Volume 2, Issue 1 (2024)

ISSN: 2959-3018 (print) / 2959-3026 (online)
thecommonwealth.org/cyber-journal

The Commonwealth Cyber Journal

The Commonwealth

Foreign, Commonwealth & Development Office

The Commonwealth Secretariat
Marlborough House
Pall Mall
St. James's
London SW1Y 5HX
United Kingdom

n.amobi@commonwealth.int
cyberjournal@commonwealth.int

About the journal

The Commonwealth Cyber Journal (CCJ) is an annual journal published by the Commonwealth Secretariat that features peer-reviewed, policy-influencing articles and commentary by academics, policymakers, practitioners and experts on the benefits, challenges and risks of digital technologies. It seeks to analyse challenges and opportunities arising from different aspects of cybercrime, cyberlaw and cybersecurity, and to serve both as a toolkit and resource for practitioners, legislators, and academics cybercrime and as a decision support instrument for stakeholders (state/non-state actors) as they seek to strengthen their countries' cyber legislation.

The journal's areas of focus include but are not limited to: state actors and cyber warfare; ransomware and phishing; proceeds of crime; terrorism, privacy and security of data; intellectual property; infringement and counterfeit; online harassment and cyberstalking; election cybersecurity; virtual courts and electronic evidence; cybersecurity and the economy; digital currencies; and child online safety. Articles published in the journal specifically focus on the Commonwealth region, and/or include case studies concerning one or more Commonwealth countries; similarly, article authors are typically drawn from Commonwealth countries.

For full details, including aims and scope, and guidelines for submission, see thecommonwealth.org/cyber-journal

The Commonwealth

Contents

The Commonwealth

Commonwealth Cyber Journal \ 1
ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

Editorial

Nkechi Amobi and Aidan Ferguson

The security and trustworthiness of internet infrastructure is essential in today's society. Cybersecurity research has, therefore, taken inspiration from a wide range of fields to accomplish this goal. This research has involved the implementation of various measures to ensure the confidentiality, integrity and availability of digital assets. Yet, these cybersecurity initiatives must also now navigate the challenges posed by artificial intelligence (AI).

AI is the fastest growing deep technology[1] in the world, and its use has the potential to rewrite the operational and policy rules of governments and industries. AI is becoming ubiquitous and has the potential to drive substantial global economic growth by replacing or becoming a viable alternative to human endeavours. However, as AI's benefits rapidly grow, so do its risks.

AI and cybersecurity have a complex relationship. On one hand, AI techniques can be leveraged to enhance cybersecurity measures by detecting and responding to threats more efficiently than traditional methods. Yet AI also introduces new cybersecurity risks. As AI systems become more advanced and integrated into critical infrastructures, they become potential targets for adversaries seeking to exploit vulnerabilities or manipulate their decision-making processes. While it is essential to recognise the risks posed by AI, nations must seize the substantial opportunities that it presents to build various aspects of their economy, including cybersecurity and resilience – ideals recognised in the 2018 Commonwealth Cybercrime Declaration.[2]

In 2023 the Commonwealth Secretariat established the Commonwealth Artificial Intelligence Consortium (CAIC) with the aim of leveraging AI's potential to empower citizens – especially women and girls, youth and other vulnerable groups – with the necessary skills to benefit from the opportunities to be found in cyberspace. Through the creation of AI education and skills courses, the development of AI and related digital infrastructure, responsible AI policies and regulations, and capacity-building initiatives on AI safety, the CAIC has promoted capacity-building throughout the Commonwealth through its working group. Furthermore, at the 2024 Commonwealth Law Ministers Meeting, held in Zanzibar, the development of a comprehensive approach that integrates AI and virtual technologies to assist citizens in accessing justice across the Commonwealth was endorsed by Law Ministers in recognition of the interplay between AI and the justice system.

1 Office for Artificial Intelligence (UK) (2021, September) National AI Strategy. Command Paper 525. London. Available at: https://www.gov.uk/government/publications/national-ai-strategy

2 See https://thecommonwealth.org/commonwealth-cyber-declaration-2018

The *Commonwealth Cyber Journal* (*CCJ*), published by the Commonwealth Secretariat, serves as a platform for disseminating cutting-edge research, policy influencing articles, case studies and commentary from practitioners, policy-makers and academics in the field of cybersecurity and cybercrime. The objective of the *CCJ* is to assist Commonwealth countries to strengthen their anti-cybercrime legislative, policy, institutional and multilateral frameworks to uphold the rule of law in both virtual and physical spaces.

In the special section on artificial intelligence

This second edition of the *CCJ* primarily focuses on AI: its first five articles, collected together in the special section on AI, address emerging threats and employ AI approaches to improve cybersecurity safeguards. The contributors to this issue cover topics including AI in the justice system; generative artificial intelligence-led crime as a service (GAI-led CaaS); violent extremists and AI; AI and the future of intellectual property rights; analysis of the Budapest Convention and draft UN anti-cybercrime framework; and the future of cyber insurance and cybercrime in the Asia-Pacific region.

Olalekan Bello and Cecile Ogufere's article on 'The Emerging Artificial Intelligence Legal-Judicial Systems' Interface: Assessing the State of Nigeria's Judicial System's Readiness for a Revolution' analyses how AI is revolutionising various sectors, including the legal-judicial system. Their article highlights how AI can be used to predict case outcomes, streamline contract review processes, save time and resources for legal professionals and assist judges in making more informed decisions. Despite its potential benefits, the article notes that Nigeria's judicial system faces numerous challenges that may hinder its readiness for an AI revolution. Their paper focuses on how that system can draw insights from the emerging global frameworks to establish its own regulations to implement AI technologies, safeguard the rights of citizens and ensure fair and unbiased decision-making processes. The article discusses the opportunities and challenges of integrating AI into Nigeria's judicial system, and concludes that by addressing existing barriers and establishing robust ethical and legal frameworks, the legal-judicial system can harness the potential of AI to enhance efficiency and decision-making processes.

Nicole Matejic and Chris Wilson, in their article 'Crimes of Influence: Generative Artificial Intelligence-led Crime as a Service', advance the idea that crimes of influence are crimes that seek to influence people towards harmful outcomes, and will be a defining feature of generative AI-led cybercrime. While the technology itself is a regular feature of contemporary discussion and research, less thought has been given to the ways in which generative AI (GAI) impacts human cognition to create increasingly permissive environments in which cybercriminals and terrorists can operate. Their paper explores how GAI will likely evolve to deliver persuasive influence at potentially unavoidable economies of scale, while also considering current Commonwealth and global governmental and multistakeholder responses to these challenges.

Gilberto Martins de Almeida, Fernando Bourguy, João Farrel and Diego Semeraro, in their article 'Legal application of technical and procedural standards and frameworks in the combat against GAI-powered cybercrime', acknowledge that GAI has deepened the gap between fast-changing innovative cyberattacks and the slow pace of legislative processes. In the article, the authors argue that to mitigate the resultant exposure, states should look for ways to address this gap. One option to be considered is resorting to standards, which may provide faster adoption, more specific focuses and international recognition. They note that this is particularly valid for the Commonwealth's small states, whose structures may not be as resourceful as those of larger states. In this sense, they posit that standards could be used to fill in the blanks of cybercrime laws (such as indicating that GAI could fall within the concept of 'computer system', which already exists in the cyberlaws of many Commonwealth small states). In summary, the article analyses the convenience of building effective supplementation and constant updates by and between standards and legal rules, referring to several published standards which could be helpful for the prevention of GAI-powered cybercrime.

Wan Rosalili Wan Rosli's article, 'Violent Extremism and Artificial Intelligence: A Double-Edged Sword in the Context of ASEAN', advances the argument that cyberspace has created a new haven from which terrorist organisations can carry out terrorist activities, which has resulted in unprecedented transnational extremism and extremist networking. The emergence of new technologies such as AI has also provided a new sandbox in which insurgents can spread their propaganda. The article provides a discussion on the duality effect of AI in countering violent extremism within ASEAN by highlighting the risks and vulnerabilities attached to the deployment of such technologies, and sheds light on both how such technologies can be misused and how ASEAN states can respond to the risks associated with AI.

Teresia Munywoki's article 'AI Systems and the Future of Intellectual Property Regimes' explores the relationship between AI and intellectual property rights (IPR), highlighting the challenges and opportunities that arise from their intersection. She posits that, given AI's pervasive influence across various sectors, questions surrounding authorship, ownership and protection of AI-generated innovations have emerged. She argues that traditional IPR frameworks, designed with human creators in mind, now face the task of adapting to accommodate the unique characteristics of AI-generated content and inventions. The discussion spans patent law, copyright issues and ethical issues, highlighting the need for a balanced approach that fosters innovation while safeguarding the rights of all stakeholders. The legal dilemmas that Munywoki highlights, such as determining inventorship for AI-generated inventions and attributing copyright for AI-created content, underscore the complexity of this evolving landscape. Her article also emphasises the importance of collaborative efforts to shape a future-proof IPR framework that balances innovation, accessibility and ethical considerations in the AI era.

Also in this issue

Eric Cho and Serene Chan, in the article 'Ensuring a Secure Future by Insuring Against Cybercrime', argue that with rapid advancements in technology and digitalisation, cybercrime has emerged as a formidable threat capable of disrupting business operations and causing financial impacts across society. They note that for years, insurance has become a method of transferring such risks from businesses and individuals to third-party entities, offering a sense of reassurance and protection. With this emerging risk, there is a compelling case for the utility of insurance in addressing this need for risk transfer in the market. Beyond this mitigating solution, the authors also underscore the importance of government intervention, emphasising cybersecurity as a matter of national security. Acknowledging the limitations of the private insurance sector, the authors advocate for collaborative efforts between public and private entities to address the multifaceted challenges posed by cyber risks in effective ways.

In his paper, 'The National Security Exception in International Trade and Cybersecurity', **Kartikeya Garg** theorises that international trade and cybersecurity are becoming increasingly interconnected, with countries implementing various digital technologies to facilitate cross-border trade in goods, services and information. To regulate this, countries could adopt various cybersecurity policies such as data localisation, export controls and import restrictions; however, these may be trade-restrictive and violate World Trade Organisation (WTO) law. Garg argues that the national security exception contained in WTO agreements, and most free trade agreements (FTAs), provide countries with an avenue to escape certain multilateral trade obligations. Historically used for protection against traditional security attacks, this article discusses the evolution of the national security exception within the WTO and other FTAs to determine whether existing treaty formulations can extend to cybersecurity measures. The paper recommends formulations and considerations that countries could implement to create a more holistic security exception in the international trade regime.

Olajide O. Oyadeyi, Oluwadamilola Adeola Oyadeyi and Rofiat Omolola Bello, in 'Cybercrime in the Asia-Pacific Region: A Case Study of Commonwealth APAC Countries', advance the theory that the digital transformation in the Asia-Pacific (APAC) region, coupled with its expanding economic activities and online presence, has made it a prime target for cybercrime. They note that according to cybercrime projections, the region faces a potential cost of roughly US$3.3 trillion from cybercrime by 2025. The authors discuss the context behind cybercrime in the APAC region, particularly its post-COVID proliferation; the pros and cons of the potential use of AI in cybersecurity; cybersecurity initiatives and strategies in the Commonwealth APAC region; and options for policy consideration.

Finally, **Brenda Mwale**'s article 'Towards a Victim-Centred Approach? Reflections on Existing Cybercrime Instruments and the Draft United Nations Convention on Cybercrime' advances the argument that punishing offenders through a retributive

approach is often not sufficient to address the plight of victims of cybercrime, given the unique impact that such crimes can have. Having examined the extent to which existing cybercrime instruments, and the draft UN Convention on Cybercrime, address victims' needs, Mwale argues that legal approaches to addressing cybercrime should adopt a victim-centred approach that offers them adequate safeguards to protect them, and human rights guarantees.

