is the creation of the framework by the regulators for enhanced security and further oversight [46], for example, to watch for the potential challenges presented by the forthcoming launch of Diem, Facebook's cryptocurrency [47].

A uniform framework and government oversight are needed to prevent misinformation and misuse of cryptocurrencies, for example, The White House is considering a wide-ranging oversight of the cryptocurrency market, and the US national security advisers will gather officials from 30 countries [48]. New regulations have, however, also resulted in the shutting down of crypto start-ups leaving the market after the implementation of the fifth AML directive in Europe. This has culminated in harming the domestic economies by driving out innovation [49].

In response to the ethical challenges faced by using honeytraps, increasing funding for LEAs' hackers allows them to gather the tools and knowledge necessary to shut down websites on the dark web [50]. Taking down websites may be a more moral alternative to deception, and also it cannot be predicted by the website owners. For example, after the original Freedom Hosting was shut down in 2013, Freedom Hosting II was set up as a website hosting service on the dark web. In February 2017, hackers allegedly connected to Anonymous took down Freedom Hosting II. Over 50% of the content on Freedom Hosting is associated with child pornography. Some

of the website data that was leaked might now be used to identify users of those websites. Security researchers estimated that Freedom Hosting II held 1500–2000 covert services [18]. This response may be seen as more ethical as it does not involve any deception or manipulation. It is also the swift takedown of illegal websites which is also an unpredictable approach. This means that online criminals cannot prepare in advance whereas they can with the honeypot trap methods. This response was also carried out by vigilante civilian groups, which means with enough funding, it should be easily incorporated into government responses.

In terms of addressing civil liberty issues resulting from policing the dark web, digital policing should allow for sites such as the New Yorker to continue functioning and not abolish the Tor network entirely. This enables freedom of speech and also improves human rights, for example, for those in countries such as China, which carries out extensive surveillance on the public and media to control its citizens. It is, therefore, immensely important that individuals governed by authoritarian governments have an untraceable output to share the horrors they face; otherwise, these would go undetected [51]. Instead of banning Tor networks, digital policing should aim to increase training and knowledge of how to combat anonymous users. For instance, it has been reported that the FBI has gained the ability to de-anonymise TOR servers, revealing identities and locations of illicit websites. This allows Tor to continue to run for legitimate users [52].

A further response to the challenges discussed in the previous section concerns using tactical cyber threat intelligence tools as a method to enhance digital policing operations. This makes it possible for organisations to tighten their general risk management strategies and adopt a position of proactive cybersecurity management. It supports a cybersecurity stance that is anticipatory, enables enhanced detection of advanced threats, and informs better decision-making to help LEAs to identify how illicit activities such as money laundering are taking place [53]. Learning how to predict cybercrime allows police to foresee how it may occur in the future and create ways to prevent it. In addition, further education on the use of cryptocurrencies and the dangers it poses is important, especially when online currency is growing at a rapid pace [54]. By preventing attacks in the first place, it means other technical, ethical, and legal challenges are also avoided to a large degree. It is important, therefore, to educate the public on how to keep their online information and banking safe through the use of firewalls and VPNs [55], for example, and through learning how to maintain anonymity. As a result, this could nullify the need that many civilians might feel to use Tor servers.

5 Conclusion

The study's main findings are that the dark web's nature of anonymity allows crime to thrive and go undetected. However, this characteristic also allows for a place where people are free to use the internet without being traced or watched over, which is an ongoing issue in society today, and therefore the TOR network is vital

in the preservation of civil liberties [18]. Finding the right balance between civil liberties and policing, proposes challenges for digital police when using methods such as honeypot traps, hacking devices and trying to target money launderers using cryptocurrency [21]. These methods often require digital police to follow seemingly unethical protocols with the intention of fighting crime, such as deception for example.

The recommendations suggested in this article have highlighted the need to improve regulations and training for officers looking to fight cybercrime, while bettering the techniques and knowledge of those trying to fight cybercrime on the dark web, in ways that allow the TOR network to continue to function [19]. However, the safety of the population is the top priority for police and fighting crime may be argued to have more importance than preserving an area of anonymity. There are many alternatives for anonymous sites and ways to whistleblow without exposing one's identity. Furthermore, cross-jurisdictional law enforcement co-ordination and international co-operation are essential. In addition, the results suggest that in order for digital policing to be effective, the techniques used must also be unpredictable to criminals. As shown in the takedown of Freedom Hosting II by Anonymous group hackers [18].

These findings are paramount to further research on how to improve digital policing with methods that result in less harm to the innocent and are more effective in the long term. Conducting research into past studies and literature helps to improve techniques and methods that will result in a reduction in crime and will also aid digital police in the arrest of dark web terrorists, paedophiles, and money launderers.

This study has limitations including the fact that the literature is of a restricted breadth, and although it presents solid recommendations to combat the challenges presented, more research on exactly how these can be implemented is needed. In the future, deeper research into the technical aspects of improving hacking and de-anonymising websites on the TOR browser is required, allowing digital police to formulate a methodology to carry out ameliorated plans. To fully understand the threat it poses, it is important to further examine extensive literature on cybersecurity and the workings of the dark web.

The reality of the dark web is very complex, authoritarian governments will continue their attempts to prevent access to the Dark Web, while liberal civil societies will continue to campaign that Tor remains unmonitored and unpoliced to defend free expression and privacy. Thus, a nuanced and balanced response is required to thwart illegal and unethical activities while preserving the advantages of the Dark Web's anonymity.

References

1. Tran NK, Sheng QZ, Babar MA, Yao L, Zhang WE, Dustdar S (2019) Internet of things search engine. Commun ACM 62(7):66–73
2. Kovalchuk O, Masonkova M, Banakh S (2021) The dark web worldwide 2020: anonymous vs safety. In: 2021 11th International conference on advanced computer information technologies (ACIT). IEEE, pp 526–530
3. Moggridge E, Montasari R (2022) A critical analysis of the dark web challenges to digital policing. Artificial intelligence and national security. Springer International Publishing, Cham, pp 157–167
4. Montasari R, Boon A (2023) An analysis of the dark web challenges to digital policing. In: Cybersecurity in the age of smart societies: proceedings of the 14th international conference on global security, safety and sustainability, London. Springer International Publishing, Cham, pp 371–383
5. Wilmot McIntyre M, Montasari R (2022) The dark web and digital policing. Artificial intelligence and national security. Springer International Publishing, Cham, pp 193–203
6. Staley B, Montasari R (2021) A survey of challenges posed by the dark web. In: Artificial intelligence in cyber security: impact and implications: security challenges, technical and ethical issues, forensic investigative challenges, pp 203–213
7. Lyon D (2007) Surveillance studies: an overview
8. Lavorgna A, Antonopoulos GA (2022) Criminal markets and networks in Cyberspace. Trends Organ Crime 1–6
9. Lewandowski D, Mayr P (2006) Exploring the academic invisible web. Libr Hi Tech
10. Moore D, Rid T (2016) Cryptopolitik and the darknet. Survival 58(1):7–38
11. Okyere-Agyei S (2022) The dark web—a review
12. Broadhurst R, Lord D, Maxim D, Woodford-Smith H, Johnston C, Chung HW, Chung HW, Carroll S, Sabol B (2018) Malware trends on 'darknet' crypto-markets: research review. Available at SSRN: 3226758
13. Li B, Erdin E, Gunes MH, Bebis G, Shipley T (2013) An overview of anonymity technology usage. Comput Commun 36(12):1269–1283
14. Bartsch R (2020) The relationship of drug and human trafficking and their facilitation via Cryptomarkets and the dark web: a recommendation for cryptocurrency regulation
15. Alghamdi H, Selamat A (2022) Techniques to detect terrorists/extremists on the dark web: a review. Data Technol Appl
16. Murty CA, Rughani PH (2022) Dark web text classification by learning through SVM optimization. J Adv Inf Technol 13(6)
17. Ahmad A, Maynard S (2018) The dark web as a phenomenon: a review and research agenda
18. Finklea K (2017) Dark web. Congressional Research Service, Washington, pp 1–19. https://fas.org/sgp/crs/misc/R44101.pdf
19. Chertoff M (2017) A public policy perspective of the dark web. J Cyber Policy 2(1):26–38
20. Adams DM (2016) The 2016 amendments to criminal rule 41: national search warrants to seize cyberspace, particularly speaking. Univ Richmond Law Rev 51:727
21. Ghappour A (2017) Searching places unknown: law enforcement jurisdiction on the dark web. Stanford Law Rev 69:1075
22. Hathaway OA, Crootof R, Levitz P, Nix H, Nowlan A, Perdue W, Spiegel J (2012) The law of cyber-attack. Calif Law Rev 817–885
23. Kittichaisaree K, Kittichaisaree K (2017) Cyber espionage. Public Int Law Cyberspace 233–262
24. Madison E (2014) News narratives, classified secrets, privacy, and Edward Snowden. Electron News 8(1):72–75
25. Tonkovich EA (1987) Survey of trends in search and seizure law
26. Brown I, Korff D (2014) Foreign surveillance: law and practice in a global digital environment. Eur Hum Rights Law Rev 3:243–251
27. Fondren E (2017) Snowden. Am Journalism 34(3):381–383

28. Omar ZM, Ibrahim J (2020) An overview of Darknet, rise and challenges and its assumptions. Int J Comput Sci Inf Technol 8(3):110–116
29. Committee to Protect Journalists (2015) Attacks on the press: journalism on the world's front lines. Wiley
30. Schneier B (2013) Attacking Tor: how the NSA targets users' online anonymity. The Guardian. https://www.theguardian.com/world/2013/oct/04/tor-attacks-nsa-users-online-anonymity
31. Holt TJ, Cale J, Leclerc B, Drew J (2020) Assessing the challenges affecting the investigative methods to combat online child exploitation material offenses. Aggress Violent Behav 55:101464
32. Wortley RK, Smallbone S (2006) Child pornography on the internet. US Department of Justice, Office of Community Oriented Policing Services, Washington, pp 5–2006
33. Bleakley P (2019) Watching the watchers: Taskforce Argos and the evidentiary issues involved with infiltrating dark web child exploitation networks. Police J 92(3):221–236
34. Blass T (ed) (1999) Obedience to authority: current perspectives on the Milgram paradigm
35. Miloshevska T (2019) Dark web as a contemporary challenge to cyber
36. Knaus C (2017) Australian police sting brings down paedophile forum on dark web. The Guardian. https://www.theguardian.com/society/2017/oct/07/australian-police-sting-brings-down-paedophile-forum-on-dark-web
37. Beshiri AS, Susuri A (2019) Dark web and its impact in online anonymity and privacy: a critical analysis and review. J Comput Commun 7(03):30
38. Bajaj K, Gochhait S, Pandit S, Dalwai T, Justin M (2022) Risks and regulation of cryptocurrency during pandemic: a systematic literature review. WSEAS Trans Environ Dev 18:642–652
39. Tewari SH (2021) Abuses of blockchain and cryptocurrency in dark web and how to regulate them (No. 4995). EasyChair
40. Westoll N (2019). Dark web vendor 'Mr. Hotsauce' busted for selling hard drugs to Canadians, RCMP say. Global News. https://globalnews.ca/news/4245971/mr-hotsauce-dark-web-drug-trafficking-rcmp/
41. Haasz A (2015) Underneath it all: policing international child pornography on the dark web. Syracuse J Int Law Commer 43:353
42. Bertola F (2020) Drug trafficking on darkmarkets: how cryptomarkets are changing drug global trade And the role of organized crime. Am J Qual Res 4(2):27–34
43. Hooper C, Martini B, Choo KKR (2013) Cloud computing and its implications for cybercrime investigations in Australia. Comput Law Secur Rev 29(2):152–163
44. Kadzik PJ (2015) Assistant Attorney General, Letter to Senator Charles E. Grassley, Chairman, Senate Committee on the Judiciary. Unclassified. National Security Archive. https://nsarchive.gwu.edu/document/22893-document-07-peter-j-kadzik-assistant-attorney
45. Allsopp R, Noto La Diega G, Onitiu D, Rasiah S, Thanaraj A (2019) Digital currencies: an analysis of its present regulation in the UK: a collaborative essay by NINSO, the Northumbria Internet & Society Research Interest Group
46. Turner NW (2014) The financial action task force: international regulatory convergence through soft law. NYLS Sch Law Rev 59:547
47. Pocher N, Veneris A (2022) Central bank digital currencies. In: Handbook on blockchain, pp 463–501
48. Morgan G, Finney C (2018) Initial coin offerings. The good, the bad, and the ugly. Retrieved from https://www.foxwilliams.com/uploadedFiles/FEATURE
49. Teichmann FMJ, Falker MC (2020) Money laundering via cryptocurrencies–potential solutions from Liechtenstein. J Money Laundering Control
50. Kerr OS, Murphy SD (2017) Government hacking to light the dark web: what risks to international relations and international law. Stanford Law Rev Online 70:58
51. Szadziewski H (2020) The push for a Uyghur human rights policy act in the United States: recent developments in Uyghur activism. Asian Ethn 21(2):211–222
52. Kumar A, Rosenbach E (2019) The truth about the dark web: intended to protect dissidents, it has also cloaked illegal activity. Financ Dev 56(003)
53. Saunders J (2017) Tackling cybercrime–the UK response. J Cyber Policy 2(1):4–15