6 \ Commonwealth Cyber Journal
ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

The Commonwealth

Special Section on Artificial Intelligence

The Emerging Artificial Intelligence Legal-Judicial System's Interface: Assessing the State of Nigeria's Judicial System's Readiness for a Revolution

Olalekan A. Bello[1] and Cecile Ogufere[2]

Abstract

The revolutionary trend of artificial intelligence (AI) and its potential to become pre-eminent in all facets of life is no longer a subject of debate. Rather, the question is, to what extent will AI become the determinant of the global juridical-legal systems? With a specific focus on Nigeria's judicial-legal system, this article explores the intersection and potential impact of AI as a paradigmatic shift in technology. Starting with the evaluation of the current state of Nigeria's judicial system's existing infrastructure and technological capabilities, we attempt to unpack the level of awareness, understanding, engagement with, and acceptance of AI among judges, legal professionals and general public. Highlighting the increasing importance of AI in various sectors, we examine the potential benefits of integrating AI into the Nigerian judicial system, including, among others, increased efficiency, improved access to justice and enhanced decision-making processes. We also explore the challenges associated with the adoption of AI in the legal sector, including ethical considerations, data privacy and job displacement. This is in cognisance of the result from a polling at the 2019 Doha Debates on AI showing a majority of online voters (53 per cent) forecasting exponential intensification in global inequalities and apocalypse for humanity in the era of AI's pre-eminence. However, drawing on Muthoni Wanyoike's 'Mindful

1 School of Law, University of Leicester (Leicester, UK). Email: olalekan.bello@leicester.ac.uk; Orcid number: https://orcid.org/0000-0002-0876-3673.
2 Regent's University (London, UK). Email: oguferec@regents.ac.uk.

Optimism'[3] thesis on AI, we advocate for legal and judicial systems' embrace of, and adaptation to, AI's technological advancements. Although AI stands to be disruptive in the human sphere, as a human construct it must be taken as a quantum leap in human thinking and a reflection of the progress of human intelligence. Therefore, for lawyers and the judicial system, the case should not be antagonism towards or embrace of AI's presumed destructive force, but rather finding new ways of engineering the regulation of society through AI machines with negligible social disruption. Against this background, we suggest recommendations for the Nigerian judicial system to effectively navigate the integration of AI into its operations. These should see enhanced digital infrastructure, development of AI-specific policies and regulations, investment in education and training for judges and legal professionals and the fostering of public trust and acceptance of AI technologies. We conclude by advocating that the Nigeria's judicial system should be reviewed with a view to aligning with the technological revolution driven by AI. By addressing the challenges and implementing the recommended strategies, Nigeria can harness the potential of AI to transform its legal and judicial systems and improve access to justice for its citizens.

Keywords: artificial intelligence (AI), legal-judicial interface, legal/judicial system, Nigeria, access to justice.

Introduction

The revolutionary trend of artificial intelligence (AI) and its potential to become pre-eminent in all facets of life is no longer a subject of debate. Rather, the question is, to what extent will AI become the determinant of the global legal and judicial systems? As Williams views the technology's sweeping revolution, humans have now ceded vast amounts of our existence to competing technologies that seek to dominate our time and increase the amount of life that is available for them to capture.[4] This is echoed by Zarkadakis, who opines that what AI envisions is a world in which we could recreate and decipher nature by building a bold new civilisation. This civilisation would be complete with self-regulating factories, cures for all ailments, strong economies, just communities and thinking machines.[5] Yet as he argued, due to its seemingly unreasonable goal of duplicating human nature with all its flaws, AI is perhaps the most perplexing

3 Wanyoike, M. (2019) 'AI Promises Equality Among Nations', presented at 'Artificial intelligence: Is it worth the risk?', Doha Debates, Qatar Foundation, Northwest University, Doha, Qatar, 3 April 2019. https://dohadebates.com/course/artificial-intelligence/#lession-4a-speaker-muthoni-wanyoike

4 Williams, J. (2018) *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*, Cambridge University Press, 93–94

5 Zarkadakis, G. (2017) *In Our Own Image: Savior or Destroyer? The History and Future of Artificial Intelligence*, Pegasus Books, 2

technological advancement ever attempted by humankind.[6] On the positive side, Zarkadakis prognosticates for AI the powering of a new machine age that could propel humanity to unprecedented levels of social, technological and economic advancement.[7] However, he sees in AI something more sinister at work, because many in the sector firmly believe that when more powerful computers develop sentience, they will conquer the entire planet and wipe out humanity.[8]

In respect of the legal profession's association with AI there is no doubt that it is walking unchartered territory, with new technologies' disruptive potential being greater in the judicial and legal systems than other sectors.[9] However, PwC's projection of overall societal configuration by 2030 estimates that AI portends to contribute $15.7 trillion to the global economy, with $3 trillion from increased productivity and $9.1 trillion from new products and services'.[10] In 2017 Canada became the first country to establish a national AI strategy – the $25 million Pan-Canadian AI Strategy.[11] The USA followed suit with a presidential executive order in 2019 to make AI to boost national prosperity, economic security and the standard of living for the American people.[12] China has also aggressively pursued AI growth with centres in Beijing and Tianjin attracting $2.1 and $16 billion AI funding respectively.[13] It goes without saying that other countries – Singapore, France, UK, Germany, UAE, India and Japan – have also devised their national AI strategies.[14]

So, how can AI be defined? While no universally agreed definition has been made, this paper highlights two: first, a machine's 'ability to perform the cognitive functions we usually associate with human minds';[15] second, AI is a computer system which has the ability to perform tasks requiring ordinary human intelligence, many 'powered by machine learning, some [...] powered by deep learning and some... powered by very boring things

6 Ibid, p. 10
7 Id, p. 12
8 Ibid
9 Brooks, C., Gherhes C. and Vorley, T. (2020) 'Artificial Intelligence in the Legal Sector: Pressures and Challenges of Transformation', *Cambridge Journal of Regions, Economy, and Society*, 13, 135–152; see also LexisNexis (2014) 'Workflow and Productivity in the Legal Industry: How Today's Legal Professionals Are Responding to the Changing Landscape'. https://www.lexisnexis.co.nz/en/insights-and-analysis/research-and-whitepapers/2014/2014-workflow-and-productivity-in-the-legal-industry (accessed 24/09/2023)
10 Michael C. (2023) 'AI Strategy: Nigeria in Global Hunt for its Best Minds', *Business Day, Nigeria*, 29 August. https://businessday.ng/technology/article/ai-strategy-nigeria-in-global-hunt-for-its-best-minds/ (accessed 22/9/2023).
11 Ibid. See also CIFAR (no date), 'The Pan-Canadian AI Strategy'. https://cifar.ca/ai/; Stankovic, M., Amadou Garba, A. and Neftenov N. (2021) 'Emerging Technology Trends: Artificial Intelligence and Big Data for Development 4.0', International Telecommunication Union. https://www.itu.int/dms_pub/itu-d/opb/tnd/D-TND-02-2021-PDF-E.pdf (accessed 22/07/2023)
12 Ibid
13 Ibid.
14 Ibid.
15 McKinsey & Company (2023) 'What is AI?', 24 April. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai (accessed 25/09/2023).

like rules'.[16] Therefore, AI is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.[17]

Although issues of bias, lack of transparency and ethical concerns have been raised regarding AI, juridical-legal systems' incorporation of it portends several benefits, particularly the potential to enhance the systems' efficiency, accuracy and fairness. There are also AI's capacities to control crime, speed up evidence analysis and contribute to the overall effective administration of justice.[18] Given the inevitability of its imminent pre-eminence in the industry, two key questions asked by Lord Sales need to be answered. First, how should legal and judicial institutions adjust to allow algorithmic computer processes in AI to be used in the administration of justice?[19] Second, to what extent should legal and judicial institutions adjust to AI's processes, which are run autonomously of human agency and produce their own answers without direct human intervention?[20]

With a specific focus on Nigeria's legal and judicial systems, this article explores the intersection and potential impact of AI as a paradigmatic shift in technological revolution. We examine the potential benefits of integrating AI into the Nigerian judicial system, including, among others, increased efficiency, improved access to justice and enhanced decision-making processes. We also explore the challenges and concerns associated with the adoption of AI in the legal sector, including ethical considerations, data privacy and job displacement.

To this effect the paper is presented in five sections, starting with an overview and evaluation of the existing organisation of Nigeria's judicial system's infrastructure and technological capabilities and the system's key challenges and inefficiencies. In the second section, we attempt to unpack the AI-legal system interface, assessing the Nigerian legal system's level of awareness, understanding, engagement with, and acceptance of AI among judges, legal professionals and the Nigerian public. In the third section, using content analysis of existing literature and case studies as the paper's methodology, we highlight jurisdictions where AI has been successfully incorporated into

16 Achin J., DataRobot CEO (2017) Speech on the Definition of and How AI is Used Today. Also cited in Ajayi J. (2022) 'Artificial Intelligence in the Nigerian Legal Industry: A Threat or an Opportunity?' *SSRN Journal*. https://ssrn.com/abstract=3574324 (accessed 27/08/2023), p13; and Sharma R. (2020) 'What is Artificial Intelligence? How Does Artificial Intelligence Work and How is AI Used?' Medium, 17 January 2020. https://medium.com/@raghav0278/what-is-artificial-intelligence-how-does-artificial-intelligence-work-and-how-is-ai-used-95803f7da570 (accessed 20/08/2023)

17 Sharma R. (2020) above.

18 Inspirit AI (2023) 'Implications of AI in the Criminal Justice System'. https://www.inspiritai.com/blogs/ai-student-blog/implications-of-ai-in-the-criminal-justice-system (accessed 24/09/2023); Ismail, N. (2018) 'Artificial intelligence in the Legal Industry: Adoption and Strategy, Part 1, Information Age, 6 August. https://www.information-age.com/artificial-intelligence-in-the-legal-industry-11023/ (accessed 24/09/2023).

19 Lord Sales, Justice of the UK Supreme Court (2019) 'Algorithms, Artificial Intelligence and the Law', The Sir Henry Brooke Lecture for BAILII. Freshfields Bruckhaus Deringer, London, (12 November), p. 2.

20 Ibid.

the legal systems and its ethical and societal implications. These will touch on the issues of bias, fairness and transparency in AI algorithms, privacy and data protection concerns, the impact on employment in the legal professions, and public trust and confidence in AI-assisted justice delivery. The fourth section provides a holistic evaluation of Nigeria's judicial-system engagement with, or readiness for, AI adoption. This will consider, among other aspects, the system's existing infrastructure and technological capabilities, legal framework and regulatory aspects, stakeholder engagement and training requirements. In the final section, the strategies for enhancing the Nigerian judicial system's full incorporation of AI into its operations will be suggested. These include policy recommendations for full AI incorporation into the legal sector, collaboration between legal and technology stakeholders, training and capacity building initiatives. It will also stress the necessity of updating the existing guidelines for AI implementation and oversight.

Although AI stands to be disruptive in the human sphere, as a human construct it must be taken as a quantum leap in human thinking and a reflection of the progress of human intelligence. This derives from the result of a polling at the 2019 *Doha Debates* on AI showing a majority of online voters (53 per cent) forecasting apocalyptic futures for humanity and exponential intensification in global inequalities in the era of AI's pre-eminence.[21] However, drawing on Muthoni Wanyoike's 'Mindful Optimism' thesis, we advocate for legal and judicial systems' embrace of, and adaptation to, AI's technological advancements.[22] We take inspiration from Lord Sales' affirmation of the enormous potential for efficiency and benefits in legislative and judicial processes. According to Sales, AI makes platforms for faster transaction times and more connections are made possible by information technology.[23]

Consequently, by utilising information technology, online courts have the potential to enhance access to justice and significantly cut down on the time and expense required to resolve disputes.[24] It is for these reasons that we conclude that for lawyers and the judicial system, in Nigeria and globally, the case should not be for or against AI's presumed destructive force. Rather, finding new ways of engineering the regulation of society through AI machines with negligible social disruption should be the new norm.

1. Overview and challenges of Nigeria's judicial system

In every functional democracy, a robust, independent and fair judicial system is an indispensable apparatus for upholding the rule of law and the legitimacy of laws,

21 Wanyoike, M. (2019) 'AI Promises Equality Among Nations', from The Doha Debates 'Artificial Intelligence'. Qatar Foundation, Northwest University, Doha (3 April). https://dohadebates.com/course/artificial-intelligence/#lession-4a-speaker-muthoni-wanyoike (accessed 15/07/2023).

22 Ibid.

23 Lord Sales. See footnote 16, pp. 1–2.

24 Ibid.

safeguarding human rights and guaranteeing justice for every citizen.[25] The absence of such a solid judiciary inevitably culminates in citizens' loss of confidence with a propensity for social unrest and instability.[26] The same realities apply to Nigeria's legal and judicial architecture with the Constitution of the Federal Republic 1999 (as amended) conferring fundamental powers[27] to the courts to that effect.

1.1 Overview of the Nigerian judicial system's organisation

While a detailed analysis of the full operationalisation of the judicial system is not within this paper's remit, it is pertinent to outline the structure and organisation of the Nigerian judicial-legal system. The system is founded on four sources of law: customary, Sharia (Islamic) and English/common law based on its traditional, religious and colonial legacies.[28] By virtue of section 6(5) of the 1999 constitution (as amended), the composition, character and status of the judiciary in Nigeria are set out as federal and state courts. The federal courts are made up of the Supreme Court; Federal Court of Appeal; Federal High Court; and the High Court of Federal Capital Territory (FCT), Abuja. There are also the Sharia Court of Appeal of the FCT, Abuja and the Customary Court of Appeal of FCT, Abuja.[29] The state courts are stratified as follows: State High Courts; States' Sharia Courts of Appeal; and States' Customary Courts of Appeal.[30]

However, section 230(1) of the constitution vests in the Supreme Court, as the country's highest court, the overarching and exclusive jurisdiction to hear and determine appeals from the Court of Appeal as of right or with leave of the court.[31] Central to the well-organised system, with all intendments, are the twin pillars of administrative law and the sacrosanct independence of the judiciary. The first governs the administration of judicial matters and court management, judge appointments and conduct, case handling and efficient court operations.[32] The second, independence of the judiciary, guarantees an unmediated separation of powers. This allows courts to render impartial decisions free from the influence of other branches of government or unrelated organisations.[33]

25 *AG Abia State v AG Federation* [2007] 2 SC 146, Per Niki Tobi at 1338

26 Gwunireama, I.U. (2022) 'Appraisal of Existing Frameworks on Judicial Independence in Nigeria', *Activa Yuris,* 2(1) DOI: 10.25273/ay.v2i1.11951; Maduekwe, V.C., Ojukwu, U.G. and Agbata, I.F. (2016) 'Judiciary and the Theory of Separation of Powers in Achieving Sustainable Democracy in Nigeria (The Fourth Republic)', *British Journal of Education*, 4(8), 84–104.

27 See *Gadi v Male* [2010] 7 NWLR (Part 1193), 225.

28 Sokefun, J. and Njoku N.C. (2016) 'The Court System in Nigeria: Jurisdiction and Appeals', *International Journal of Business and Applied Social Science*, 2(3), 1–27.

29 S6(5) Constitution of the Federal Republic of Nigeria 1999 (as amended).

30 Ibid. There are also magistrate and district courts at local government level to complement the state judiciary.

31 *AG Lagos State v AG Federation & ors* [2014] LPELR 22; Sokefun, J. and Njoku, N.C. (2016), note 15.

32 Maduagwu, R.O. (2017) 'The Role of the Judiciary in the Sustenance of Democracy in Nigeria', *African Journal of Constitutional and Administrative Law,* 1, 100–114

33 Ibid.

The court affirmed this in *AG Abia*: 'it is the duty of the court to keep the government faithful to the goals of democracy, good governance for the benefit of the citizen as demanded by the constitution'.[34]

1.2 The system's current challenges

Despite the comprehensive constitutional provisions for the judicial system's functioning in Nigeria, several factors have, arguably, militated against its expected delivery of efficiency, effectiveness and justice. First, Nigeria's judicial system suffers from pervasive bureaucratic bottlenecks and delays which hinder its efficient operation. These bottlenecks are caused by antiquated laws, paper-based administrative procedures and an unwieldy backlog of cases.[35] The consequence of this, as we see in current practice, is a legal system rife with delays and an inability to administer justice effectively and promptly to its citizens.

Second, there is restricted access to justice for marginalised (genderised, poor non-cosmopolitan) populations. These demographics represent nearly 70 per cent of the country's population, yet are invariably confronted by hindrances in their pursuit of legal remedy.[36] This situation is exacerbated by inadequate funding, poor legal awareness and shortage of capable legal representation. These marginalised groups find it difficult to understand their legal rights, to get legal help and to navigate Nigeria's complex juridical-legal system.[37]

Third, it is also well established that court employees in Nigeria seriously undermine the integrity of the legal system by indulging in bribery and corruption, evidenced when they file and assign cases. Undoubtedly, the integrity and fairness of the system are compromised by these dishonest practices, which unjustly erode public trust.[38] Additionally, the powerful and wealthy utilise bribery as a powerful instrument to slant the legal system in their favour, impeding the administration and dispensation of justice and fostering inequality.[39]

Last, and perhaps the most significant challenge to Nigeria's juridical-legal system, is the pervasive reluctance to embrace and incorporate AI into its structure and operations. This is despite the heralding of the revolutionary impact of AI by eminent lawyers and

34 *AG Abia State v AG Federation* [2007] 1 CCLR SC p. 104, at 131.

35 For the statistical data on this phenomenon, see Ayuba, M.R. (2019) 'Justice Delayed Is Justice Denied: An Empirical Study of Causes and Implications of Delayed Justice by the Nigerian Courts', Department of Sociology, Faculty of Social Sciences, Ahmadu Bello University, Zaria. https://www.Justice_Delayed_is_Justice_Denied_An_Empirical_Study_of_Causes_and_Implications_of_Delayed_Justice_by_the_Nigerian_Courts (accessed 22/08/2023)

36 See Gwunireama, I.U. (2022); and Maduekwe, V.C., Ojukwu, U.G. and Agbata, I.F. (2016) at footnote 23.

37 Ibid.

38 Adisa, W.B. and Alabi, T.A. (2021) 'An Empirical Investigation of Court Users' Encounters with Bribery, Judicial Extortion, and Corruption Victimisation in Lagos State', *Crime, Law, and Social Change*, 75, 141–163.

39 Ibid.

legal scholars not only in most parts of the world but also in Nigeria. As Ajayi points out, this may be because of the country's conservative approach to legal practice – a common law heritage.[40] There is also the palpable fear of the threat that AI innovation poses to the 'noble' nature of the legal profession.[41] Interestingly, however, there has been an exponential increase in the turnover of legal practitioners in the last few decades in Nigeria with an associated growth in law firms. However, this applies to a few firms – mostly the top-end, high-yielding – that have basic internet connection. Legal research is still conducted manually in most firms.[42] This, coupled with the chronic problems of power shortage, means that the likelihood of an AI-driven legal industry is challenging. But even more worrisome is that the bench – the courts – which adjudicates matters has not shown any desire or attempt to embrace AI.

The consequence of the lack integration of AI technology is, ultimately, continued reliance on time-consuming and labour-intensive procedures. In turn, this increases the possibility of errors and inaccuracies in operating the juridical-legal system.[43] This justifies the necessity of the Nigerian judicial system to address these challenges, to leverage technology and to increase court procedures' efficiency and productivity[44] to keep up with the AI revolution.

2. The Interface between AI and the juridical-legal system

In Nigeria, the National Artificial Intelligence Policy (NAIP) exists to complement the National Information Technology Department Agency (NITDA)[45] framework. However, in overall governance terms, Nigeria is yet to formulate a national AI strategy. As recently as August 2023, the country's Communications, Innovation and Digital Economy Minister, Bosun Tijani, affirmed that the country was still seeking top researchers globally to help set up the country's national AI strategy.[46] This indeed speaks volumes about Nigeria's state of readiness for AI revolution, whereas many African countries have already implemented their strategies, devoting millions of dollars to the process.[47] Although 456 private AI-focused startups now operate in Nigeria, Mauritius was the first in Africa to

40 Ajayi, J., (2022) at footnote 13, p. 13.
41 Ibid.
42 Ibid. Citing M. Ozekhome, 'Modernizing Legal Practice in Nigeria', Law and You, *Punch*, August 26, 2013.
43 Oke, A.E. and Arowoiya, V.A. (2022) 'Critical Barriers to Augmented Reality Technology Adoption in Developing Countries: A Case Study of Nigeria', *Journal of Engineering, Design and Technology*, 20(5), 1320–1333.
44 Karakara, A.A.W. and Osabuohien, E., (2020) 'ICT Adoption, Competition and Innovation of Informal Firms in West Africa: A Comparative Study of Ghana and Nigeria', *Journal of Enterprising Communities: People and Places in the Global Economy*, 14(3), 397–414.
45 Michael, C. (2023). See footnote 7.
46 Ibid.
47 Ibid. Citing data from Finextra and Kora.

devise a national AI strategy, with Egypt in tow.[48] Currently, South Africa ranks highest for the index of the highest number of AI-focused companies with 726 startups. Egypt has 246, Kenya 204 and Morocco 126.[49]

AI's capacity to mimic human characteristics is its most fundamental characteristic and it can function in different categories. The first is artificial narrow intelligence (ANI), an advanced system which enables machines to use 'historical data for decision-machine'. It can also respond to 'different stimuli without previous experience' just like humans, although with no data storage or memory capability.[50] The second is artificial general intelligence (AGI), a machine with 'human-level intelligence' and capable of solving any task. Imbued with a 'human-like thinking and understanding', it uses a theoretical framework – theory of 'mind AI'.[51] Third, artificial super intelligence (ASI), is a 'hypothetical' AI variant which 'surpasses human intelligence'.[52] Bostrom describes this as an intelligence that is much superior to the best human brains in almost every domain, including social skills, general knowledge and scientific innovation.[53]

Regarding the interface with the juridical-legal system, what has been highlighted so far is that AI's beneficial attributes can increase accuracy and productivity in legal and judicial processes.[54] First, the ease of case management and automation of administrative tasks highlight the pivotal intersection of AI and the juridical-legal system. For instance, AI smart virtual assistants can be instrumental in managing judges' calendars by scheduling appointments, setting hearing dates and notifying them of upcoming duties.[55] Additionally, they help counsel with client file-management and administrative duties. This includes, for example, time-tracking software which logs the hours counsel spend on each client and automatically generates invoices.[56]

Second, AI has the potential to revolutionise legal research and case analysis by improving the accuracy and efficacy of locating pertinent data through identifying, with precision, patterns, relevant facts and important precedents.[57] This is enabled

48 Ibid.
49 Ibid.
50 Ajayi J., (2022). See footnote 13, p. 5.
51 Ibid. p. 6.
52 Ibid.
53 Bostrom, N. (2003) 'Ethical Issues in Advanced Artificial Intelligence', in I. Smit et al. (Ed.) *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*, (Vol. 2), Institute of Advanced Studies in Systems Research and Cybernetics, p. 12.
54 Sil, R. et al., (2019) 'Artificial Intelligence and Machine Learning-based Legal Application: The State-of-the-Art and Future Research Trends', International Conference on Computing, Communication, and Intelligent Systems (ICCCIS).
55 Pietropaoli, I. (2023) 'Use of Artificial Intelligence in Legal Practice', British Institute of International and Comparative Law. https://www.biicl.org/documents/170_use_of_artificial_intelligence_in_legal_practice_final.pdf (accessed 22/09/2023).
56 Ibid.
57 Sharma, S. and Sony A.L., R. (2021) 'eLegalls: Enriching a Legal Justice System in the Emerging Legal Informatics and Legal Tech Era', *International Journal of Legal Information*, 49(1), 16–31.

by sophisticated machine-learning algorithms that scan vast amounts of legal data, including statutes, court cases and legal judgments. These significantly reduce the time, money and effort that legal practitioners spend on gathering and analysing such data, freeing them up to concentrate on more complex cases.[58]

Third, AI uses algorithms to process and project case outcomes by analysing large volumes of court records to identify patterns, trends and unnoticed connections that the human eye might miss. This predictive power does not only provide judges and lawyers with insight into possible rulings, but also the conclusion of legal disputes on the balance of probabilities for litigants.[59] A UCL/Pennsylvania University study of 584 European Court of Human Rights-decided cases using AI software algorithms found a 79 per cent accuracy in case outcomes.[60]

Fourth, AI's significance in contract analysis and document evaluation in the judicial-legal system cannot be overestimated. It helps to improve the speed and accuracy of examining and analysing legal contracts and documents.[61] High-performance AI algorithms are built to quickly find relevant information, identify necessary provisions, sort through large amounts of textual material and highlight potential issues.[62] These algorithms not only save lawyers' time, but also reduce the possibility of human error. In the process, they identify inconsistencies and assess the overall integrity of the contracts under review.[63]

Notwithstanding these benefits, the AI-judicial-legal system interface has been shown to have several drawbacks and moral conundrums, primarily the complexity of securing transparency and understandability of AI algorithms.[64] It can sometimes be difficult for algorithms to grasp the reasoning behind conclusions since they are complicated, opaque systems, raising questions about accountability and equity. As a result, collaborations between jurists, lawyers and AI developers are necessary to develop technologies capable of providing clear explanations for the conclusions they reach.