54. Hawkins S, Yen DC, Chou DC (2000) Awareness and challenges of Internet security. Inf Manag Comput Secur
55. Abie H (2000) An overview of firewall technologies. Telektronikk 96(3):47–52

Assessing Current and Emerging Challenges in the Field of Digital Forensics

Zaryab Baig and Reza Montasari

Abstract This paper critically assesses the current and emerging challenges encountered in the field of Digital Forensics (DF) with reference to Cloud Forensics, the Internet of Things (IoT) Forensics, admissibility of digital evidence, lack of standardisation, limitations of tools in the field, as well as the significant problems associated with case backlogs within DF. Following the evaluation, the paper offers a set of recommendations that can be adopted to address or mitigate the stated challenges. To this end, a particular focus will be placed on the analysis of the recent report by His Majesty's Inspectorate of Constabulary and Fire & Rescue Services (His Majesty's Inspectorate of Constabulary and Fire and Rescue Services in Digital forensics: an inspection into how well the police and other agencies use digital forensics in their investigations, 2022) and applications of the recommendations discussed therein. This chapter contributes to the existing body of research with the inclusion of the recent recommendations made by the HMICFRS (His Majesty's Inspectorate of Constabulary and Fire and Rescue Services in Digital forensics: an inspection into how well the police and other agencies use digital forensics in their investigations, 2022) report, making this research relevant in relation to the current and emerging challenges faced in the field of DF.

Keywords Digital forensics · Cloud forensics · The IoT forensics · Digital investigation · Cybercrime investigation · Cybercrime · Digital evidence · Challenges · Recommendations

Z. Baig (✉) · R. Montasari
Department of Criminology, Sociology and Social Policy, School of Social Sciences, Swansea University, Swansea SA2 8PP, UK
e-mail: zaryabbaig1912@gmail.com

R. Montasari
e-mail: Reza.Montasari@Swansea.ac.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_8

1 Introduction

Due to the large digital footprint that now permeates every aspect of an Internet user's daily life, the likelihood of illegal activity leaving behind digital evidence has become increasingly high. According to The European Union Agency for Cybersecurity [1], 85% of all criminal investigations involve electronic devices and digital evidence underpinning almost all modern crime scenes. According to Locard's Exchange Principle, formulated by Edmond Locard, when two objects come into contact with one another, there is an exchange of materials between them. Subsequently, every criminal can be connected to a crime through trace evidence [2]. This evidence can therefore be used as forensic evidence. Since the early 2000s, the field of DF has gradually expanded and evolved with regulations [3]. As a result, the demand for forensic investigators has increased leading to an expansion of the academic education and certifications available in the DF domain.

The responsibility of a forensic scientist involves the important task of establishing facts related to questions such as what happened, how it happened, who was involved or when it took place. In order to answer these questions, a forensic scientist must use the digital evidence available to which different methodologies apply. However, there are current and emerging challenges in the field of DF due to the complexity of the tasks and the establishment of strict protocols in relation to legal requirements. Considering the significance of digital evidence in the investigation process within the majority of modern-day crime, this chapter aims to critically examine the current and emerging challenges faced in the field of DF. Following this analysis, the chapter also aims to offer associated recommendations that can be adopted to address or mitigate these challenges.

The remainder of this chapter is structured as follows. Section 2 provides a set of background information with a view to placing the study in context. Section 3 discusses and assesses several current and emerging challenges that DF investigators encounter in the field. Section 4 provides a set of recommendations to help address or mitigate these challenges whilst Sect. 5 discusses the findings. Finally, the chapter is concluded in Sect. 6.

2 Background

DF refers to forensic science applied to digital information. There does not currently exist an agreement in relation to a universal definition of the term. However, one of the most commonly used definitions is that proposed by Palmer [4], who defines DF as

> the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorised actions shown to be disruptive to planned operations'. [4]

Crime reconstruction is the establishment of the actions and events surrounding the commission of a crime [5]. It is used to determine a hypothesis about the sequence of events that led up to the crime and tests whether the hypothesis can be used as a possible explanation towards the crime. If the hypothesis is proved, it can be used, however, if it is refuted, a new hypothesis must be tried until a solution is found. Carrier and Spafford [6] outline the five-step process in conducting crime reconstruction. This involves evidence examination, role classification, event construction and testing, event sequencing and hypothesis testing. In order to test the hypothesis, investigators will need to utilise digital evidence which can be defined as the digital data that contains reliable information that can support or refute a hypothesis of an incident or crime [7].

Characteristics of good digital evidence are that it must be: admissible, authentic, complete, reliable and believable [8]. The admissibility of digital evidence refers to the collection of only the evidence that is allowed in court. Every digital investigation must be authorised, and an investigator must only investigate what they are authorised to do so. The evidence should also be believable and understandable in court to the laymen. This digital evidence should be acquired in a forensically sound manner. Forensic soundness refers to "the application of a transparent digital forensics process that preserves the original meaning of data for production in a court of law" [9]. Collecting digital evidence in a forensically sound manner denotes that any investigation of the evidence must be entirely reproducible by a third party. Regardless of who investigates the case, they should arrive at the same conclusion using forensic soundness. There are also two fundamental principles that must be considered with great importance when conducting DF investigation. These are "chain and custody" and "evidence integrity" [10]. Chain of custody requires investigator to document all actions that have taken place in relation to acquiring digital evidence. Evidence integrity refers to the preservation of evidence in a complete form without any intentional or unintentional changes. Whilst this is ideal in DF, it is not always achievable as data inevitably changes in live computer systems and networks during investigations.

The DF is employed in the public sector as law enforcement agencies are increasingly dependent on DF to process digital evidence in the context of the crime under investigation. This includes criminal cases involving identity theft, child pornography, terrorism, hacking and computer or insurance fraud. Both the public and private sector employ DF as a tool for supporting legal actions in criminal cases. DF is deployed in the investigation of various crimes such as human trafficking, employment disputes, counterfeiting and forgeries, internet financial and invoice fraud, and intellectual property theft [10]. In order to collect digital evidence to investigate the variety of crimes, different branches of DF must be used. These include mobile forensics, cloud forensics, network forensics, malware forensics, Internet of Things (IoT) forensics, drone forensics and car forensics.

DF is a rapidly expanding scientific field that continues to evolve as a result of the rapid advancement of technology. DF comprises of several branches that specialise in specific areas. Figure 1 represents examples of the sub-domains of DF.

Computer Forensics	Mobile Forensics	Cloud Forensics	IoT Forensics
Drone Forensics	Car Forensics	Internet Forensics	Database Forensics
Malware Forensics	Social Media Forensics	Multimedia Forensics	Email Forensics

Fig. 1 Different branches of DF

Cloud Forensics is the "application of scientific principles, practices and methods for organising events by identifying, collecting, storing, testing and reporting digital evidence" [11]. Due to its wide access for storage, economical solution and dynamicity, it has gained significant attention by the forensic experts. Network Forensics is a critical branch of DF in relation to, "computer network traffic monitoring and analysing it for data collection, legal artifacts, or detecting intrusion" [11]. Mobile Forensics is applied for "the extraction of digital evidence from portable and/or mobile devices" [12]. The processes in Mobile Forensics include seizure, acquisition, and examination. IoT Forensics deals with IoT related cybercrime and includes investigation of the devices, related sensors and data stored on possible platforms related to cybercrime [11]. IoT technologies include: "unmanned aerial vehicles (UAVs), smart swarms, the smart grid, smart buildings and home appliances, autonomous cyber-physical and cyber-biological systems" [13]. Whilst these branches of DF have significantly helped to advance the field, they have also posed some challenges discussed in the next section.

3 Challenges

3.1 Cloud Forensics

A significant challenge posed by cloud computing to forensic investigation is its flexibility and scalability [14]. As cloud computing is used in smart mobile devices, the data may be easily accessed everywhere, therefore posing a challenge to the investigators in locating the data whilst ensuring the privacy and rights of the user remains. Similarly, Tiwari et al. [11], identified the challenge of decentralised data, as cloud computing, 'permits data generation, storage, processing and distribution over different data centres, storage devices and physical machines' [11]. Due to this, law enforcement agencies, as recommended by Montasari and Hill [13], need to rely on local laws in order to conduct digital evidence acquisition. However, from a DF

perspective, this also poses a significant challenge as there is a discrepancy in the legal systems of different jurisdictions [13]. Another challenge in Cloud Forensics is that cloud-based applications allow users to access data from multiple devices. It is therefore difficult for an investigator to identify a possible source of change if, for instance, one of the two devices from a singular user were to be compromised [14]. Consequently, cloud-based environments can facilitate the identity and credentials theft since potential changes which would be used as evidence might remain unknown.

3.2 IoT Forensics

According to Casino et al. [12], IoT Forensics presents several challenges in relation to the "management of different streams of data sources", "the preservation and acquisition of evidence considering its volatility and value of data" and "the adoption of routine forensic tasks in the IoT ecosystem" [12]. They further note that data encryption and cryptographically protected storage systems are "one of the most significant barriers" for IoT Forensic investigators and it hinders efficient forensic analysis. Another significant challenge concerning IoT investigation include the legislative frameworks adopted in specific regions such as the GDPR in Europe [15]. This is supported by Montasari and Hill [13], who state that significant privacy and security challenges are posed by IoT-connected devices as they store personal data about individuals.

Security for secure operation of IoT-connected devices is also of paramount importance considering the fact that IoT devices present a wide range of attack surface. These cyberattacks can include: intercepting and hacking cardiac devices, "launching DDoS attacks using compromised IoT devices", and "hacking various CCTV and IP cameras" [13]. Montasari and Hill [13] note that, from a DF perspective, the extraction of evidential artefacts from IoT devices in a forensically-sound manner is a complex process and virtually impossible. The authors add that this is due to the proprietary hardware and software, data formats, the loss and overwriting of data and the spread of data across multiple devices and platforms. From a legal perspective, these issues are similar to those of traditional DF. However, from a technical perspective, further research is suggested in order to develop the capabilities of forensic tools. Forensic tools should be developed to support various IoT devices in the market. Further research is needed to resolve issues of ambiguity concerning network boundaries as well as the location of stored data.

3.3 Admissibility of Digital Evidence

There are legal challenges faced by DF in relation to the admissibility of digital evidence. Many cyber laws and regulations fail to consider the complexity involved

in collecting evidence. For example, the suspect's computer might include personal information that is crucial to an investigation. However, accessing this information is likely to be considered a violation of user privacy [16]. Written reports presented to law practitioners and judges should be readable in order to effectively communicate with these stakeholders. This is due to their potential lack of technical knowledge regarding the forensic tools and underlying technologies analysed [12]. The digital evidence should be authentic, accurate, complete and convincing to the juror. Following a proper chain of custody methodically and consistently also poses challenges to DF owing to the dynamic nature of digital evidence. The admissibility of evidence considers legal ramifications, and DF investigations approach needs an evaluation of the source of digital devices at the digital crime scene [17]. Another challenge concerns the acquisition of a "pre-search and post-seizure warrant that meet the required legal objective" [17]. This results from the lack of formalised procedures and stakeholder involvement. Therefore, it is essential to improve the reporting mechanisms in order to bridge the gap between law enforcement agencies, the judiciary and forensic investigators.

3.4 Lack of Standardisation and Limitations of Tools

Researchers in the field of DF have made efforts to agree on "formats, schema, and ontologies on DF artefacts" even though little progress has been made [13]. The issue of diversity problem arises due to the lack of cooperation between DF experts and DF researchers to create "standardised methods and guidelines to detect, acquire, store, examine, analyse and present digital evidence" [13], posing significant challenges to DF investigators. There is also a lack of standardised procedures used to facilitate forensic investigators in generating credible reports to be used in the court during legal proceedings [18]. This causes disparities in the way forensic reports are to be prepared, generated, and presented to various stakeholders. Thus, a standardised procedure is required, that has considered the processes defined within the existing investigation models, guidelines and standards. An example of these includes the ISO/IEC 27043 standard, which describes the processes relevant to different types of investigations.

In order to effectively extract and collect digital evidence, it is crucial to have relevant and adequate tools. DF tools are adopted in order to help investigators to conduct the four stages of a DF process. The extraction stage of the process requires the encrypted data to be extracted. Due to the vast amount of data stored in modern digital devices, this process usually results in the extraction of raw data. However, the current tools and techniques used in DF are limited in their functionality and inadequate in the identification of data which is highly complex or subtly modified [13]. As new technologies emerge, traditional DF tools, techniques and methods are unable to address the challenges presented by these technologies. Whilst existing DF tools may be able to handle cases with many terabytes of data, they are unable to compile terabytes of data into a clear report. Because of the inadequacies of the

current DF tools to conduct data analysis, event and timeline reconstruction is often conducted manually during a DF investigation.

3.5 Backlog of Digital Forensic Cases

The emergence of new technologies has presented law enforcement agencies with significant challenges in relation to a backlog of DF cases [19]. This is caused by the increase in the volume of evidence data which requires processing within modern forensic cases. Newer technologies allow higher storage capacity which law enforcements and legal systems are lagging behind. The time and effort needed to extract and analyse this data has also become increasingly difficult due to the absence of appropriate forensic tools. The backlog of DF cases has a wide range of implications including the suppression of evidence, delays in prosecution, and law enforcement prioritising high profile cases [20]. The delays in prosecution could enable criminals to commit additional offences, potentially impacting the public trust and confidence in the legal system.

A recent report by His Majesty's Inspectorate of Constabulary and Fire and Rescue Services [21], "An inspection into how well the police and other agencies use digital forensics in their investigations", addresses and explores this backlog in detail. The report elucidates that, in some forces, senior leaders did not completely understand the current demand for DF and that not enough forces have the right number of trained and equipped staff to meet the demand and address the backlog. The DF backlog problem is evidently a significant challenge posed to DF and there needs to be a collaboration amongst all sectors of law enforcement to address this issue urgently and efficiently.

4 Recommendations

To deal with the aforementioned challenges posed by IoT Forensics, it is recommended that there should be a collaboration between the government, industry, and academia in order to develop a robust IoT framework. Cloud cybersecurity may also be reviewed as IoT devices produce data which is stored in the Cloud [13]. It is important that cloud cybersecurity policies be blended with IoT infrastructure in order to produce timely responses in response to suspicious activities. Montasari and Hill [13] propose the idea of "developing new investigation methods which can track and filter the transfer of data across IoT-connected devices", supported by [22]. This recommendation may also apply to the challenge that cloud-based applications allow users to access data from multiple devices. The development of new investigation methods would facilitate investigators in identifying possible sources of change between two devices. Cloud storage can also be used instead of traditional storage

methods in DF as it improves confidentiality, integrity as well as availability of the digital evidence. It is also less expensive than using external drive for storage [12].

To address the limitation of tools, Beebe [23] argues that there should be communication between forensic practitioners and DF tool developers in order to ensure the technological needs and requirements regarding forensic tools are fulfilled. This would allow forensic practitioners to effectively investigate crime through their possession of adequate tools. This is an appropriate solution given that the use of inadequate tools to analyse data has led to the use of manual reconstruction of events during a DF investigation, which may also be impacting the DF backlog problem. The challenges of standardisation and admissibility of digital evidence can be addressed using similar recommendations. The lack of standardised procedures to help forensic investigators to produce succinct reports to be used in the court affects the admissibility of digital evidence. Both challenges can therefore be addressed by improving the reporting mechanisms and it would also be beneficial to use a common standardised framework. If this framework is adopted, it would be vital for all staff included in the DF process to be trained and equipped for it to be properly understood.

The HMICFRS [21] report on the investigation into how well DF is used by the police and other agencies includes recommendations on how to address the issues of DF case backlogs. These recommendations may also be used as indication of what may be causing the other challenges explored in this research. The report outlines nine recommendations, which include that by April 2023, the National Police Chiefs' Council (NPCC) should appoint a dedicated lead for DF, the Home Office should review DF budgets and funding, and the College of Policing should ensure all its digital courses focus on investigations and victims' needs. By June 2023, the NPCC lead for DF, the Home Office and relevant support services should provide guidance to all forces on the use of cloud-based storage and computing power. By September, the NPCC and all forces within England and Wales should include the management of DF kiosks within their governance and oversight frameworks. By December 2023, each force in England and Wales should better understand the demand for DF services. By April 2024, the NPCC should encourage an increase in the number of trained digital media investigators. Finally, by November 2024, chief constables should integrate DF services under existing forensic science structure, and the Home Office should work with NPCC, the College of Policing, and the private sector to design an operating model which would assist DF services with police investigations [21].

5 Discussion

This paper examines the current and emerging challenges facing the field of DF. Some of the discussed challenges of the lack of standardisation and the admissibility of evidence might be addressed in 2024 as the HMICFRS [21] report outlines plans to design an alternative operating model. This challenge is therefore supported and recognised as significant to the field of DF. The limitations of tools used in DF investigation to extract evidence causing challenges to the field of DF should be

addressed using the additional funding provided in 2024. This could assist with extracting evidence in a forensically sound manner, potentially without falling behind the capabilities of new emerging technologies. This will, in turn, assist with ensuring a successful prosecution. As aforementioned in the chapter, many cyber laws currently fail to consider the complexity involved in collecting digital evidence. However, as recommended in the HMICFRS report, if all forces within England and Wales include the management of DF kiosks in their governance and oversight frameworks, the importance of recognising the complexities involved in collecting digital evidence may be identified and implemented in future cyber laws as all forces would be involved. The funding could also allow for the development of a robust IoT framework through collaborations between the government, industry, and academia. This would also help to address the issues of DF case backlogs as if there is a development of improved frameworks within IoT, and cases involving the latest technology can be more easily investigated.

6 Conclusion

This chapter explored the use and application of DF discussing its key principles such as chain of custody and evidence integrity. The challenges facing DF were examined followed by recommendations to address or mitigate these challenges. Cloud storage was recommended instead of traditional storage methods in order to address cost issues and to improve data confidentiality. Recommendations were also made with regards to issues resulting from the backlog of DF cases. These recommendations provide directions for further research such as utilising the funding provided in 2024 to improve the tools employed in DF investigation methods. Collaborations between law enforcement agencies and DF experts are encouraged with a view to exploring the design and development of specific tools required.

References

1. The European Union Agency for Cybersecurity (2021)
2. Thornton JI, Peterson J (1997) The general assumptions and rationale of forensic identification. In: Modern scientific evidence: the law and science of expert testimony, vol 2, p 13
3. Pollitt M (2010) A history of digital forensics. IFIP international conference on digital forensics. Springer, Berlin, pp 3–15
4. Palmer G (2001) A road map for digital forensic research. In: First digital forensic research workshop, Utica, New York, pp 27–30
5. Chisum WJ, Turvey BE (2007) A history of crime reconstruction. In: Crime reconstruction, pp 1–35
6. Carrier B, Spafford E (2004) An event-based digital forensic investigation framework. Digit Invest
7. Stoykova R (2021) Digital evidence: unaddressed threats to fairness and the presumption of innocence. Comput Law Secur Rev 42:105575

8. Brezinski D, Killalea T (2002) Guidelines for evidence collection and archiving (No. rfc3227)
9. McKemmish R (2008) When is digital evidence forensically sound? In: IFIP international conference on digital forensics. Springer, Boston, pp 3–15
10. Årnes A (ed) (2017) Digital forensics. Wiley
11. Tiwari A, Mehrotra V, Goel S, Naman K, Maurya S, Agarwal R (2021) Developing trends and challenges of digital forensics. In: 2021 5th International conference on information systems and computer networks (ISCON). IEEE, pp 1–5
12. Casino F, Dasaklis TK, Spathoulas G, Anagnostopoulos M, Ghosal A, Borocz I, Solanas A, Conti M, Patsakis C (2022) Research trends, challenges, and emerging topics in digital forensics: a review of reviews. IEEE Access 10:25464–25493
13. Montasari R, Hill R (2019) Next-generation digital forensics: challenges and future paradigms. In: 2019 IEEE 12th international conference on global security, safety and sustainability (ICGS3). IEEE, pp 205–212
14. Zuhri FA (2019) The illusion of the cyber intelligence era. ZAHF.ME
15. Stoyanova M, Nikoloudakis Y, Panagiotakis S, Pallis E, Markakis EK (2020) A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues. IEEE Commun Surv Tutorials 22(2):1191–1221
16. Carrier BD, Spafford EH (2006) Categories of digital investigation analysis techniques based on the computer history model. Digit Invest 3:121–130
17. Yeboah-Ofori A, Brown AD (2020) Digital forensics investigation jurisprudence: issues of admissibility of digital evidence. J Forensic Legal Invest Sci 6(1):1–8
18. Karie NM, Kebande VR, Venter HS, Choo KKR (2019) On the importance of standardising the process of generating digital forensic reports. Forensic Sci Int: Rep 1:100008
19. Homem I (2018) Advancing automation in digital forensic investigations. Doctoral dissertation, Department of Computer and Systems Sciences, Stockholm University
20. Brandt J, Wärnling O (2020) Addressing the digital forensic challenges within modern law enforcement: a study of digital forensics and organizational buying behavior from a DF-company perspective
21. His Majesty's Inspectorate of Constabulary and Fire and Rescue Services (2022) Digital forensics: an inspection into how well the police and other agencies use digital forensics in their investigations
22. Hegarty R, Lamb DJ, Attwood A (2014) Digital evidence challenges in the internet of things. In INC, pp. 163–172
23. Beebe N (2009) Digital forensic research: the good, the bad and the unaddressed. In: Advances in digital forensics V: fifth IFIP WG 11.9 international conference on digital forensics, revised selected papers 5, Orlando, Florida, USA. Springer Berlin Heidelberg, pp 17–36

A Critical Analysis: Key Strategies of Far-Right Online Visual Propaganda

Nina Kelly

Abstract The approach from the far-right in producing and disseminating visual propaganda has allowed for a persistent online presence to be maintained, despite efforts to remove extremist and hateful content. This chapter will critically explore the academic literature which considers how far-right actors are taking advantages of the affordances of online communication routes to spread visual propaganda. Three key strategies which emerge from the literature will be critically discussed to understand the role of visual imagery in facilitating and maintaining far-right online discourse. Firstly, the use of imagery to other out-groups through boundary construction will be considered. From understanding how such representations drive online engagement, the second strategy of image and information manipulation will be discussed. This strategy will consider how the far-right take advantage of social media systems to garner more visibility through manipulating and framing imagery. Lastly, considerations will be made towards how humour through meme images and board subculture have been used as a strategy to lower the boundary for the participation in extremist ideology.