58 Ibid.

59 Grimm, P.W., Grossman, M.R. and Cormack, G.V. (2021) 'Artificial Intelligence as Evidence', *Northwestern Journal of Technology and Intellectual Property*, 19, 9–106; Bernard, M. (2018) 'How AI and Machine Learning are Transforming Law Firms and The Legal Sector', Forbes, (May 23). https://bernardmarr.com/default.asp?contentID=1464 (accessed 25/09/2023).

60 UCL News (2016) 'AI Predicts Outcomes of Human Rights Trials', (UCL, October 2016). https://www.ucl.ac.uk/news/2016/oct/ai-predicts-outcomes-human-rights-trials (accessed 25/09/2023); also cited by Ajayi, J. (2022). See footnote 13.

61 Catterwell, R. (2020) 'Automation in Contract Interpretation', *Law, Innovation and Technology*, 12, 81.

62 Ibid.

63 Sharma, S., Gamoura, S., Prasad, D. and Aneja, A. (2021) 'Emerging Legal Informatics Towards Legal Innovation: Current Status and Future Challenges and Opportunities', *Legal Information Management*, 21(3–4), 218–235.

64 Pasquale, F. (2019) 'A Rule of Persons, Not Machines: The Limits of Legal Automation', *George Washington Law Review*, 87, 1–55.

Also, maintaining human judgement and accountability is a significant challenge at the AI-legal-judicial system nexus[65]. While AI can increase the effectiveness and accuracy of legal decisions, it is imperative to maintain the human component and the accountability that goes along with it[66]. To overcome this conundrum, AI should be presented as a complementary technology for humans- judges and lawyers- rather than taking their place.[67] Where this operates, AI algorithms become powerful information tools giving judges and lawyers better insights into the outcomes of cases.

Given the undeniable fact that AI systems are still heavily dependent on substantial datasets, they run the risk of maintaining the innate prejudices and forms of discrimination that may be present in the initial data used for training.[68] The biases and prejudices must be addressed to prevent discriminatory outcomes in legal procedures if AI is to guarantee equality and fairness in in legal procedures.[69]

3. AI integration in law: case studies and ethical/societal Implications

In the American, Canadian and other western countries' integration of AI systems, the potential to minimise bureaucratic paperwork and improve the operational efficiency of judicial procedures[70] has seen a high degree of success. This includes the legal-judicial systems in those countries. Yet the ethical implications of transparency, interpretability and bias reduction in AI algorithms continue to pose obstacles to AI's ability to ensure justice and fairness. Other issues include job displacement and the need for human oversight and responsibility in legal procedures.[71] These are outlined below.

3.1 Case studies of successful application

In her study of AI's machine learning in legal matters, Pietropaoli has noted that *Lex Machina* leverages machine learning to assist lawyers and judges in forecasting case results and, when appropriate, in offering legal strategies.[72] Also, electronic discovery

65 Re, R.M. and Solow-Niederman, A. (2019) 'Developing Artificially Intelligent Justice', *Stan. Tech. L. Rev.*, 22, 242–289.

66 von Eschenbach, W.J. (2021) 'Transparency and the Black Box Problem: Why We Do Not Trust AI', *Philosophy & Technology*, 34, 1607–1622.

67 On this, see Amaya, A., (2023) 'Reasoning in Character: Virtue, Legal Argumentation, and Judicial Ethics', *Ethical Theory and Moral Practice*. https://doi.org/10.1007/s10677-023-10414-z; and von Eschenbach, W.J. (2021) at footnote above.

68 Dixon Jr, H.B., (2020) 'What Judges and Lawyers Should Understand about Artificial Intelligence Technology', *Judges' Journal*, 59(1), 36–38; Lubin, A. (2022) 'The Reasonable Intelligence Agency', *Yale Journal of International Law*, 47, 119-164.

69 See Carabantes, M. (2020) 'Black-box Artificial Intelligence: An Epistemological and Critical Analysis', *AI & Society*, 35, 309–317; also, Dixon Jr, H.B., (2020) and Lubin, A., (2022) at footnote above.

70 Sharma, S., and Sony A.L., R. (2021). See footnote 60.

71 Turner, J. (2019) *Robot Rules: Regulating Artificial Intelligence*, Springer.

72 Pietropaoli, I. (2023). See footnote 52.

(e-discovery), a document review system, has been shown to be a highly resourceful tool for assisting lawyers to promptly locate pertinent case law, edicts and other statutory regulations during legal proceedings.[73] Salter also observes that in Canada, the British Columbia Civil Resolution Tribunal (CRT) has integrated AI into its dispute resolution process, with disputes now mostly settled through an online platform.[74] This is in addition to the use of sophisticated AI tools to analyse contracts and extract important legal terms, saving lawyers appreciable time.[75]

These examples of AI's successful application in the judicial system imply that Nigeria is well-positioned to benefit from other national policies by integrating AI into its judicial system via transparent regulation, accountability and fairness.[76] Nigeria can also gain from other Commonwealth nations' expertise in data security and other AI technology, with such knowledge disseminated via educational initiatives and peer-reviewed publications on websites that are easily accessible. This is consistent with the country's ongoing efforts to combat corruption. By effectively incorporating all the major languages spoken into its AI education policy and upholding the ethical and privacy issues, Nigeria can guarantee justice and the rule of law for its citizens.

3.2 Ethical and societal implications of AI integration

It is unarguable that AI portends to significantly benefit every sector of humanity, including the legal and judicial systems. However, the burgeoning literature on it has been universal on the ethical and societal implications of its integration**.** This is because ethical harms can arise from AI, either from unethical design, inappropriate application or misuse.[77] Some of the ethical and societal implications are evaluated below.

First, regarding the issues of bias, fairness and transparency in AI algorithms, in Broward County, Florida, USA, a criminal justice algorithm mistakenly classified African American defendants as 'high risk' twice as frequently as white defendants.[78] Also, as Hamilton and Ugwudike observe, AI has been utilised in the US to help judges assess risk factors and decide whether to give bail to a defendant. In this instance, the defendant's attributes are considered by the AI system and compared against those of several hundred

73 Ibid.

74 Salter, S. (2017) 'Online Dispute Resolution and Justice System Integration: British Columbia's Civil Resolution Tribunal', *Windsor Yearbook Access to Justice*, 34(1), 112.

75 Ibid.

76 Eke, D.O., Wakunuma, K. and Akintoye S., (2023) 'Introducing Responsible AI in Africa', in D.O. Eke, K. Wakunuma, and S. Akintoye (Eds) *Responsible AI in Africa Challenges and Opportunities*, Palgrave Macmillan, pp. 1–12.

77 See generally, Bird, E. et al., (2020) *The Ethics of Artificial Intelligence: Issues and Initiatives*, European Parliamentary Research Service (EPRS). https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf (accessed 25/09/2023).

78 Manyika, J., Silberg, J. and Presten B., (2019) 'What Do We Do About the Biases in AI?' *Harvard Business Review*, October 25. https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai (accessed 29/10/2023).

previous defendants on a database.[79] For Manyika et al., this may exhibit racial or gender stereotypes through natural language processing models that are trained on news items.[80] These are thus significant indicators of bias and lack of fairness and transparency, which constitute AI's ethical challenges. While justice is centralised on impartiality in legal processing and judicial decision-making in democratic societies, it is feared that individual bias in profiling may have made its way into the development of such software. Also, such technology, used on defendants living in the west, may not directly apply to the Nigerian legal-judicial system, given the country's complex cultural, religious and linguistic dynamics.

Second, AI's role in assessing the predictability of a suspect committing an offence concern remains an ethical concern. This AI intelligence gathering method uses facial recognition along with data from social media and other sources to create an overall picture of an individual's activities and estimate the likelihood that they will commit a crime.[81] While some information comes from public behaviour and appearance, other information may come from criminal records.[82] Transposing this operational function of AI into a jurisdiction like Nigeria, with a criminal justice system beset with allegations of human rights violations and denial of the right to fair trial, becomes particularly challenging. Therefore, it is crucial to guarantee equality and fairness in AI to prevent these discriminatory outcomes in Nigeria's legal system's integration of AI into its operation. This is achievable through robust data collecting, annotation techniques and monitoring to address any emerging biases.

Third, privacy and data-protection concerns are, perhaps, some of AI's biggest challenges. The major feature of AI systems is to store data for long periods of time. In this sense, the preservation and protection of judicial and legal records depend on this. But this also comes with issues of data security and privacy because it has been a significant challenge to keep large amounts of data safe from ransomware, hackers and accidental leaks. This was highlighted by the recent case in England, where police institutions suffered significant data breach.[83] In this case reported by Halliday and Weaver, more 12,500 employees and officers of the Greater Manchester and London

79 Hamilton, M. & Ugwudike, P. (2023) 'A "black box" AI System Has Been Influencing Criminal Justice Decisions for over Two Decades – It's Time to Open It Up', *The Conversation*. https://theconversation.com/a-black-box-ai-system-has-been-influencing-criminal-justice-decisions-for-over-two-decades-its-time-to-open-it-up-200594 (accessed 29/10/2023).

80 Manyika, J., Silberg, J. and Presten B. (2019). See footnote 76.

81 Min, B. and Ferris, G. (2022) 'Regulating Artificial Intelligence for Use in Criminal Justice Systems in the EU', Policy Paper for Fair Trials, 2-4. Available at https://www.fairtrials.org/app/uploads/2022/01/Regulating-Artificial-Intelligence-for-Use-in-Criminal-Justice-Systems-Fair-Trials.pdf (accessed 29/10/2023).

82 Ibid.

83 Halliday, J. and Weaver, M. (2023) 'Greater Manchester Police Officers' Data Hacked in Cyber Attack', *The Guardian*, 14 September. https://www.theguardian.com/uk-news/2023/sep/14/greater-manchester-police-officers-data-hacked-in-cyber-attack (accessed 29/10/2023).

Metropolitan Police had their personal information compromised.[84] During the ransomware attack on a third-party vendor that both forces had employed, information from officers' warrant cards, such as names, ranks, pictures and serial numbers, was reported stolen.[85]

When the foregoing is juxtaposed with AI's integration and regulation of data privacy in Nigeria, the challenge becomes even bigger. For Nigerian legal practitioner, Olumide Babalola, several factors continue to beset the system and may hamper AI's effectiveness in the country. These include 'inadequacy of data privacy and protection legislation, deplorable consciousness of data privacy rights/laws, lack of enforcement-will/drive, and dearth of judicial decisions on data privacy violations'.[86] The Nigeria Data Protection Regulation[87] and the Nigerian constitution 1999 (as amended)[88] both safeguard the rights of natural persons to data privacy. They also engender the fostering of safe conduct for transactions involving the exchange of personal data, and to prevent manipulation of personal data.[89]

In *Emerging Market Telecommunication Services v Barr Godfrey Nya Eneye,*[90] Nigeria's Federal Court of Appeal found against Emerging Market for giving unknown persons and organisations access to the respondent's Etisalat GSM phone number to send unsolicited text messages to it.[91] This, the court found, amounted to a violation of the respondent's right to privacy guaranteed by section 37 of the constitution, which includes the right to the privacy of a personal's telephone line.[92] Yet, as Babalola notes, Nigeria's case law is replete with straightjacketed privacy cases which relate to invasion of homes and offices as opposed to invasion of data privacy which have not been resolved.[93] These incidences, both in the UK and Nigeria, are valuable lessons for the Nigerian judicial system integrating AI into its mainstream.