Keywords Online propaganda · Far-right · Online radicalisation · Extremism · Online discourse · Visual propaganda

1 Introduction

Social media is a well-established tool for communication. However, the opportunity for direct spread of information and the low threshold for participation allows social media to be used as a tool by actors who intend to spread malicious agendas and harmful ideology [1, 2]. The UK government has consistently recognised the complexity in the harms of this usage, with 2018 Prime Minister Theresa May

N. Kelly (✉)
School of Social Sciences, Swansea University, Singleton Park, Swansea SA2 8PP, UK
e-mail: nina.devi.kelly@gmail.com
URL: http://www.swansea.ac.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_9

declaring the need for technology companies to move more swiftly to remove terrorist content, and further governmental reports identifying the significant role of the internet in the radicalisation process [3, 4].

Academic literature has highlighted how the persistent and flexible response from the far-right toward attempts to limit their online influence has allowed them to maintain a persistent online presence [5]. Research into the strategies behind far-right online propaganda has flourished in an attempt to understand the strategies that enable this, however fewer studies focus on the visual aspect as opposed to the textual aspect [6, 7]. When considering the meaning of propaganda, the following definition as developed by Jowett and O'Donnell [8] will be adopted for this chapter and directed towards visual imagery: "*the deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behaviour to achieve a response that furthers the desired intent of the propagandist*" ([8], p. 7). The use of the term 'far-right' throughout this chapter will be used as an umbrella term to encompass the range of groups and individuals which are driven by the rhetoric that Western civilisation is threatened by non-native people or ideas and anti-elitist sentiment [9].

This chapter will draw together three key strategies arising from the literature which will be used to critically discuss the key online visual propaganda strategies employed by the far-right, with effectiveness being considered as key in terms of responses and engagements as referenced in the previously mentioned definition of propaganda by Jowett and O'Donnell [8]. Firstly, the construction of otherness in an online context will be critically discussed to identify the underlying mechanisms of nativist discourse and boundary construction which drives engagement with propaganda. Secondly, the use of mis and disinformation through imagery and the engagement it fosters through the manipulation of emotions and algorithms will be considered. Lastly, the culture of memes and humour will be discussed and how harmful rhetoric can be inconspicuously spread to mainstream platforms.

2 Construction of Otherness Through Imagery

The online space in which far-right groups occupy is sometimes referred to as the extreme right-wing sphere [10, 11]. However, this description lacks in applicability to an online context with alternative terms such as 'far-right ecosystem' offering a more conceptualised definition [12]. Baele et al. [12] use this framework to identify four elements in increasing analytical depth: whole network (the macro level which constitutes the elements), biotopes (groups which share a sub-identity), communities (collectively linking entities), and entities (such as a social media platform). Understanding the online space as a multidimensional framework allows insight into the journey of far-right propaganda from the point of dissemination to wider communities. This section will discuss the initial point of entry for this type of propaganda.

The transition of far-right ideology into the online medium has been greatly benefitted by far-right ideas growing to be a part of the political mainstream [13, 14]. Prominent far-right issues such as terrorism, immigration, and crime have become a focus in mainstream politics over the last two decades, with language such as the 'war on terror' and an unwavering focus on 'illegal asylum seekers' contributing to normalising and legitimising far-right ideas [14]. Across the West, far-right political parties have rallied electoral support through these issues, further legitimising their ideas in the mainstream [15]. This trend became known as the rise of far-right populism—a political ideology combining right-wing politics and anti-elitist sentiment as defined in a European Parliament Report [15]. This rise in far-right involvement in politics has been extensively covered from political, academic, and journalistic perspectives, with the excessive use of the term "far-right populism" also being seen as problematic.

Muddle [16] argues that the term 'far-right populism' takes the emphasis away from the core of radical right ideology, with discourse analysis focusing on populism in the past decade supporting this concern [17]. Brown and Mondon [17] also highlight several concerns over the uncritical use of the term when used generally throughout literature. Effects such as disproportionate coverage and the deflection of responsibility onto communities with less agenda setting power than media actors and politicians were recognised as problematic. Brown and Mondon [17] also suggest that another detrimental effect stemming from the use of the term 'far-right populism' is the trivialisation and blurring of racism and nativist narrative, which is a concern also echoed by Newth [18]. In the context of far-right online propaganda, understanding this theme of nativism and othering as a strategy will provide a foundation for understanding how far-right actors apply further key strategies through visual propaganda.

With nativism being an ideological pillar of far-right politics and discourse, research has illustrated how the othering and stigmatisation of minority groups have spread through online visual propaganda [6]. Nativism is commonly defined as the negative portrayal of culturally deviant outgroups while positively emphasising one's own culture and traditions—a combination of nationalism and xenophobia [16]. However, this definition has been criticised by Newth [18] as requiring a certain level of nuance as to not be used as a euphemistic term for racism. Newth [18] also emphasises the importance of understanding nativism as a discourse rather than an ideology to gain an insightful understanding of the wider social and political context in which it is present. This point will also be considered though this chapter as nativism is considered in the wider discourse surrounding online propaganda.

When seeking to promote nativist discourse as a strategy, the construction of boundaries is used to create a desirable and exclusive in-group and other out groups such as immigrants, who are perceived to threaten these exclusive ideas [6]. In the context of far-right populism, this is typically done to propagate nationalist group identities and mobilise a sense of solidarity within their audience through the medium of memes, photographs, and visual symbols [6, 19].

Engagement with this visual propaganda comes from both the in-group constituents who bond through the anti-immigration narrative, and the wider mainstream audience who engage in the discourse [19]. Against this backdrop, the capacity of visuals to address these audiences through their open-ended characteristics allows for a calculated ambivalence rhetoric strategy to further nativist discourse [20]. Studies adopting a multimodal visual analysis demonstrates how this is done online through categorising visual content and analysing audience engagement [21]. Awad et al. [6] use this methodology to analyse how boundaries are constructed by the Danish People's Party using 1120 images posted on their Facebook (now Meta) page between the period 2012 and 2020. 45.8% of the images were categorised as positive in-group representation of the party with images promoting the party leaders beside fellow "ordinary" Danes, characterising a warm and friendly approach within the in-group [6]. In contrast to this, the "threat" category was second largest at 13.6%, with 89 images being used to other Muslim migrants [6]. Othering was frequently presented by using the headscarves of Muslim women as a symbolic boundary to distance them from the in-group, with the continuous use of images over time containing the headscarf garnering an increase in engagement and visibility over the period of study [6] (Fig. 1).

Although this study focused on the Danish People's Party, the use of the headscarf as a symbolic boundary has also been seen echoed throughout far-right parties globally, such as the Alternative for Germany (AfD). AfD's Facebook page has been seen to consistently utilise headscarves, niqab, and burka as symbolic images to construct an image of Muslim women being voiceless, passive, and oppressed—threatening the freedom of German women [7, 22]. Far-right parties have also used this construction to distance and contrast their in-group from these portrayals, which can be seen

Fig. 1 Refugees shall not integrate into Denmark! They must go home! *Source* [6]

Fig. 2 PVV image first posted on Twitter. *Source* [7]

in Fig. 2 from the Netherlands Partij voor de Vrijheid (PVV) [7]. By having freedom represented by Dutch women, it implicitly references the contrast between representations of oppressed Muslim women, aiding their long-term goal to shrink the cultural independence of residing Muslims [7].

This strategy of boundary construction through symbolic imagery is also regarded as effective in garnering engagement and mobilising followers [6]. Images othering migrants and Muslims accumulated the most reactions and second most shares on Facebook in the analysis carried out by Awad et al. [6]. Although this study is limited as the category of reactions (such as sad, angry or happy) were not recorded, it demonstrates the attractive strategy of using boundary construction and nativist discourse to garner attention and engagement online. This engagement leads to further online visibility and more individuals participating with the propaganda, whereby the impact of participatory propaganda allows far-right groups to influence perceptions through iconic boundary construction [23].

3 The Manipulation of Imagery

A second key strategy employed in far-right actors' use of visual propaganda takes advantage of the potential visibility of mainstream platforms and the vulnerabilities within the news and media ecosystem through the spread of misinformation [24]. The term "misinformation" can be used interchangeably with "propaganda" in respect to a misleading message or far-right narrative being spread under the appearance of informative content [25]. Debates surrounding the truthfulness of political claims grew in the wake of the 2016 U.S. elections, where a vast amount of news claims were exposed as inaccurate or incorrect [26]. This breakdown of trust in democratic institutions allowed far-right actors to take advantage of the damaged media ecosystem

to propagate their own ideas, leading to them being identified as one of the primary creators and distributors of manipulated information [27–30]. Using visuals to spread online misinformation also comes with the benefit of images being perceived as a more direct image of reality compared to text [31]. In addition to photo manipulation tools being easily accessible, images therefore have the potential to be powerful tools for spreading misinformation [26].

In view of this, Klein [30] conducted a multimodal analysis of 342 images posted by the British National Party between 2017 and 2018. Results illustrates that the majority of visuals (52.32%) were coded as fallacious—with high facticity but misleading with an exaggerated partisan frame [30]. The second most common category was factual at 37.13% [30]. The type of user interaction with these categories were also recorded, with users' engagement with fallacious images showing more angry (😡)responses and an increased number of comments compared to factual or funny posts [30]. The manner in which imagery is constructed is can also be attributed to this higher level of engagement, as seen below in Fig. 3. The pained expression on "Lilly's" face contrasting with the masses of less visible immigrants with a black and white filer aims to evoke a sense of compassion from their audience by constructing the image with strong nativist rhetoric [30]. The use of darker colours to elicit automatic negative connotations through cognitive analysis has also been identified in literature [32, 33].

In the study from Klein [30], multimodal analysis was used as the method to study social media content. This method has merit due to the dependent relationship between visual and text statements, as by themselves they are deemed less effective with less context [34]. However, studies which seek to understand how credible multimodal disinformation is perceived suggests that this is only true for certain topics [26]. An online survey sample of 1404 participants were used to assess perceived credibility of textual and multimodal disinformation [26]. Multimodal disinformation (consisting of an image and text) was perceived as significantly more credible for the

Fig. 3 Example of a fallacious image with a partisan frame. *Source* [30]

topic of refugee involvement in terrorism compared to sole text alone [26]. However, in the second topic of school shootings, this difference was not significant, which suggests a future area for research in the context of far-right social media content to assess what topics of mis and disinformation are perceived as most credible.

Attempts to combat this growth of mis and disinformation in mainstream social media has emerged through fact checkers, with evidence suggesting that simple and factual rebuttals can be effective in countering false information [35]. Hameleers et al. [26] consider the use of textual fact checkers in the rebuttal of multimodal disinformation and textual disinformation, with results demonstrating that the impact on both conditions is similar across the topics of refugee terrorism, school shootings, and climate change. However, challenges are presented when considering that the followers of radical far-right actors or groups on social media view news stories that match with their beliefs. Literature suggests that under the elaboration likelihood model (ELM) of processing, the peripheral route will be used to analyse the misinformation if shared beliefs exist [33, 36]. The peripheral route offers less scrutiny and a low level of elaboration compared to the central route, which involves a high level of elaboration and scrutiny [33, 37]. Therefore more motivation for critical analysis is used when processing information which does not match prior beliefs [33, 37]. Considering the number of followers of far-right pages or profiles online, there is a high risk for misinformation to go unobserved in the far-right ecosystem.

The use of emotional manipulation as shown in Klein's [30] study and Fig. 3 to increase engagement through visual misinformation is also well supported by literature [38, 39]. Bakir and McStay [40] contribute through an economic model of emotion which demonstrates how empathetic media generates increased attention and views which can be converted into revenue, increasing the incentive to disseminate misinformation. This mutually benefits both far-right actors through visibility, and the platforms such as Facebook themselves who gain advertising revenue from increased engagement [41]. This leads to a complex issue where calls for change around the social media algorithms which feed off dis/misinformation are being made, with concerns that the business structure of Facebook's platform itself is increasing radicalisation and political polarisation [41]. Although Facebook claims to filter out disinformation through the use of AI filters, there are exhaustive examples which contradict these claims, and criticism which points towards AI needing to also consider ethics in order for filters to be more effective in detecting mis/disinformation [41, 42].

It is also important to consider how the propaganda reaches mainstream platforms in the first instance. Examinations into the journey of information through the far-right and mainstream media eco-systems offer an insight into the mechanisms which allow the strategy of spreading misinformation to be so effective [43]. Studies have shown how trading up the chain allows for information to be deliberately propagated from the alt-right imageboard forum 4chan to far-right mainstream media actors [24].

Krafft and Donovan [43] examined which strategies were used in a case study of a disinformation campaign. This campaign took place after the 2017 Charlottesville Unite the Right Rally, where a white supremacist deliberately drove a car into counter protesters, killing activist Heather Heyer and injuring dozens more [44]. The disinformation campaign began prior to the public release of the attacker's identity and took place against a left leaning "Joel V." who was not present at the event and therefore falsely accused [43]. Images were identified as a key strategic element within this campaign, with the crowdsourced investigation on 4chan being triggered by an image of the attacking vehicle that included a licence plate linked to Joel through social media posts [43]. As more information was gathered through Joel's public photos identifying him as outspoken and left leaning, these images were being used to create a collage [43]. This evidence college comprised of screenshot images was noted as the key strategic element in the spread of this disinformation campaign spreading to alterative conservative media websites and Twitter, with the screenshots offering independent verification through public information that can be found through search engines, thus increasing credibility [43]. The omission of any references to 4chan as the origin of these images also decontextualised and obscured the source of the narrative—a strategic move which takes further advantage of the lack of journalistic process and factual checks, illustrating the importance of fact-checking [25, 43].

4 Constructing Hate Through Humour and Visual Memes

In addition to fringe sites manipulating information for the mainstream, another key strategy of visual propaganda lies within the /pol/ board subcultures of 4chan and 8chan (subsequently 8kun). This final key strategy which will be discussed uses humour and extremist meme images to spread racist and hateful speech while communicating white-supremacist narratives. Memes are characterised as images, videos, and texts (or combinations of all) which use humour to spread ideas [45]. In the context of the far-right imageboard chan culture, they are seen as an important aspect of communication, as the humorous element is used to mask the overtly racist narratives thus lowering the boundary for participation in the extreme ideology [24].

The function of memes within far-right imageboards is initially suggested to be humour—to be evocative and make fun of politics while also creating inside jokes which bonds the in-group by visualising stereotypes of shared enemies [10, 13, 14]. However, this subversive humour contributes towards the normalisation of countercultural opinions, creating a space which can facilitate radicalisation [13]. This potential has been observed in case studies from collected Discord logs released by a left-wing media organisation, where one user comments that his involvement in white supremacy started by being entertained by Nazi memes on the /pol/ board [46, 47]. Further evidence suggesting that the culture behind these memes has a role in radicalisation comes from incident case studies. The 2022 Buffalo shooter had featured dozens of memes within his manifesto, and also references the Nazi ideology board as the point at which he engaged in racist hate speech in Discord

Fig. 4 Example of the "Yes Chad" meme. *Source* [52]

chat logs [10, 48]. When considering this and the impact of memes in creating inside jokes and group bonding, an alignment with elements of belonging and in-group cohesion as identified in classical theories of radicalisation can be seen [45, 49, 50]. This suggests that there is potential for a causal link between board meme culture and far-right extremism [45, 49, 50].

However, gaining a deeper insight into how visual propaganda relates to radicalisation through meme culture presents numerous challenges. Firstly, there is the disadvantage of needing to catch-up to trends [51]. Increasing amounts of far-right campaigners and actors have moved over to alt-tech platforms such as 8kun and Gab in recent years to facilitate their use of extremist rhetoric [51]. Consequently, response systems have not been effective, with trend predictions for new technology and communicative strategies being too slow to predict new meme trends [51].

A second issue pertaining to the study of image based memes is the level of digital literacy required to correctly understand the nuances conveyed by the images [45]. This is needed due to the malleable nature of images where the extremist nature is not overt, and therefore requires a certain level of familiarity with chan culture to uncover the underlying context [45, 52]. An example of this can be seen in Fig. 4—the Yes Chad reaction meme which was the most observed images across all chan sites [52]. Although appearing as innocuous, it was noted as originating on 4chan as part of a series which compared various race and nationalities—primarily Nordic and Mediterranean [52]. Throughout the site, it was observed as a representation of the white face of /pol/ culture and used to affirm racist statements [52].

Figure 5 displays the mobilisation of Yes Chad in another form where the overt context is in the perceived acceleration of a race war in the US following Black Lives Matter protests [52]. This is just one example of how many innocuous meme images can propagate extremist values, and demonstrates the challenges faced by researchers and policymakers in understanding the underlying context.

However, digital literacy is not only needed for researchers, but the youth who are also at risk of being targeted by meme visuals from the far-right [11, 52, 53]. With memes being part of a wider pop-cultural aesthetic that allows them to be easily consumed and shared, instances of teenagers being targeted with images of extreme neo-Nazi propaganda through mainstream social media such as Instagram have come to light [54, 55]. Furthermore, as the editor of the neo-Nazi website Daily Stormer confirmed that memes were being used as a strategy to indoctrinate children

Fig. 5 Example of the "Yes Chad" meme glorifying accelerationism. *Source* [52]

as part of the website's design, it creates a further need for the youth to be equipped with the critical thinking skills to immunise themselves against media with extreme undertones [11, 56].

Recommendations for a collaborative approach between NGOs and educators have been suggested to allow non-state actors to reach the target audience through classroom workshops which critically analyses political Internet memes [57, 58]. However, although an approach by a German NGO named "The Fair Skills approach" was made and referred to as recently being implemented in three European countries in Pauwel ([57], p. 8.), publications from the project's researchers and the project website itself shows the most recent iteration of the project (CEE Prevent Net) ended in 2020 [59, 60].

In terms of preventative methods, effort to reduce the spread of this type of extreme visual propaganda is challenging. Far-right actors have a significant advantage by creating a transnational, cross-ideological allegiance, and the use the of chan boards subculture in creating ingroup solidarity, planning terrorist attacks, and disseminating the violent images from the attacks through other platforms without having their pathways disrupted [51, 61, 62]. The pathways in which images are spread throughout the far-right ecosystem are also severely complex. One study used a novel image processing pipeline and cluster analysis to provide a multi-platform measurement of the meme ecosystem, analysing the connections between 160 million images [63]. 4chan's /pol/ was identified as having the largest influence over racist and political memes, but the information pathway led to the fringe subreddit r/The_ Donald as having the most success in pushing the most memes into both the fringe and mainstream [63].

The implications of these flexible networks are plentiful, as through these networks, content will re-emerge on different platforms and take advantage of external file hosting websites to keep propaganda accessible [63]. A case study which demonstrates this in action is the 2019 Christchurch terror attack, where the attacker posted an outlink to a Facebook livestream of the attack on 8chan's /pol/ board, in which meme references were made throughout [64]. With 8chan users archiving and resharing the footage to YouTube and Reddit, the visual propaganda was being engaged by masses of individuals, with the consequences of this resulting in further

copycat attacks [64]. This case study further exacerbates the need for a multi-level approach involving governments, private entities, and social media firms to disrupt the pathways which draws individuals into chan culture through disturbing meme humour and the amplification of harmful terrorist propaganda [13]. As evidenced through the aforementioned case studies, it is apparent the success of this strategy lies within the lack of a response to effectively disrupt these information pathways.

5 Conclusion

While discussing the key strategies employed in far-right visual propaganda, a holistic manner has been used to highlight the intricacies of the issues and debates surrounding the far-right ecosystem, the media ecosystem, and the social media ecosystem. A key pillar of far-right ideology was observed as nativism and the ways in which it was used to create discourse through social media posts of far-right populist parties. Mechanisms such as othering were seen to use visuals to explicitly show who the desirable and elite in-group are by representing them in happy scenes of nature and sunlight, while also constructing boundaries by using headscarves as icons of oppression and a threat to the nation's free values. Understanding these representations led to a second key strategy to be understood—the manipulation of visual imagery and information.

By framing images in a misleading way, far-right actors were seen to construct emotive and partisan propaganda which took advantage of Facebook's engagement systems, garnering more attention on fallacious images an opposed to factual. The wider harms behind mis and disinformation in the far-right ecosystem was also considered through the disinformation campaign in relation to the attack during Charlottesville rally, where a lack of fact-checking mechanisms enabled the deliberate spread of disinformation.

The last key strategy of visual meme humour strategically complicates visual propaganda by assigning extremist connotations to widespread innocuous images. The harm from this is plentiful, with individuals being drawn in by the humour and potentially being led down the path of radicalisation. Also, the deliberate targeting of children through this form of visual propaganda and the complexity of digital literacy required to understand the subculture also poses issues when formulating responses to the threat. The lack of a coordinated response for all three of these strategies allows for violent propaganda footage of devastating terrorist attacks to be deliberately spread throughout fringe and mainstream platforms at the will of terrorists, signalling the need to combat these effective strategies from far-right actors.