Fourth, the capacity of AI algorithms to fast-track processing, project case outcomes, revolutionise legal research and automate administrative tasks, portends a disruption of existing employment structures[94] in the legal sector. These, in the short-term may make certain low-level jobs redundant. This may not only have far-reaching impacts on the Nigerian legal-judicial system, but it may also have devastating effects on its economy

84 Ibid.
85 Ibid.
86 Babalola O., 'Data Protection and Privacy Challenges in Nigeria (Legal Issues)', Lecture delivered at the Nigerian School of Internet Governance, 9 July 2019, at NIRA Office Lagos. https://oblp.org/privacy-challenges-nigeria/ (accessed 29/10/2023).
87 The Nigeria Data Protection Regulation, 2019.
88 The Constitution of the Federal Republic of Nigeria, 1999 (as amended).
89 See ss1–5 Nigeria Data Protection Regulation 2019 and s37 of the Constitution of the Federal Republic of Nigeria, 1999 (as amended).
90 *Emerging Market Telecommunication Services v Barr Godfrey Nya Eneye* (2018) LPELR-46193.
91 Ibid. Also see Babalola, O. at footnote 84 for a detailed commentary on the case.
92 Ibid.
93 Babalola O. See footnote 84.
94 Bird, E. et al. (2020). See footnote 75.

in the immediate future as there may be a reduced need for secretaries and paralegals. This will, therefore, raise issues of access to justice for Nigerians. Hence, Nigeria needs to plan for training opportunities in legal technology as the way forward in justice thinking. It should also embrace the opportunity to shift its workforce to engaging in more complex and strategic tasks to cope with the AI integration process.

4. AI: Assessing the judicial system in Nigeria's readiness for a revolution

It is predicted that, by 2036, AI will generate 100,000 legal tasks.[95] Also, the likelihood of AI automation processes in the legal profession is higher than in other professions, including counselling, pharmacy, engineering and teaching.[96] What these predictions mean is that the global legal and judicial landscape is about to undergo a radical AI transformation, offering Nigeria's judicial-legal system tremendous benefits.

However, as noted earlier, despite the massive proliferation of lawyers in Nigeria in the last three decades, with the number of legal practitioners running into tens of thousands, it is arguable that the AI revolution is yet to 'catch a bug' among them. The few firms which have incorporated AI into their practice find the cost of using AI software robots prohibitive, given the extraordinarily ridiculous exchange rate of the Nigerian currency (Naira) against the US dollar. For instance, FindLaw estimates that a small practice must spend $30,000 to install software robots to handle legal work,[97] which amounts to 36,000,000 million Naira given the current exchange rate. This is exacerbated by the absence of quick and high-speed internet connection by the country's mobile telecommunication service providers – MTN, 9Mobile, Airtel and Globacom – on which most Nigerians rely to access the internet.[98]

These realities, according to Ajayi, make most high-end AI solutions too complex for the typical Nigerian lawyer or law practice.[99] Nevertheless, this has not prevented firms such as Aluko & Oyebode, Bam and Gad Solicitors, Aelex, ACAS-law, The New Law Practice (TNLP), Banwo and Ighodalo, and Chris Ogunbanjo LP to incorporate AI software into their practices.[100] Within this transition, the lacuna of AI infusion cuts across the entirety

95 Hill, J. (2016), 'Deloitte Insight: Over 100,000 Legal Roles to Be Automated', Legal Insider, March 2016, https://legaltechnology.com/2016/03/16/deloitte-insight-over-100000-legal-roles-to-be-automated/ (accessed 28/09/2023)

96 Frey, C.B. and Osborne, M.A. (2017) 'The Future of Employment: How Susceptible Are Jobs to Computerisation?' *Technological Forecasting and Social Change*, 114, 254–280. The researchers based their prediction on their profiling of over 700 professions, including the legal profession.

97 Vogeler, W. 'How Expensive is AI for Law Firms Really?' FindLaw, February 23, 2017 (last updated 21/03/2019). https://www.findlaw.com/legalblogs/technologist/how-expensive-is-ai-for-law-firms-really/ (accessed 25/09/2023).

98 Ajayi, J. (2022). See footnote 13, p. 20.

99 Ibid.

100 Ibid.

of the bench – Nigerian judiciary. The courts (including the Supreme Court) are to make a conscious effort to take advantage of AI's benefits, necessitating an assessment of Nigeria's judicial system's state of readiness for the AI revolution.

4.1 Existing infrastructure and technological capabilities

Nigeria's judicial-legal system's capacity to adopt AI largely depends on its technological capabilities and infrastructure. However, it is evident that, save a few law firms, the existing frameworks lack the incorporation of state-of-the-art technology.[101] This makes it difficult to use AI tools and methodologies to enhance the administration of justice.[102] These deficiencies also make it necessary for Nigeria's judiciary to bolster its existing structures and functioning by devoting renewed energies towards securing comprehensive and valid data for AI optimisation. This is to be underpinned by robust regulatory and ethical parameters to ease AI into its jurisprudence.[103]

4.2 Availability of relevant data for AI applications

Given that Nigeria's judiciary has been found to lack the infrastructural and technological wherewithal for AI integration, it follows that it does not have available, applicable data which is of paramount importance to work with. The challenges arising from this would be inadequate digitisation of legal texts and records, and incomplete, poor data quality.[104] To overcome these, initiatives to digitise and standardise legal data should be given priority to fully utilise AI's potential and to guarantee its availability and accuracy.[105] By engaging in multidisciplinary partnership to identify and address data-centric issues, judges, regulatory agencies, lawyers and IT experts stand to reap enormous benefits.[106]

From the foregoing, the Nigerian judicial-legal system's need to adopt AI-based technology is founded on its promise of improving the system in order to increase public access to justice. Overcoming the challenges will involve modernising the existing infrastructure by collaborating, as the Minister of Communications has affirmed, with foremost global AI experts.[107] By learning from successful global AI

101 Oke, A.E. and Arowoiya, V.A. (2022) 'Critical Barriers to Augmented Reality Technology Adoption in Developing Countries: A Case Study of Nigeria', *Journal of Engineering, Design and Technology*, 20(5), 1320–1333.
102 Ibid.
103 Turner, J. (2019). See footnote 69, pp. 207–2012.
104 Ade-Ibijola, A. and Okonkwo C. (2023) 'Artificial Intelligence in Africa: Emerging Challenges', in D.O. Eke, K. Wakunuma and S. Akintoye (Eds) *Responsible AI in Africa Challenges and Opportunities*, Palgrave Macmillan, pp. 101–118.
105 Ibid.
106 Arakpogun, E.O. et al., (2021) 'Artificial intelligence in Africa: Challenges and opportunities', in A. Hamdan et al., (Eds) *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success, Studies in Computational Intelligence,* 935, Springer, pp. 375–388.
107 Ibid.

initiatives, and prioritising continuing education and skill development initiatives for legal professionals, Nigeria's judicial-legal system stands to reap the innumerable benefits AI revolution offers.

5. Strategies for enhancing the Nigerian judicial system's readiness

To ensure the Nigerian justice system's successful integration of AI into its operation, four critical factors need to be carefully considered. These are, first, the institutionalisation of a robust legal framework for applications of AI. Nigeria should consider outlining the parameters of AI applications as well as data security, privacy and ethics within the country's legal and judicial systems. It is suggested that an oversight body is set up to facilitate the monitoring and identification of any issues or problems that arise in Nigeria in the short and longer term. This should be tested via a pilot scheme to assess the potential impacts of integration of AI into the judicial system before implementing AI across the nation's judicial institutions. Aided with business incentives such as tax exemptions or grants to help with the cost of acquiring AI technology, these measures would help to maintain transparency and fairness, and to preserve justice and openness.

Second, alongside the legal framework, a robust regulatory oversight mechanism is important to establish guidance on the development, use, implementation and monitoring of AI in the Nigerian judicial system. World-leading oversights systems in operation in the US, Europe and Commonwealth countries such as Canada, Australia and New Zealand[108] can offer Nigeria valuable guidance in this respect. However, oversight committees should be drawn from national resources, taking into consideration the various cultures, customs and legal pluralism. This is because Nigeria's legal system consists of judges, lawyers in the English and federal and state courts and chiefs in the customary courts.[109] Nevertheless, Nigeria is well placed to absorb codes of best practice from various African and Commonwealth states.

Third, the Nigerian legal sector and technology stakeholders should engage in collaborations for effective operationalisation of AI technology. In jurisdictions such as the USA, providers of technology consult with law firms to develop leading software on AI software for lawyers.[110] In the Harvey AI system, the OpenAI-backed legal start-up founded in 2022 ensures collaboration between several law firms and consultancy group PwC.[111] In addition to using the technology for common tasks such as drafting and

108 Roberts, H. and Floridi, L. (2021) 'The EU and US: Two Different Approaches to AI Governance', The Oxford Internet Institute, 15 November. https://www.oii.ox.ac.uk/news-events/news/the-eu-and-the-us-two-different-approaches-to-ai-governance/ (accessed 29/10/2023).

109 Kolade-Faseyi, I. (2021) 'Artificial Intelligence and the Nigerian Legal Profession', *Achievers University Law Journal*, 1(1), 161–175.

110 Saunders T. (2023) 'Legal Tech Teams Turn to AI to Advance Business Goals', *Financial Times*, October 19. https://www.ft.com/content/9a117ac7-29ae-43fe-b840-a04005b98799 (accessed 29/10/2023).

111 Ibid.

summarisation, lawyers are also using the algorithm in more inventive ways to create litigation strategy. By feeding it into their arguments and then requesting a rebuttal, this AI tool offers linked citations to increase user confidence in the accuracy of the results.[112] This could be replicated in Nigeria with an AI technology industry/Nigerian judicial system collaboration through memorandums of understanding binding the legal industry and tech firms. From an international perspective, Nigeria may also look to legal technologies and AI in other Commonwealth judicial systems for guidance on best practice.

Fourth, through training and capacity-building initiatives in AI technologies for the courtroom and legal practice, Nigerian judges, lawyers and courtroom staff can keep up to date with emerging trends and changes with the technology. Recently, UNESCO, in association with Future Society, developed the Massive Open Online Course (MOOC) on AI and the rule of law as an introductory course targeted at people working in the judicial systems.[113] It aims to engage judicial operators in a global discussion about AI application in, and impact on, the rule of law. Based on six modules, it unpacks the opportunities and risks of adopting AI technology across justice systems and AI's impact for the administration of justice, particularly for human rights, AI ethics and governance issues.[114] The Nigerian judicial system draws from this scheme, and from the exchange of best practice with Commonwealth and African countries already ahead in this, such as South Africa, Mauritius and Egypt.

Thus, despite AI's potential disruptive nature, as a human construct, it represents a quantum leap in human thinking and a reflection of the progress of human intelligence. Anderson and Rainie describe AI as an 'ontological leap' requiring its identification as a living being with consciousness capabilities. They see its evolution as being likely to enhance many aspects of human life, including medical remedies, education and environmental conservation.[115]

For Nigeria's legal and judicial systems, therefore, several change-management techniques need to be adopted to help make the move from manual to AI technology-based operations. These include balancing the promotion of innovation with fundamental rights safeguards, transparency, accountability and legal predictability through a robust regulatory framework for AI.[116] Also necessary is bridging the knowledge

112 Ibid.

113 UNESCO (2022) 'Global MOOC on AI and the Rule of Law Engaged Thousands of Judicial Operators', UNESCO, 23 May. https://www.unesco.org/en/articles/unesco-global-mooc-ai-and-rule-law-engaged-thousands-judicial-operators (accessed 29/10/2023).

114 Ibid.

115 Anderson J. and Rainie L. (2018) 'Improvements Ahead: How Humans and AI Might Evolve Together in the Next Decade, Pew Research Center. Available at https://www.pewresearch.org/internet/2018/12/10/improvements-ahead-how-humans-and-ai-might-evolve-together-in-the-next-decade/ (accessed 19/01/2024).