References

1. Krämer B (2017) Populist online practices: the function of the Internet in right-wing populism. Inf Commun Soc 20(9):1293–1309. https://doi.org/10.1080/1369118X.2017.1328520
2. Schwarzenegger C, Wagner AJ (2018) Can it be hate if it is fun? Discursive ensembles of hatred and laughter in extreme right satire on Facebook. Stud Commun Media 7:473–498. https://doi.org/10.5771/2192-4007-2018-4-473
3. Kenyon J, Binder J, Baker-Beall C (2021) Exploring the role of the internet in radicalisation and offending of convicted extremists (Ministry of Justice Analytical Series). HM Prison and Probation Service. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1017413/exploring-role-internet-radicalisation.pdf
4. May T (2018) Theresa May's Davos address in full. World Economic Forum. https://www.weforum.org/agenda/2018/01/theresa-may-davos-address/
5. Conway M, Scrivens R, Macnair L (2019) Right-wing extremists' persistent online presence: history and contemporary trends. ICCT Policy Brief. https://doi.org/10.19165/2019.3.12
6. Awad S, Doerr N, Nissen A (2022) Far-right boundary construction towards the "other": visual communication of Danish People's Party on social media. Br J Sociol. https://doi.org/10.1111/1468-4446.12975
7. Sayan-Cengiz F, Tekin C (2022) Gender, Islam and nativism in populist radical-right posters: visualizing 'insiders' and 'outsiders.' Patterns Prejudice 56(1):61–93. https://doi.org/10.1080/0031322X.2022.2115029
8. Jowett GS, O'Donnell V (2012) Propaganda and persuasion. SAGE
9. Intelligence and Security Committee of Parliament (2022) Extreme right-wing terrorism (HC 459). House of Commons. https://isc.independent.gov.uk/wp-content/uploads/2022/07/E02710035-HCP-Extreme-Right-Wing-Terrorism_Accessible.pdf
10. Abbas T, Somoano I, Cook J, Frens I, Klein G, McNeil-Willson R (2022) The buffalo attack—an analysis of the manifesto. International Centre for Counter-Terrorism. https://icct.nl/publication/the-buffalo-attack-an-analysis-of-the-manifesto/
11. Farinelli F (2021) Conspiracy theories and right-wing extremism—insights and recommendations for P/CVE (radicalisation awareness network). European Commission. https://home-affairs.ec.europa.eu/system/files/2021-04/ran_conspiracy_theories_and_right-wing_2021_en.pdf
12. Baele S, Brace L, Coan T (2020) Uncovering the far-right online ecosystem: an analytical framework and research agenda. Stud Confl Terrorism. https://doi.org/10.1080/1057610X.2020.1862895
13. Liang C, Cross M (2020) White crusade: how to prevent right-wing extremists from exploiting the internet (strategic security analysis: July 2020 issue 11). Geneva Centre for Security Policy. https://dam.gcsp.ch/files/doc/white-crusade-how-to-prevent-right-wing-extremists-from-exploiting-the-internet
14. Trade Union Congress (2020) The rise of the far right: building a trade union response. TUC. https://www.tuc.org.uk/sites/default/files/2020-12/TUC%20Rise%20of%20the%20Far%20Right%20FINAL.pdf
15. Liger Q, Gutheil M (2022) Right-wing extremism in the EU (PE 700.953—May 2022). European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/700953/IPOL_STU(2021)700953_EN.pdf
16. Mudde C (2007) Populist radical right parties in Europe. Cambridge University Press
17. Brown K, Mondon A (2021) Populism, the media, and the mainstreaming of the far right: the guardian's coverage of populism as a case study. Politics 41(3):279–295. https://doi.org/10.1177/0263395720955036
18. Newth G (2021) Rethinking 'nativism': beyond the ideational approach. Glob Stud Cult Power: Identities 30(2):161–180. https://doi.org/10.1080/1070289X.2021.1969161
19. Doerr N (2017) Bridging language barriers, bonding against immigrants: a visual case study of transnational network publics created by far-right activists in Europe. Discourse Soc 28(1):3–23. https://doi.org/10.1177/0957926516676689

20. Wodak R (2015) The politics of fear: what right-wing populist discourses mean. Sage
21. Müller MG (2011) Iconography and iconology as a visual method and approach. In: Margolis E, Pauwels L (eds) The SAGE handbook of visual research methods. Sage Publications, pp 283–297
22. Berg L (2019) Between anti-feminism and ethnicized sexism. Far-right gender politics in Germany. In: Fielitz M, Thurston N (eds) Post-digital cultures of the far right online actions and offline consequences in Europe and the US. Transcript Verlag, pp 79–91
23. Asmolov G (2019) The effects of participatory propaganda: from socialization to internalization of conflicts. J Des Sci (6). https://doi.org/10.21428/7808da6b.833c9940.
24. Marwick A, Lewis R (2017) Media manipulation and disinformation online. Data and Society. https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf
25. Guess A, Lyons B (2020) Misinformation, disinformation, and online propaganda. In: Persily N, Tucker J (eds) Social media and democracy: the state of the field, prospects for reform (SSRC Anxieties of Democracy). Cambridge University Press, pp 10–33
26. Hameleers M, Powell TE, Van Der Meer TGLA, Bos L (2020) A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. Polit Commun 37(2):281–301. https://doi.org/10.1080/10584609.2019.1674979
27. Bennett WL, Livingston S (2018) The disinformation order: disruptive communication and the decline of democratic institutions. Eur J Commun 33(2):122–139. https://doi.org/10.1177/0267323118760317
28. Faris R, Roberts H, Etling B, Bourassa N, Zuckerman E, Benkler Y (2017) Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election (Berkman Klein Center for Internet and Society Research Paper). Cambridge. https://cyber.harvard.edu/publications/2017/08/mediacloud
29. Humprecht E (2018) Where 'fake news' flourishes: a comparison across four western democracies. Inf Commun Soc 1–16. https://doi.org/10.1080/1369118X.2018.1474241
30. Klein O (2020) Misleading memes. The effects of deceptive visuals of the British National Party. Open J Sociopolit Stud 13(1):154–179. https://doi.org/10.1285/i20356609v13i1p154
31. Messaris P, Abraham L (2001) The role of images in framing news stories. In: Reese S, Gandy OH, Grant AE (eds) Framing public life: perspectives on media and our understanding of the social world. Routledge, pp 231–242
32. Meier BP, Robinson MD, Crawford LE, Ahlvers WJ (2007) When "light" and "dark" thoughts become light and dark responses: affect biases brightness judgments. Emotion 7(2):366–376. https://doi.org/10.1037/1528-3542.7.2.366
33. Singh VK, Ghosh I, Sonagara D (2021) Detecting fake news stories via multimodal analysis. J Am Soc Inf Sci 72(1):3–17. https://doi.org/10.1002/asi.24359
34. Hakokõngäs E, Halmesvaara O, Sakki I (2020) Persuasion through bitter humor: multimodal discourse analysis of rhetoric in internet memes of two far-right groups in Finland. Soc Media Soc 6(2). https://doi.org/10.1177/2056305120921575
35. Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J (2012) Misinformation and its correction: continued influence and successful debiasing. Psychol Sci Public Interest 13(3):106–131. https://doi.org/10.1177/1529100612451018
36. San José-Cabezudo R, Gutiérrez-Arranz AM, Gutiérrez-Cillán J (2009) The combined influence of central and peripheral routes in the online persuasion process. Cyber Psychol Behav 299–308. https://doi.org/10.1089/cpb.2008.0188
37. Geddes J (2016) Elaboration likelihood model theory: how to use ELM. Interaction Design Foundation. https://www.interaction-design.org/literature/article/elaboration-likelihood-model-theory-using-elm-to-get-inside-the-user-s-mind
38. Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio L, Brashier N, Kendeou P, Vraga E, Amazeen M (2022) The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psychol 1:13–29. https://doi.org/10.1038/s44159-021-00006-y
39. Martel C, Pennycook G, Rand DG (2020) Reliance on emotion promotes belief in fake news. Cogn Res: Principles Implications 5(47). https://doi.org/10.1186/s41235-020-00252-3

40. Bakir V, McStay A (2018) Fake news and the economy of emotions. Digit Journalism 6(2):154–175. https://doi.org/10.1080/21670811.2017.1345645
41. Lauer D (2021) Facebook's ethical failures are not accidental; they are part of the business model. AI Ethics 1(4):395–403. https://doi.org/10.1007/s43681-021-00068-x
42. Lauer D (2021) You cannot have AI ethics without ethics. AI Ethics 1:21–25. https://doi.org/10.1007/s43681-020-00013-4
43. Krafft PM, Donovan J (2020) Disinformation by design: the use of evidence collages and platform filtering in a media manipulation campaign. Polit Commun 37(2):194–214. https://doi.org/10.1080/10584609.2019.1686094
44. Lavoie D (2021) Woman recalls total 'terror' of Charlottesville car attack. AP News. https://apnews.com/article/sports-violence-lawsuits-race-and-ethnicity-racial-injustice-1e2d3e8ee3662494093ff9ebfc829a50
45. Crawford B, Keen F, Suarez-Tangil G (2020) Memetic irony and the promotion of violence within Chan cultures (ES/N009614/1). Centre for Research and Evidence on Security Threats. https://kclpure.kcl.ac.uk/portal/en/publications/memetic-irony-and-the-promotion-of-violence-within-chan-cultures(51e9a948-2191-4ba9-8657-2c4491b2c0dd).html
46. Crawford B (2020) The influence of memes on far-right radicalisation. Centre for Analysis of the Radical Right. https://www.radicalrightanalysis.com/2020/06/09/the-influence-of-memes-on-far-right-radicalisation/
47. Higgins E (2021) We are Bellingcat. Bloomsbury
48. Ling J (2022) How 4chan's toxic culture helped radicalize buffalo shooting suspect. The Guardian. https://www.theguardian.com/us-news/2022/may/18/4chan-radicalize-buffalo-shooting-white-supremacy
49. Crenshaw M (1987) Theories of terrorism: instrumental and organizational approaches. J Strateg Stud 10(4):13–31. https://doi.org/10.1080/01402398708437313
50. Fielitz M, Ahmed R (2021) It's not funny anymore. Far-right extremists' use of humour (radicalisation awareness network). European Commission. https://utveier.no/wp-content/uploads/sites/6/2021/10/ran_ad-hoc_pap_fre_humor_20210215_en.pdf
51. Ebner J (2019) Counter-Creativity: innovative ways to counter far-right communication tactics. In: Fielitz M, Thurston N (eds) Post-digital cultures of the far right. Transcript Verlag, pp169–181
52. Crawford B, Keen F, Suarez-Tangil G (2021) Memes, radicalisation, and the promotion of violence on Chan sites. In: Proceedings of the international AAAI conference on web and social media, vol 15, no 1, pp 982–991. https://doi.org/10.1609/icwsm.v15i1.18121
53. Askanius T (2021) On frogs, monkeys, and execution memes: exploring the humor-hate nexus at the intersection of neo-Nazi and alt-right movements in Sweden. Telev New Media 22(2):147–165. https://doi.org/10.1177/1527476420982234
54. Gibson C (2019) 'Do you have white teenage sons? Listen up.' How white supremacists are recruiting boys online. The Washington Post. https://www.washingtonpost.com/lifestyle/on-parenting/do-you-have-white-teenage-sons-listen-up-how-white-supremacists-are-recruiting-boys-online/2019/09/17/f081e806-d3d5-11e9-9343-40db57cf6abd_story.html
55. Grierson J (2021) Neo-Nazi groups use Instagram to recruit young people, warns Hope Not Hate. The Guardian. https://www.theguardian.com/world/2021/mar/22/neo-nazi-groups-use-instagram-to-recruit-young-people-warns-hope-not-hate
56. Hayden ME (2018) Neo-Nazi website Daily Stormer is 'designed to target children' as young as 11 for radicalization, editor claims. Newsweek. https://www.newsweek.com/website-daily-stormer-designed-target-children-editor-claims-782401
57. Pauwels A (2021) Contemporary manifestations of violent right-wing extremism in the EU: an overview of P/CVE practices (radicalisation awareness network). European Commission. https://home-affairs.ec.europa.eu/system/files/2021-04/ran_adhoc_cont_manif_vrwe_eu_overv_pcve_pract_2021_en.pdf
58. Somoano IB (2022) The right-leaning be memeing: extremist uses of internet memes and insights for CVE design. First Monday 27(5). https://doi.org/10.5210/fm.v27i5.12601

59. CEE Prevent Net (n.d.) CEE Prevent Net—Central and Eastern European network for the prevention of intolerance and group hatred. CEE Prevent Net. Retrieved from https://ceepreventnet.eu/project-summary.html. Accessed on 7 Jan 2023
60. Weilnböck H, Kossack O (2020) Prevention of group hatred and right-wing extremism in Germany and Central and Eastern European—experiences, lessons learnt and ways forward from the European Fair Skills, Fair*in and CEE Prevent Net projects. In: Heinzelmann C, Marks E (eds) International perspectives of crime prevention, 11th edn. Forum Verlag Godesberg GmbH 2020, pp 159–189
61. Evans R (2019) The El Paso shooting and the gamification of terror. Bellingcat. https://www.bellingcat.com/news/americas/2019/08/04/the-el-paso-shooting-and-the-gamification-of-terror/
62. Wong JC (2019) 8chan: the far-right website linked to the rise in hate crimes. The Guardian. https://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website
63. Zannettou S, Caulfield T, Blackburn J, De Cristofaro E, Sirivianos M, Stringhini G, Suarez-Tangil G (2018) On the origins of memes by means of fringe web communities. In: Proceedings of the internet measurement conference 2018, pp 188–202. https://doi.org/10.1145/3278532.3278550
64. Macklin G (2019) The Christchurch attacks: livestream terror in the viral video age CTC Sentinel 12(6):18–29. https://ctc.westpoint.edu/wp-content/uploads/2019/07/CTC-SENTINEL-062019.pdf

Investigating Online Propaganda Strategies Employed by Extremist Groups Through Visual Propaganda

Georgina Butler

Abstract This chapter aims to assess the key online propaganda strategies employed by extremist groups through visual propaganda. To this end, the chapter discusses visual propaganda as a growing phenomenon within these extremist groups, whilst considering the comparison with textual propaganda and why extremist groups such as ISIS have increased their use of visual content. In terms of the key online strategies, the chapter specifically focuses on investigating the use of video games to radicalize and recruit individuals into extremist groups' organizations. This use of visual propaganda is altering the ways in which these groups approach the radicalization process as well as changing how law enforcement and governments attempt to control it.