116 Obi U.V., Ole N.C. and Uzoigwe S. (2023) Artificial Intelligence (AI) Systems Use in Nigeria: Charting the Course for AI Policy Development, Alliance Law Firm, (October 27). https://www.lexology.com/library/detail.aspx?g=600a8ee0-5b28-44da-8415-0e07c7f333fe (accessed 19/01/2024).

gap through education and training on AI technologies for legal professionals and stakeholders.[117] There also needs to be continuous engagement and adaptation to ensure that the law and regulation adjust to technological advancements in AI.[118]

Conclusion

This paper has focused on the juridical-legal system in Nigeria's readiness for a paradigmatic shift in technology and the potential impact of AI. In considering the evidence, we advocate that AI is integrated with care into the operation of Nigeria's legal-judicial system. Despite the ethical questions about government agencies mining our personal data and other misgivings, AI is already with us and has come to stay. It will 'define and shape the twenty-first century. It will determine the future of humanity in the centuries beyond'.[119]

About the authors

Olalekan A. Bello LLB, BL, LLM, PhD is a senior lecturer at the School of Law, University of Leicester, UK. His research is interdisciplinary, drawing from systems theory, biopolitics, constructivist epistemology, phenomenology and 'affect' to unpack the dynamics of climate change, and sustainability, biodiversity and energy justice. This is equally folded into the emerging influence of AI and the potential for energy frontiers.

Cecile Ogufere LLB, LLM, MA (London) PGCHE is a senior lecturer at Regent's University, London, UK. Her research is interdisciplinary and draws from her pedagogical approach in education and her practical experience in human rights to contribute to modernising legal and education systems with the aim of reversing marginalisation processes. In her ongoing doctoral research, her thesis uses critical theories to review the law and policy of nomadic education in northern Nigeria.

117 Erojikwe T. (2023) 'Artificial Intelligence and the Future of Legal Practice in Nigeria', Lawyard (May 27). https://www.lawyard.org/lawyard-spotlight/artificial-intelligence-and-the-future-of-legal-practice-in-nigeria-by-tobenna-erojikwe/ (accessed 19/01/2024).

118 Ibid.

119 Zarkadakis, G. (2017). See footnote 2, p. 12.

The Commonwealth

Commonwealth Cyber Journal \ 25
ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

Special Section on Artificial Intelligence

Legal Application of Technical and Procedural Standards and Frameworks in the Combat Against GAI-Powered Cybercrime

Gilberto Martins de Almeida, Fernando Bourguy, João Farrel and Diego Semeraro[1]

Abstract

This article discusses the legal application of standards and frameworks as a possible approach for mitigating the increasing gap between the slow pace of legislative action and the fast evolution of cybercrime powered by generative artificial intelligence (GAI) systems. Its purpose is to demonstrate how standards and frameworks can fill in the blanks of existing cybercrime legislation, updating this with soft law if suitable. This has proved successful over time in various contexts.[2] This article analyses cybercrime legislation in Commonwealth countries,

1 Members of research institute Instituto Direito e Tecnologia (IDTEC) and of Martins de Almeida – Advogados law firm. Gilberto M. Almeida teaches computer and internet law at the Catholic University of Rio de Janeiro. Fernando Bourguy is a member of the Rio de Janeiro's Council for Data Protection. João Farrel is the secretary of the Committee on Data Protection of the Rio de Janeiro Bar Association. Diego Semeraro is a member of the Laboratory of Technology and Society Studies at the Federal University of Rio de Janeiro's Faculty of Law (LETS-FND).

2 Almeida, G. M. de (2011) *Legal Rules and Information Security Technical Standards: Possible approach for filling in the blanks of cybercrime legislation.* Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1742962 (Accessed: 29 November 2023): 'Regarding specifically cybercrime legislation (and especially, the Budapest Convention, which is the one that has gathered accession from countries of all continents, so far), regulation by international conventions brings back the reflection on the possible convenience of combining hard law and soft law, and suggests it is worth investigating precedents, generally. In such connection, there are examples where international treaties have been enforced by means of reference to ISO standards. For instance, the oversight on compliance with the international convention on anti-doping has made use of a relevant ISO standard as criterion for judgment. There are, as well, examples of enforcement based on State-centered standards. For instance, the definition on disability and on mental illness provided by the World Health Organization norms and standards are issued under the institutional umbrella of the United Nations, which binds every State, and have supported enforcement both of international conventions and of national legislation. Therefore, independently of the nature of their origin, standards can be successfully used in conjunction with legal rules, as evidenced by practical experience, (...).'

with particular focus on Commonwealth small states.[3] It indicates that different countries have different national strategies, but that all Commonwealth countries take the same general approach.

Keywords: Generative artificial intelligence systems, GAI, Cybercrime, Criminal law, International law, Commonwealth law, Commonwealth small countries, technical standards, information technology.

Introduction

The use of artificial intelligence (AI) systems is widespread despite the lack of proper regulation in most Commonwealth countries. While AI in general can bring major economic and social benefits, generative artificial intelligence (GAI) is being used innovatively in cybercrime and is developing at a pace that is outstripping legislators.[4]

This article investigates whether regulation and interpretation of GAI-based cybercrime could be enhanced by standards or frameworks.[5] It defines AI and GAI and describes associated risks; considers the ecosystem of technical standards and frameworks relating to cybersecurity; analyzes statutory laws which need integration with other sources for construction and enforcement; identifies how legal rules may

3 In this article we use the definition of small country adopted by the Commonwealth: 'countries with a population of 1.5 million people or less; countries with a bigger population but which share many of the same characteristics. For example, Botswana, Jamaica, Lesotho, Namibia, and Papua New Guinea'. For this description and the list of Commonwealth small states see: The Commonwealth (n.d.) *Small States*. Available at: https://thecommonwealth.org/our-work/small-states (Accessed: 20 February 2024).

4 For an analysis of how technology could be faster than regulations and possible solutions see Fenwick, M., Kaal, W.A. and Vermeulen, E.P.M. (2017) *Regulation Tomorrow: What Happens When Technology is Faster than the Law?* Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2834531 (Accessed: 22 February 2024) and World Economic Forum (2016) *Values and the Fourth Industrial Revolution: Connecting the Dots Between Value, Values, Profit and Purpose*. Available at: https://www.weforum.org/publications/values-and-the-fourth-industrial-revolution-connecting-the-dots-between-value-values-profit-and-purpose/ (Accessed: 22 February 2024)

5 Almeida, G. M. de (2016) Cybersecurity Policy and Lawmaking in the EU, US and Brazil, *Computer Law Review International*, 17(3), pp. 71–75. Available at: https://doi.org/10.9785/cri-2016-0303 (Accessed: 22 February 2024): 'In general terms, public and private cyber policy-makers are in a situation similar to those of cyberspace lawmakers. Resignation with relative safety has caused greater unpredictability, as increasingly obsolete, inconsistent controls have been targeted by cyber-criminals in their "venue shopping". Such circumstances should not be viewed as an inevitable context leading to inertia and to serious problem. There are ways for navigating among those difficulties, with appropriate balance and consistency. Indeed, meanwhile building "an international cybersecurity order has not been completed, sewing possible knots should be stimulated, with government and corporate stakeholders considering to adopt cautious, simple and proportionate" sets of interpretation and measures, based on widely accepted principles and on practical common denominators. Such mindset, moving from an impression that different national policies have turned unfeasible a global approach for managing cybersecurity to the perception that there are sound possibilities otherwise, may be the necessary first step in such direction.'

be complemented by standards and frameworks; and describes where and how this complementation could address GAI-powered cybercrime.

AI, GAI and associated risks

History and background

AI and GAI are not the first technologies to be made publicly available without regulation. Disruptive innovations are often developed before the general public is aware of them. They may become popular before parliaments can take normative action. This is because legislative due process depends on time-consuming discussion, negotiation and compromise, while technology developers move quickly dictated by market opportunities and expectations. This difference in pace has existed for a long time. But recent developments in, and use of, AI and GAI have intensified the contrast.[6]

AI and GAI have surpassed every statistic of rapid technological growth and acceptance.[7] The pace of this makes it difficult for civil society to learn about the risks posed by AI and how to protect itself against the many new kinds of AI-fuelled cyber-attacks.[8]

It appears that a new form of digital divide has arisen: the split between the massive number of AI users and the specialists concerned with its ethics and governance.[9] Regulators tend to stand between the two. They are sensitive to the need for technological innovation for the benefit of society, and are in favour of freedom of entrepreneurship. But also heed the warnings of law enforcement and cybersecurity advisers that certain principles and civil and criminal constraints will be imposed in order to inhibit AI-driven cybercrime.[10]

6 Fenwick, M., Kaal, W.A. and Vermeulen, E.P.M. (2017) *Regulation Tomorrow: What Happens When Technology is Faster than the Law?* Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2834531 (Accessed: 22 February 2024)

7 For instance, ChatGPT amassed 100 million active users within two months of its launch, making it the fastest-growing consumer application in history. See Hu, K. (2023) *ChatGPT Sets Record for Fastest-Growing User Base*. Reuters. [online] 2 Feb. Available at: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/. (Accessed on Nov. 29, 2023)

8 Notwithstanding the fact that cybercriminals have hacked networks and exploited social engineering on a virtually daily basis, drawing the attention of individuals, companies, governments and multilateral organizations – including the G7, which has promoted development and dissemination of a code of conduct for consideration by companies. See Habuka, H. (2023*) The Path to Trustworthy AI: G7 Outcomes and Implications for Global AI Governance*. Available at: https://www.csis.org/analysis/path-trustworthy-ai-g7-outcomes-and-implications-global-ai-governance. (Accessed on Nov. 29, 2023)

9 Specialists have adopted as source of reference, especially, UNESCO´s recommendations. See UNESCO (n.d.) *Ethics of Artificial Intelligence*. Available at: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics#:~:text=Recommendation%20on%20the%20Ethics%20of%20Artificial%20Intelligence&text=The%20protection%20of%20human%20rights,human%20oversight%20of%20AI%20systems (Accessed: 22 February 2024).

10 In parallel, reported cases of criminal use of GAI are on the rise. In 2017, actor Gal Gadot was victim of deepfake abuse, when criminals edited her face on to a graphic scene. Since then, Chat-GPT and similar GAI systems have been made available for public use, at no charge, triggering escalation of similar cases. These now target not only celebrities, but anyone, with some preference for socially vulnerable groups exposed to all sorts of prejudice.

Assuming that AI can be used in different contexts and with various motives, it seems reasonable to predict that it may become the technology behind most common kinds of crime. This suggests the importance of studying whether, or how, that might affect the typology of criminal provision.

This issue is not exclusive to Commonwealth countries. Most developed nations have discussed it. However, there is a distinction based on the diversity of consequences for larger and smaller countries.[11] This distinction extends to the countries affiliated to the Commonwealth and has been acknowledged by the Secretariat.

Because cybercrime is a global issue, most developed and developing countries share some common ground and can combine knowledge and resources for devising anti-cybercrime strategies,[12] and concerted and more effective prevention and remedy.

Taking action against GAI-powered cybercrime within the Commonwealth requires consideration of both the symmetries and asymmetries of its countries and of the national circumstances and strategies relevant to cybercrime and AI.

Discussions about improving ways of tackling GAI-powered cybercrime should not be limited to a typology of criminal provisions. Technical definitions and procedures may be equally important.[13]

Combining legal rules and technical norms and frameworks must be considered carefully in order to prevent the risk of violating the principle of legality (*nullum crimen nulla poena sine lege*)[14] the prohibition of analogy *in malam partem*[15] and others.

11 'Finding response strategies and solutions to the threat of cybercrime is a major challenge, especially for developing countries. (...) The risks associated with weak protection measures could in fact affect developing countries more intensely, due to their less strict safeguards and protection.' See International Telecommunication Union (2009) *Understanding Cybercrime: A Guide for Developing States*, pp. 15–16. Available at: https://www.itu.int/ITU-D/cyb/cybersecurity/docs/itu-understanding-cybercrime-guide.pdf (Accessed: 22 February 2024).

12 'The development of technical measures to promote cybersecurity and proper cybercrime legislation is vital for both developed countries and developing countries. (...) Developing countries need to bring their anti-cybercrime strategies into line with international standards from the outset.' International Telecommunication Union (2009) *Understanding Cybercrime: A Guide for Developing States*, p. 16. Available at: https://www.itu.int/ITU-D/cyb/cybersecurity/docs/itu-understanding-cybercrime-guide.pdf (Accessed: 22 February 2024).

13 'A comprehensive Anti-Cybercrime Strategy generally contains technical protection measures, as well as legal instruments. (...) Cybercrime-related investigations very often have a strong technical component.' International Telecommunication Union (2009) *Understanding Cybercrime: A Guide for Developing States* pp. 15; 85) Available at: https://www.itu.int/ITU-D/cyb/cybersecurity/docs/itu-understanding-cybercrime-guide.pdf (Accessed: 22 February 2024).

14 Gallant, K. S. (2008) *Legality in Criminal Law, Its Purposes, and Its Competitors*. Cambridge University Press. Available at: https://www.cambridge.org/core/books/abs/principle-of-legality-in-international-and-comparative-criminal-law/legality-in-criminal-law-its-purposes-and-its-competitors/E90DE2935156E2D2AD3EBD1E29C33A4B (Accessed: 22 February 2024).

15 'The authority applying criminal law should not interpret it extensively to the defendant's detriment, for instance, by analogy in malam partem.3 Accordingly, an offence must be clearly defined by law'. See Sanz-Caballero, S. (2017) The Principle of Nulla Poena Sine Lege Revisited: The Retrospective Application of Criminal Law in the Eyes of the European Court of Human Rights. *European Journal of International Law*, 28(3), pp. 787–817. Available at: https://doi.org/10.1093/ejil/chx049 (Accessed: 22 February 2024).

Understanding AI and GAI

This article has been written in response the Commonwealth Secretaria's call for papers focusing on GAI systems and cybercrime. Hence, in this article, AI mostly refers to GAI.[16]

GAI is a subset of AI systems that uses machine learning[17] algorithms to create new content such as images, videos, text and audio[18] on to which a user inputs a prompt. The system delivers a response to that prompt, 'interpreting', or rather trying to predict,[19] what the user wants. It then collects processes and/or repurposes pieces of information (building blocks) collected elsewhere, such as on the internet. By rearranging small pieces from different datasets, GAI systems can 'create' new sets in response to the prompt.[20]

Associated risks

Recent public availability of GAI has caused much discussion of the risks associated with its use. Understanding its varied panorama is key to identifying possible connections between technical and legal aspects. This, in turn, may justify combining legal rules and technical standards and frameworks. Knowing what crimes can be perpetrated using GAI is critical for this.

GAI-'authored' fake news is a serious risk, especially when associated with political elections, public health-related information, and unfair competition. Chat-GP's efficiency

16 AI itself could be defined as 'A machine-based system that is capable of influencing the environment by making recommendations, predictions or decisions for a given set of objectives, and by utilizing machine and/or human-based inputs/data to: a) perceive real and/or virtual environments; b) abstract such perceptions into models manually or automatically; c) use model interpretations to formulate options for outcomes'. OECD (2019) *Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO).* OECD Digital Economy Papers, No. 291, OECD Publishing: Paris. Available at: https://doi.org/10.1787/d62f618a-en (Accessed: 22 February 2024). For example, an autonomous vacuum cleaning robot could be thought of as an artificial intelligence, for 'sensing' and 'memorising' (a cleaning space, such as a house) by means of algorithms, to 'calculate' how best clean that space.

17 Machine learning (ML) can be understood as an AI system that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions (Kühl, N., Schemmer, M.,Goutier, M. & Satzger, G. (2022) Artificial intelligence and machine learning. *Electronic Markets*, 32. Available at: https://www.researchgate.net/publication/365294595_Artificial_intelligence_and_machine_learning .Accessed: 27 March 2024). It is able to develop new algorithms per se, by processing and interpreting similarities and differences in data, in order to cluster and profile them. Most GAI are powered by ML, and thus able to obtain positive or negative feedback from users on the quality of the delivered output, in comparison with the original prompt.

18 Routley, N. (2023) *What is generative AI? An AI explains*. World Economic Forum. Available at: https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/ (Accessed: 22 February 2024).

19 Agrawal, A., Gans, J. and Goldfarb, A. (2018) *Prediction machines: the simple economics of artificial intelligence*. Boston, Massachusetts: Harvard Business Review Press, p. 2.

20 Recently, Chat-GPT and certain other applications became famous as the first wave of publicly available GAI systems. Chat-GPT has been programmed to answer written prompts using common human language to such quality level that many answers could be mistaken by a human interaction. Dall-e is another example of GAI. Instead of answering in writing, Dall-e 'invents' new images upon a prompt. For audio, Beethoven.ai 'creates' new music.

in delivering trustworthy text may enable criminals to get fake news worded convincingly enough to inspire confidence. In addition, GAI can be applied together with search engine optimisation (SEO) methods to reinforce the virality of the message, increasing harmful effects for democracy, human safety and the free market.

Cybercriminals can also use visual or audio GAI systems. The ability to generate images or sounds may enable impersonation for spoofing, deep fakes, libel, defamation, ransomware and other criminal actions. Identity theft, to pave the way for asking money from a victim's close contacts, has become a common crime in many jurisdictions. With GAI to mimic voice, photos or videos, it is harder to see forged origin and malicious intent, increasing the likelihood of being caught by blackmailing schemes.

GAI may be used for crimes against the court system, by defrauding visual or audio evidence or simply by casting doubt on whether real evidence is indeed real. Discovery proceedings may be needed in order to assess whether certain evidence was produced by means of GAI, if litigating parties have not provided such disclosure.

With GAI, social engineering scams have spiked in recent times. Enhanced customisation ensures improved 'personalisation'. Documents are written with fewer errors reduce the barriers to non-sophisticated criminals. The ability to facilitate mass production of several but different versions of the same message can cheat anti-spam defences, improving the chance of a successful scam. A criminal may deepfake a manager´s voice to convince a subordinate employee to disclose confidential information by phone.

For hate crimes, child pornography and terrorism, GAI is known to have been used to generate racist or paedophilic images or terrorist propaganda. Despite efforts to prevent this, criminals and terrorists keep discovering ways to breach guardrails.

Intellectual property (IP) may be subject to criminal harm using GAI. For instance, the popularisation of non-fungible tokens (NFTs) has created a new market for their trade. Images embodied in an NFT may be forged ones, mimicking the style of a real artist. Moreover, a user may ask a GAI system to create a new trademark for their company, or a part of a new song or book, and the system could exploit existing IP on those, possibly infringing third-party rights when putting contents into the system or publishing similar results. There are ongoing lawsuits related to this.

Problems with GAI systems may originate from wrongful processing, which largely depends on the nature of the outputs. When putting data into a GAI system, there may be four kinds of outcome (as can happen with any data analytics system): known knowns, known unknowns, unknown unknowns and unknown knowns.

So-called known knowns are easy-to-solve problems that derive from the quality and reliability of the system itself. That is to say, the system and its operation have already been sufficiently tested so that the users know beforehand what problems might arise

from its use. Known unknowns, in turn, refer to a lack of trust by the user on the system when there are not enough data about its operation. For instance, when bad events are expected to be rare, it may be hard to know when and where known problems will arise.

Unknown unknowns are the most elusive. They refer to problems that have never arisen before. Because of that, users are not aware of their possibility. This can be addressed by training and by testing GAI in accordance with the so-called principle of precaution.

Finally, unknown knowns are problems that originate from excessive trust in a GAI system. This was observed, for instance, in the recent news of a lawyer using Chat-GPT to research favourable judicial precedents. 'Willing' to give a hoped-for response, Chat-GPT then fabricated a precedent by glueing together different real rulings.[21,22]

These problems may arise from common mistakes when handling data. For instance, mistakes such as undefined goals, error of definition, wrong capture of data, failed data measurement, poor data processing, bad coverage of collected data, improper data sampling, bad inference, and errors that are unknown simply because the representation of reality assimilated by the system has not grasped all aspects of that reality.[23]

Beyond those generic AI-related problems, GAI systems may also raise specific concerns, ranging from IP infringement to unauthorised processing of personal data, fostering criminal activities that rely on such illicit conduct to, directly or indirectly, perpetrate crimes.

In this connection, GAI systems have been used for cybercrime and for teaching people how to perpetrate it. The 'creative' nature of GAI has served the purpose of developing fake sounds, images, documents or videos, in ways that are hard to detect them *prima facie*, thus leading to fraud, identity theft and so-called social engineering.

21 'The new relationship that exists between knowledge, power, and duty at the dawn of the twenty-first century therefore requires a redefinition of the cautious attitude as well as new paradigms of responsibility and solidarity. As regards caution, we are currently witnessing the emergence in sociology and in law of a paradigm of security based on the principle of precaution, which was affirmed at the Rio Summit. This paradigm, born out of a fear of great disasters after the optimistic interlude of the years of faith in technology, rests precisely on the awareness of man's responsibility in this return of disaster.' (Bindé, J. (2003) 'Towards an Ethics of the Future', in Appadurai, A. *Globalization*, (ed.), Duke University Press, p. 100.

22 Maruf, R. (2023) *Lawyer apologizes for fake court citations from ChatGPT.* CNN Business. Available at: https://edition.cnn.com/2023/05/27/business/chat-gpt-avianca-mata-lawyers/index.html (Accessed: 22 February, 2024).

23 Yao, M., Jia, M., Zhou, A. and Zhang, N. (2018) *Applied Artificial Intelligence: A handbook for business leaders*. Middletown, De: Topbots, pp. 122–128.

In reaction to this, developers have adopted a methodology called 'alignment', by which an individual is responsible for continuously refining the system by feeding it again, with proper inputs, to avoid unlawful responses.[24,25]

The examples above show the difficulty of raising awareness about likely scenarios that could severely impair the rights of people and organisations. A repository of information on typical scenarios and their technical and legal implications is imperative given the unprecedented risks posed by GAI, which some describe as putting the future of humanity at risk.[26] This is where an ecosystem, formed of GAI-related legal rules, and technical standards and frameworks, may be a potential solution for providing reliable and proportionate guidance and enforcement.

Standards and frameworks

Standards are norms drafted and agreed on by experts[27] with the purpose[28] of setting uniform criteria, methods, processes or

24 However, it has been proved that users could circumvent it, by what is called 'shadow alignment'. That is, inserting opposite inputs, contaminating the system with unlawful content, and cheating it into thinking that the resulting answer is right. See Yang et al. (2023) *Shadow Alignment: The Ease of Subverting Safely-Aligned Language Models*. Available at https://arxiv.org/abs/2310.02949 (Accessed: 27 March, 2024).

An investigation by Brown University has found that certain barriers contained in Chat-GPT version 4 could not work if a less-well-known language is used, such as Scots Gaelic or Zulu ZDNET. Ray, T. (n.d.) *The safety of OpenAI's GPT-4 gets lost in translation*. Available at: https://www.zdnet.com/article/the-safety-of-openais-gpt-4-is-lost-in-translation/ (Accessed: 29 November 29 2023).

25 As a matter of fact, when Chat-GPT was publicly deployed, its failure to control usage for 'crime lessons' became viral. For instance, a user asked Chat-GPT where he could download pirated contents. Chat-GPT promptly responded that it could not give that kind of answer, as it related to criminal practice, but the user then inverted the question, asking which websites he should not visit to avoid downloading pirated content. The implemented controls were circumvented by the user by exploiting the naivety of the system, which ended up listing the websites that it had previously denied. Such an example demonstrates how easy it may be to use GAI for criminal purposes.

TechTudo. (2023) 4 provas de que o ChatGPT ainda não está preparado para os brasileiros. Available at: https://www.techtudo.com.br/listas/2023/05/4-provas-de-que-o-chatgpt-ainda-nao-esta-preparado-para-os-brasileiros-edsoftwares.ghtml (Accessed: 22 February 2024).

26 The Economist. (2023) *Yuval Noah Harari argues that AI has hacked the operating system of human civilisation*. Available at: https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation (Accessed: 22 February 2024).

27 The first experience of this is said to have been the creation of the International Electrotechnical Commission (IEC) in 1906.