Keywords Online propaganda · Extremist groups · Visual propaganda · Textual propaganda · Law enforcement · Internet · Terrorism · Radicalization

1 Introduction

There are multiple tactics and strategies employed by extremist groups in order to disseminate their beliefs through propaganda online. Over the past few decades, propaganda has taken different roles from being physical posters around a century ago to now being in the form of social media posts, videos and video games. Extremist groups have exploited the Internet as a venue for their beliefs and, in particular, have used social media platforms to spread visual propaganda. Extremist groups, or extremism, is a phenomenon that has undergone extensive debate due to the complexity and subjective nature of the topic. However, for the purposes of this chapter, extremism is defined as "activities (beliefs, attitudes, feelings, actions, strategies) of a character far removed from the ordinary" ([1]: 2). Therefore, a violent extremist group can be defined as a collection of individuals that share these

G. Butler (✉)
Department of Criminology, Sociology and Social Policy, School of Social Sciences, Swansea University, Singleton Park, Swansea SA2 8PP, UK
e-mail: georgina.butler500@hotmail.co.uk

R. Montasari (ed.), *Applications for Artificial Intelligence and Digital Forensics in National Security*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-031-40118-3_10

beliefs, attitudes and feelings and act upon them in the form of violence. The term visual propaganda describes the use of contemporary visual content with the goal to influence public opinion [2].

It is contended that the goal of visual propaganda is to increase persuasiveness and to trigger an emotional response in order to fuel the process of radicalization. Visuals are key to establishing these group identities, as images have the ability to shape beliefs and evoke emotions. Each group is distinguished by their unique visual aesthetic, and this refers to their fundamental preference in the content they wish to present and the target audience they wish to reach. This insinuates the diversity among extremist groups in terms of the content they share and where they choose to share it. This chapter will particularly focus on the Islamic State (ISIS).

The ISIS has gained a reputation for using a wide range of social media platforms to spread and propagate violent jihad since its formation in 2013. Around a year later, it then went on to establish the group as a 'caliphate', meaning it is a state which runs under Islamic law. A few examples of platforms which are used as venues for violent Islamic content consist of Twitter, Facebook, YouTube and Telegram [3]. The reason to focus on ISIS as the extremist group is due to the fact that this organization has spent 38 times more funding than other extremist organizations on visual online propaganda [4]. ISIS has also put more emphasis on hard propaganda as their main strategy to recruit and radicalize, whilst also looking to expand their visual propaganda on a variety of platforms. This chapter will discuss visual propaganda as a growing phenomenon within this extremist generation, whilst considering the comparison with textual propaganda and why ISIS have increased their use of visual content. Moreover, in terms of the key online strategies, this chapter will specifically focus on the use of video games to radicalize and recruit individuals into their organization. This use of visual propaganda is altering the ways in which extremist groups approach the radicalization process as well as changing how law enforcement and governments attempt to control it.

2 Visual Propaganda

In the past, propaganda came in the form of posters that aimed to recruit individuals for political purpose and to incite them into joining a particular movement, or in some cases, war. More recently, digital visuals have taken the place of these old-fashioned forms of propaganda, with Hasic stating that recent and future propaganda has the ability to influence and persuade far beyond propaganda of the past [5]. Additionally, Hasic also argues that the modern advertising world owes the majority of its success to visual propaganda, and this is as a result of the advancement in technology contributing to the influence of visual content. Communication via visual content has been a crucial part of history in terms of the social and political environment. Not only has social media aided this movement to extensive visual content, but many companies, news agencies and governmental organizations have taken advantage of the growth of digital communication and used it to spread their agenda [6].

Unfortunately, social media platforms have also been venues for extremist groups to post content and share their extreme beliefs and attitudes to their audiences. The reason as to why these groups are spreading more visual content is due to the fact that when individuals consume information visually, it produces a response in the subconscious in ways that textual information simply cannot do [5]. It simplifies the message avoiding the complexity that comes with analysing textual data. Therefore, more credible and rapid conclusions can be drawn from visual content compared to the inaccuracy of text. Visual imagery produces a particular emotion which is needed for the purposes of persuasion [6], which is key in terms of recruitment and the radicalization process. It could be argued that words are not persuasive enough in order to convince an individual to change their beliefs. This is what renders visuals so effective at spreading a particular message. Visuals are universal and have the ability not only to produce an emotional response but also to be interpreted differently [7]. In particular, there are some viewers that will see a gruesome image and feel empathy or some kind of negative emotion, whereas there will be other people that could see this violence and feel positive emotions towards it or feel a need to justify their negative emotions. It is these types of individuals that these organizations are targeting with their strategic propaganda.

Images cannot convey a message by themselves because they depend on the image triggering emotion within the individual that already exists in their view of the world. In this case of extremist groups, and ISIS, the propaganda would need to target a specific audience to ensure the response they desire. Propaganda has this ability to reinforce and transform an individual's ideology, and literature suggests that certain emotional responses are needed to result in strengthening the existing political narrative of the individual. These involve resentment, anger, shame and fear [8]. These types of feelings are essential in also producing a violent and aggressive response which is the aim of these organizations. ISIS tend to target individuals based on whether they are experiencing identity issues. When individuals are suffering with personal identity, advocators of the group will interfere and attempt to convince them to align their beliefs with the ingroup and to join a group of like-minded individuals.

No great skills are needed to look at or watch a piece of visual propaganda, which implies that intelligence is not a necessity when targeting, as they can then mould and control the learning of the individual. Additionally, coming back to this idea of persuasiveness, researchers have introduced the idea of a framing device when it comes to visual content, which essentially ensures the message is persuasive enough by specifying the content to a certain audience. Framing is the ability to influence and manipulate how people perceive images. In other words, making the audience see what they want them to see [6]. When this cycle is repeated, framing becomes more effective and can create an 'echo chamber', which is a group of people with the same beliefs communing in one space on the Internet [9]. Using technology and framing, these organizations can create an automatic algorithm which selectively pushes content onto the platform of choice so that the individual continuously receives the same visual content and ideology [10]. They are then directed to specific webpages, group forums or social media platforms where like-minded people are posting similar content. This is where these echo chambers are formed. These are

essential within the radicalization process as it is where these individuals will be subject to more visual content and will become aware of more people who share the same beliefs and attitudes, making visual propaganda a key strategy for extremist groups online.

3 ISIS and Social Media

In 2010, multiple Middle Eastern and North African nations (MENA), such as Syria and Egypt, underwent the so-called Arab Spring. It was known to have been largely facilitated by the use of social media and online propaganda, and as a result of this, online activity increased. It seemed as if those in political power, alongside the activists, were taking advantage of the freedom of the Internet. Different social media platforms were used to raise awareness and educate users about the political situation whilst also networking to organise protests and recruit help from surrounding countries [6]. Learning from this, when ISIS was formed in 2013, they quickly gained a reputation for utilizing social media platforms to spread violent jihad. This meant that ISIS instantaneously had networks and links to the majority of well-known platforms with the most popular and favoured one being Twitter [3]. Other platforms used included YouTube, WhatsApp, Facebook and Telegram. ISIS exploited Twitter as a space to spread their content, to target specific audiences by tailoring this content, and to subsequently enhance their recruitment abilities. This particularly involves visual strategies and content through the use of videos, photographs, live streams and, in some cases, Internet magazines.

Over the years, ISIS's use of visual content has generated significant interest which has arguably overturned the earlier existing opinion that text is the favoured tool for propaganda [8]. Different social media platforms are used by organizations such as ISIS for various objectives. Services such as YouTube and Twitch put a lot of emphasis on the sharing of content through live streams or videos, whereas social media platforms such as Twitter and Facebook are more focused on sustaining and fostering social connections [6]. Additionally, there are different ways they use these platforms to spread visual propaganda, ranging from training instructions to manuals on the creations of weapons in support of lone wolf individuals. Those include individuals who act in retaliation against their state or government due to disagreeing with the policies or measures they are taking to counter terrorism [11]. They do not act for a certain extremist organization but rather on behalf of their own beliefs and attitudes. Therefore, there are also instructions for keeping secure connections and communications, as well as how to access certain VPNs and proxies to ensure security of the organization [4]. It can be deduced that the rise of visual propaganda used by ISIS has breathed new life into mainstream propaganda and radicalization all over the internet [12].

4 Narratives and Propaganda Within ISIS

More specifically, a number of researchers and studies go deeper into how propaganda leads to radicalization, and how crisis diagnosis and solution prognosis come to highlight the importance of visuals [8, 13]. It insinuates that extremist groups can promote push or pull factors to create the narrative that fits the target. In a process named as 'cycle cognitive reinforcement', extremist groups have a narrative which states crisis links to the outgroup and solutions link with the ingroup. The frequent association of these links strengthens the identity framework along with the push and pull factors that promote radical mobilisation [8]. Crisis diagnosis is this push factor, it focuses on this influence of the other. It elicits unpleasant feelings, such as rage or resentment in order to persuade people to take action, to take those negative emotions and put it to use. On the other hand, solution prognosis creates a narrative which highlights commitment and how the radicalisation process could be approached from a different perspective. A pull factor would insinuate evoking positive, prideful emotions with the intent to draw potential sympathizers, and those sympathizers will be key players in sharing visual propaganda to a wider audience. To quote Baele, Boyd and Coan, visual propaganda and images "are key tools to reinforce exclusive group identities and cement the crisis or solution attributions that fuel radicalization" ([8]: 6). The aim of this visual content is to incite fear and evoke emotions that result in large-scale influence over a specific audience. However, in order to understand the different visual strategies of these groups, it is essential to recognise how these groups are digitally communicating [7].

In terms of the two general types of propaganda, soft and hard propaganda are used within multiple instances of propaganda. In this case, they are targeted more to how extremist groups and ISIS utilize visuals in their strategies. Soft propaganda aims to show how life within ISIS or any extremist group is equivalent to the life of normal citizens anywhere in the world. An example would include photos of people, especially children, appearing to be laughing or smiling. Images of people praying could also be included, along with the posing of families cooking, medical care being given and photos of farming taking place, could convince the audience that this is a community of people that exist just like them and participate in the same day-to-day activities that they do. In addition, the use of natural or urban-looking landscapes are designed to evoke positive emotions regarding the standard of life there. An example of soft propaganda comes from the al-Hayat Media Centre, which published multiple videos showing how life under the Islamic State was normal, consisting of some of the examples stated [14].

Hard propaganda, on the other hand, involves imagery with weapons or violence. This could be the aftermath of a shooting, with blood and bodies included in the video or photo, evidence of military training, footage of conflict or burned vehicles. Some of the most well-known instances would be videos or images of masked men in black clothing standing over victims, or performing acts of punishment [4]. There is also recurring symbols which are present among this visual propaganda, including the image of people praying, holding religious documents or mosques themselves.

The One Finger is also a common symbol within ISIS propaganda in particular, and the use of the ISIS flag individually is considered soft. However, when it is seen in a video where violence or torture is occurring, then it is considered hard propaganda. An example of hard propaganda could be the images taken in some ISIS Internet magazines showing weapons or military training, or it could be in the live stream videos of beheadings [14].

5 Video Games

Extremist groups use visual propaganda in a number of different instances, not only by disseminating pictures and videos on multiple different platforms, but also through video games. This section will focus on how video games as a new phenomenon are facilitating the online radicalization process, and why it is fast becoming one of ISIS's key strategies for recruitment of youths. Online radicalization in relation to extremist and terrorist group, such as ISIS, is referred to as a process where individuals become prone to sensitive content online and, therefore, will eventually internalise these beliefs and extreme behaviours [10]. Playing video games is as old as the introduction of computers, with some of the first games having been developed in the 1960s [15]. Video games previously were a place where individuals could use their imagination and ingenuity to serve as a getaway from the outside world. Nowadays, video games are a venue for visual propaganda and online radicalization used by extremist groups such as ISIS to recruit, radicalize and train. According to Lakomy, the most attractive feature of the video game market is the pure size and global reach, as well as its ability to be interactive and creative [15]. One of the main features of gaming and the gaming culture is the 'good' versus 'bad' guys element to it. Without even playing the game, there is already an understanding of who the enemy is. The majority of players will either join and meet their friends on a game, or they will make friends on a particular game, bringing this sense of community and brotherhood. A well-designed game ensures a large personal involvement as well as immersion of themselves within the game alongside others. Gaming tends to alter the players' sense of reality, allowing them to be in a world of limited consequences and the ability to construct their own ideology. In essence, even though ISIS are manipulating and creating these games for the full purpose of recruiting people, it is designed so that the players feel as though they are crafting their own story. Combining this autonomy and the sense of belonging, playing games which are constructed by extremist groups increases the players' chances of identifying with the organization [16].

Before the increasing participation of extremist groups in the gaming industry, video games had consistently been related to communicating a political message [8]. Ottosen stresses the importance of the gamer culture, and how the growth in gaming and the computer industry has influence on the overall youth of this generation [17]. Many studies comment on the stigma within the group of people that play video games, implying that it is geared towards anti-social, isolated individuals who join

these online misfit communities [18]. Furthermore, it could be debated that whilst there is a certain stereotype of individuals that play video games, referring to young white males, this creates an overwhelming sense of brotherhood and belonging to an ingroup, which is what ISIS is attempting to achieve by mirroring the events and life that occur in the Islamic State (ibid 2021). It could also be argued that due to certain beliefs of the Islamic State and the culture and attitudes, this stereotype is the perfect candidate for the type of people they are recruiting and wish to radicalize. Both ISIS and video games emulate this sense of brotherhood, fighting for something greater than themselves, and belonging to a community. It becomes clear why this strategy of propaganda is very successful. Video games are also used to train individuals to be ISIS fighters (ibid 2021). The interactivity of these games is an essential tool for training, as the players are learning certain skills such as using firearms and aircraft training. Using this method of training is beneficial because it is more creative and enjoyable for the players whilst still containing that addictive and competitive element [15]. This method is otherwise known as gamification, and it has received attention from researchers as having a key role in radicalization [18]. Gamification, simply, is the use of elements that make the game rewarding or competitive, such as the rules of the game. Literature has also highlighted the use of a reward system through points and a leader board, making the players compete with other players in the game for the reward. It makes the game more addictive and competitive as well as enforcing the type of behaviour that ISIS are trying to produce [16, 18]. This is one of the reasons why video games are growing in popularity for online radicalization. Critics have stated, however, that gamification is overestimated in terms of causing a positive response, and instead of gamification itself, these benefits may be due to the excitement of the video games themselves [16].

One of the first games created by an extremist organization was that by Al Qaeda in 2006. This game was released for free by the Global Islamic Media Front and was called ‘Quest for Bush’ [15]. Although it does not focus on ISIS as the extremist group, it is beneficial to understand how these organizations operate the games and how they channel their ideology through the game. The start screen contained images of two of Al Qaeda’s most influential members, Abu Musab al-Zarkawi and Osama bin Laden, placed above both George Bush and Tony Blair. One of the key features of the game was that it was a first-person shooter (FPS) game and led players through six levels that were based around the US invasion of Iraq. FPS refers to the player being the killer in the game, almost as if they are the main character and so they witness everything one would if it was real life [16]. This adds to the autonomous aspect of video games, as it gives the player the option to create whatever storyline they wish to create.

Although there is no causal connection between violence within video games and violent behaviour, researchers are still debating whether FPS has seen a rise in violent, aggressive behaviour. Andrews and Skoczylis [18] note the ongoing debate surrounding school shootings and an increase in FPS video games, whereas Schlegel has highlighted how the Christchurch attacker live-streamed the event, using visuals to create this FPS aspect [16]. This element of FPS, which is used interchangeably

both within online and offline environments, is normalising certain behaviours and equally is contributing to the training of these individuals.

The 'Quest for Bush' video game had consistent jihad ideology throughout the game, and because they are so immersed in that alternate reality, these messages are only reaching their subconscious mind. Eventually, this ideology will become a part of the individuals, and the visuals they see will almost become normal. Thus, they are gradually becoming another lifeless tool of visual propaganda that is being controlled by extremist ideology. Due to this distorted sense of reality, players are unable to distance themselves between what is the game and what is real life. Although it is not a video game, Huroof was released by the Islamic State's official centre for propaganda as an educational app. It can be downloaded on both desktop computers as well as any other mobile operating systems, rendering it the only multi-platform app made by a terrorist organization. Some effective features are that the overall aesthetics are cartoonish, and it has a very childish style to it, ranging from bright colours to pictures taken right out of a children's book [15]. It is attempting to emulate some form of soft propaganda by coming with a softer approach and making it accessible and understandable to everyone. There are constant ISIS symbols throughout the app, ranging from the flag to cartoon drawings of guns and tanks. The app also aims to teach the Arabic alphabet using Nasheed music and prompting the user to match the letters with military equipment [15, 16].

These types of apps and games were becoming accessible to players of any age, with advanced or basic software, and to those who were both amateurs and professionals. Lakomy uses the term 'homegrown radical' to highlight the ease at which these individuals have access to all the instruments necessary to the radicalization process [15]. Not only does this confirm the success of video game as an online strategy, but it is also essentially proving that individuals can achieve radicalization without the extensive intervention from members of an organization. These organizations are continuing to learn and become more innovative with the way they use visual propaganda. It is apparent how effective this type of propaganda is in constructing an ideal reality, and how successful video games are as a strategy to recruit through online radicalization.

6 Conclusion

To conclude, there is limited research into the direct impact as to whether video games result in extremist ideologies. This is due to the multifaceted and complex nature of the issue. This is also a result of it being an under-researched and underestimated tool for extremist groups. It is true that video games have facilitated the recruitment of individuals into ISIS and other extremist groups. However, if video games did not exist, it would only cause a migration of people to a different platform in order to achieve their goals. As long as both the Internet and extremist groups have been alive, these groups have managed to conduct their business. If not through gaming, they will find another creative way to spread their beliefs and to share their visual

propaganda as a way to recruit and radicalize their target audience. The problem with extremist groups is that once they find a strategy that produces a successful response, they will continue to utilize and protect it, finding new ways to keep the ball rolling and to remain invisible. Notwithstanding these, video games will continue to be an effective strategy for ISIS to disseminate their visual propaganda. It plays a key and underestimated role in the recruitment and training aspect of radicalization, arguably the most important stages of the process. This is due to not only the interactivity of the games but also the gamified element, causing it to be an enjoyable and addictive way to learn. Video games allow extremist organizations to channel their ideology through the game, and as the players become immersed in the virtual world, they become comfortable and acquainted with their beliefs. This leads to online radicalization.

References

1. Coleman PT, Bartoli A (2003) Addressing extremism. White Paper. The International Center for Cooperation and Conflict Resolution, Colombia University, New York. http://www.tc.columbia.edu/i/a/document/9386_WhitePaper_2_Extremism_030809.pdf
2. Holzer A (2015) Visual Propaganda. In: Brill's digital library of World War I. Consulted online on 19 Dec 2022. https://doi.org/10.1163/2352-3786_dlws1_beww1_en_0612
3. Pearson E (2018) Online as the new frontline: affect, gender, and ISIS-take-down on social media. Stud Confl Terrorism 41(11):850–874
4. Hashemi M, Hall M (2019) Detecting and classifying online dark visual propaganda. Image Vis Comput 89:95–105. https://doi.org/10.1016/j.imavis.2019.06.001
5. Hasic A (2019) Why propaganda is more dangerous in the digital age? The Washington Post. https://www.washingtonpost.com/outlook/2019/03/12/why-propaganda-is-more-dangerous-digital-age/
6. Seo H, Ebrahim H (2016) Visual propaganda on Facebook: a comparative analysis of Syrian conflicts. Media War Confl 9(3):227–251. https://doi.org/10.1177/1750635216661648
7. Dauber CE, Winkler CK (2014) Visual propaganda and extremism in the online environment. US Army War College Press, Carlisle Barracks
8. Baele SJ, Boyd KA, Coan TG (2019) Lethal images: analysing extremist visual propaganda from ISIS and beyond. J Glob Secur Stud 1–24. https://doi.org/10.1093/jogss/ogz058
9. O'Hara K, Stevens D (2015) Echo chambers and online radicalism: assessing the internet's complicity in violent extremism. Policy Internet 7(4):401–422
10. Binder JF, Kenyon J (2022) Terrorism and the internet: how dangerous is online radicalization? Front Psychol 13:997390. https://doi.org/10.3389/fpsyg.2022.997390
11. Houssem BL (2017) How terrorists use propaganda to recruit lone wolves. The Canadian Press. https://www.proquest.com/wire-feeds/how-terrorists-use-propaganda-recruit-lone-wolves/docview/1953407624/se-2
12. Winter C (2015) The virtual 'caliphate': understanding Islamic state's propaganda strategy, vol 25. Quilliam, London
13. Ingram H (2017) An analysis of inspire and Dabiq: lessons from AQAP and Islamic state's propaganda war. Stud Confl Terrorism 40(5):357–375
14. Siboni G, Cohen D, Koren T (2015) The Islamic state's strategy in cyberspace. Mil Strat Aff 7(1):127–144
15. Lakomy M (2019) Let's play a video game: Jihadi propaganda in the world of electronic entertainment. Stud Confl Terrorism 42(4):383–406. https://doi.org/10.1080/1057610X.2017.1385903

16. Schlegel L (2020) Jumanji extremism? How games and gamification could facilitate radicalization processes. J Deradicalization (23). ISSN: 2362-9849. https://journals.sfu.ca/jd/index.php/jd/article/view/359/223
17. Ottosen R (2017) Video games as war propaganda: can peace journalism offer an alternative approach? In: Ross SD, Tehranian M (eds) Peace journalism in times of war, peace and policy, vol 13. Routledge, Abingdon, pp 73–85
18. Andrews S, Skoczylis J (2021) Video games, extremism and terrorism: a literature survey. Global network on extremism and technology (GNET). https://gnet-research.org/2021/11/16/video-games-extremism-and-terrorism-a-literature-survey/