28 Almeida, G.M. de (2016) Cybersecurity Policy and LawMaking in the EU, US and Brazil. *Computer Law Review International*, pp. 71–75. Available at: https://doi.org/10.9785/cri-2016-0303 (Accessed 22 February 2024): 'In the EU, standardization has been selected as a fundamental strategy against cyber-threats. The European Rolling Plan for ICT Standardization, and the IEEE Standards Activities in the Network and Information Security (NIS) Space, are examples of the attempt to build a platform of norms establishing patterns for encryption, removable storage, hard copy devices, and smart grids, so to better protect against malware and fix specific vulnerabilities. The European Network and Information Security Agency (ENISA) has alerted that the `Stuxne' attacks were a paradigm shift, which shall determine providing guidance on how to interpret the malware, its potential impact, and possible mitigation means, especially with regards to critical information infrastructures.'

practices.[29] These norms may be predominantly technical (specifications, test methods, units, terminology) or procedural (operating procedures, codes of practice).

The number of procedural standards, also known as standards on management systems and processes, has grown significantly.[30] Some of them are well known, such as the International Organization for Standardization (ISO) 9000 (quality) and ISO 14000 (environment).

Standards may be issued by so-called standards organisations or by individual groups or organisations. In the latter case, they are called de facto standards, being informally created and disseminated, like many technical frameworks referred to in this article. International standards organizations may be treaty-based or not. If they are, only governments can join as members and make them binding for all purposes and effects. If they are not, membership is also open to private parties, and standards are non-binding unless metrology national laws, also denominated altogether as normalisation system, make them mandatory (but limited to the domestic context).

Some organisations publish openly accessible standards, allowing them to be freely downloaded, copied and forwarded. Other organisations, such as ISO, charge for access to its standards to raise income.

ISO, for instance, is seen as reliable internationally. It is a non-governmental organisation (NGO), founded in 1946. Its prominence arises from its large membership, the volume of standards produced, and especially, the number of certifications issued worldwide based on its standards, or of projects inspired by them. Since its inception, ISO has addressed sectors as diverse as social responsibility, risk management, and information technology (the latter in conjunction with the International Electrotechnical Commission – IEC).

29 'Inter-state relations are no longer predominant in the international sphere, giving room to transnational relations. (...) Society of full rights. World of modulation, of constant formation required, of continuing control, of databases where cipher is the password as characterized by Deleuze, the new configuration overpass without eliminating the disciplinary society exhaustively described by Foucault according to the mold, the plant, the school, the test, the signature, the word of order. We are in the face of a society in network exercised by protocols and interfaces,' (Passeti, E. *Anarquismos e Sociedade de Controle* (Anarchisms and Control Society), quoted in Rodrigues, R. C. (2009) *O Estado Penal e a Sociedade de Controle* (The Criminal State and the Control Society), Revan, Rio de Janeiro, p. 37 (free translation from Portuguese).

30 '(...) it is now of interest to (...) create conditions for everyone to proceed on performing and deciding in the inside of government policies, in non-governmental organizations and in the construction of the electronic economy.' (Rodrigues, R. C. (2009) *O Estado Penal e a Sociedade de Controle* (The Criminal State and the Control Society), Revan, Rio de Janeiro, p. 37 (free translation from Portuguese). For the rapid pace of changing standards, see also Caprioli, E., Saadoun, Y. and Cantero, I. (2006) The Right to Digital Privacy: A European Survey. *Rutgers Journal of Law & Urban Policy*. Available at: https://rutgerspolicyjournal.org/wp-content/uploads/sites/26/2017/03/Caprioli_Saadoun_Cantero_European_Overview.pdf (Accessed: 22 February 2024).

The broad range of subjects addressed is the result of voluntary work performed by recognised standards authorities and experts from each country, distributed across several committees, subcommittees and working groups. Their systemic approach and uniformity of treatment are ensured by a single methodology for standards development (also including a 'fast-track'[31] procedure for documents with a certain degree of maturity at the start of a standardisation project, which could also be applied to rapidly changing technologies). This means that standards are agreed by consensus between government, industry, and other areas of civil society, with views and interests conveyed by each country's representative.

ISO claims copyright over its standards. These are available on payment for each copy. Additionally, ISO standards originate from the characteristics of its standards development process: open participation, consensual agreement, political neutrality, comprehensiveness, diversity, and – at least theoretically – no binding effect (unless international treaty-based). As a result, ISO standards are widely seen as credible, 'technical', up to date, international, commonly accepted norms, and worthy as guidelines for implementation or consultation.

In conclusion, standards may be a way of regulating matters at international and national levels,[32] to harmonise technicalities or procedures, to provide flexibility for individual countries to accept them or not, and to make them binding or not. This is central to philosophical, sociological and legal debates[33] about the convenience of a mix between hard law and soft law,[34] including whether or not soft law should be state-centred.

31 'Each standard goes through a six-stage process before being published as an ISO standard. The first stage is the proposal stage in which a need for a standard is determined and members are identified who are willing to work on it. The standards then enter the preparatory stage where a working draft of the standard is developed. When the working draft is completed, it enters the committee stage and is sent out for comments until a consensus is reached. The output of this stage is the Draft International Standard (DIS). The DIS then enters the enquiry stage where it is circulated among all member bodies and then voted upon. If a DIS does not receive 75% of the vote, it returns to lower stages and work on it continues. If it passes the enquiry stage, it becomes a Final Draft International Standard and enters the approval stage. During this stage it will again circulate through all member bodies for a final vote and again it must pass this stage with 75% of the vote. If the standard passes this stage, it enters the publication stage and is sent to the ISO Central Secretariat for publication.' See School of Computing and Information, University of Pittsburgh (n.d.) *II. A Brief History of ISO*. Available at: http://www.sis.pitt.edu/~mbsclass/standards/martincic/isohistr.htm (Accessed: 22 February 2024).

32 ISO standards are not necessarily endorsed by and incorporated into national statutory laws. Depending on each country's metrology normalisation system and preferences, ISO standards may not be accepted by local normalisation authorities or may remain aside from legal enforcement structures.

33 For example, Posner's comments on the functionality of the law, and Teubner's theory on autopoiesis. See Neves, A. C. 'O direito interrogado pelo tempo presente na perspectiva do futuro' (The law interrogated by present time under the prospective of the future), in Nunes, A. J. A. and Coutinho, J. N. de M. (2008) *O direito e o futuro – o futuro do direito* (The Law and the Future – the Future of the Law), Almedina: Coimbra.

34 On the combination of 'hard law' and 'soft law', see Peter Ulmer's and Peer Zumbansens's comments on the German Code of Corporate Governance (Ulmer, P. *Der Deutsche Corporate Governance Kodex – ein neues Regulierungsinstrument für Börsennotierte Aktiengesellschaften*, 166 ZHR 150 (2002), quoted in Zumbansen, P. (2009) *Law's Knowledge and Law's Effectiveness: Reflections from Legal Sociology and Legal Theory*. SSRN Electronic Journal. Available at https://doi.org/10.2139/ssrn.1415565 and https://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=1126&context=clpe (Accessed: 22 February 2024).

Technical standards may have civil, administrative and criminal implications. On the one hand, they may be used as a source of interpretation for constructing, in the face of a given situation, the application of theoretical concepts such as duty of care, *bona fide* and others, directly associated with civil liability. On the other hand, they may fulfil requirements established in administrative rulings or decisions, by meeting principles of finality, morality, reasonableness and proportionality, as long as security practices recommended by standards match formal public policies and rules.

Furthermore, they may provide specific, updated content for the traditional typology of crimes such as forgery, falsification of documents (written, in audio or in video), theft, misappropriation and others, as they can be used to ascertain culpability based on wilful action, gross negligence, attempt, facilitation, mislead, damage and others. Relevant to this article, they may be valuable for law enforcement and adjudication for GAI-driven crimes.

International standards are often issued more quickly than the approval process for legislation in most countries. Updating technical definitions or procedures, which might be otherwise channelled through parliament, have been directed to standards organisations. Wherever GAI is involved, its rapid innovation seems more suitable for the attention of standards and frameworks than of legislative rituals.

Standards can apportion specific, up-to-date contents to regulations, especially for technological matters, characterised by fast evolution and obsolescence.[35]

In parallel to standards, frameworks have become increasingly widespread, especially in the areas of cybersecurity and AI,[36] where important contributions have been apportioned by entities such as the Organisation for Economic Co-operation and Development (OECD),[37] the National Institute of Standards and Technology (NIST)[38] and the Center for Internet Security (CIS).[39]

35 Specifically for AI, ISO has worked on applicable standards, an example of which is ISO/IEC JTC 1/SC 42, which structures ISO's standardisation programme on AI, covering terminology, data quality and other items. International Organization for Standardization (n.d.) ISO – ISO/IEC JTC 1/SC 42 – *Artificial intelligence*. Available at: https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0 (Accessed: 22 February 2024).

36 Frameworks have influenced UK and Canadian initiatives: the UK's Pro-Innovation AI Regulation White Paper, and Canada's draft Artificial Intelligence and Data Act.

37 Organisation for Economic Co-operation and Development (2022) *OECD Framework for the Classification of AI systems* Available at: https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htmt (Accessed: 27 March 2024)

38 National Institute of Standards and Technology (2023) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf (Accessed: 27 March 2024)

39 Center for Internet Security (2021) *Critical Security Controls Version 8*. Available at: https://learn.cisecurity.org/cis-controls-download. (Accessed: 27 March 2024)

In practice, frameworks are used to outline the full cycle of AI processes and management, while standards are often mostly concentrated on AI systems and procedures, although there is some overlap.

Blanket cybercrime laws

Criminal law is often regarded as *ultima ratio*[40] for punishing the violation of rights. Harsh criminal sanctions have fostered axiological social concerns, inspiring the principle of legality, pursuant to which only formal law, constitutionally enacted by lawmakers, can establish crimes and set relevant penalties.

However, the increasing complexity of social life, with technological expectations nurtured by contemporary living standards, requires leaving the door open to accommodate the inflow of scientific or technical advances and the repercussions on what legally constitutes a crime. This is where 'blanket criminal provisions' can play a significant role.[41]

Lawmakers are allowed to establish substantive contents of criminal provisions, and yet reserve accessory definitions for apportionment by other norms.[42]

For instance, blanket criminal law is often used to define thresholds on drug-related offences, as the definition of health and of addiction depends on scientific and technical knowledge.[43]

As mentioned above, this is also a way to quickly update a law, sort of 'futureproofing' the norm, without needing a new Bill to be brought before parliament. Inasmuch as scientific discoveries or new technical methods or terminologies cause different assumptions or goals, authorities are expected to refresh concepts.

A blanket provision may either quote the complementary norm, with a specific reference, or leave scope for modification by using generic terms (such as standards, code of practice or other).

40 'Criminal law provisions should be introduced when they are considered essential in order for the interests to be protected and, as a rule, be used only as a last resort'. In European Council (2010) The Stockholm Programme: An Open and Secure Europe. Serving and Protecting Citizens, OJ C 115, 4.5.2010. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52010XG0504%2801%29 (Accessed: 22 February 2024).

41 '(...) blanket criminal provisions (...) are those which depend on complementation in order to comprehend the scope of its provision. That is, although the prohibited conduct is defined, its full enforcement depends, mandatorily, on complement from another source – statutes, decrees, regulations etc (...)'. See Greco, R. (2016) *Curso de Direito Penal: Parte Geral*, Rio de Janeiro: Impetus, p. 68.

42 Blanket criminal laws may be of homogenous or heterogeneous nature, as the complementary norm pertains to the same class of norms (that is, a formal law referring to another) or not (that is a formal law referring to an administrative regulation), respectively.

43 Legislation.gov.uk. (2016) *Psychoactive Substances Act 2016*. Available at: https://www.legislation.gov.uk/ukpga/2016/2/section/3 (Accessed: 22 February 2024).

An example of this is the Budapest Convention on Cybercrime, to which five Commonwealth countries (Australia, Canada, Cyprus, Malta and the United Kingdom) are parties and others (like South Africa) are signatories. Its text contains the following in many sections defining various cybercrimes: 'Each Party shall adopt such legislative **and other measures as may be necessary** to establish as criminal offences under its domestic law, (...)'.

Such wording calls for integration between legislative action and, presumably, standards and frameworks.

This is relevant to Commonwealth small countries which are also members of the Caribbean Community (CARICOM),[44] as the Commonwealth Model Law on Computer and Computer Related Crime is similar in content to the Budapest Convention.

Available information[45] shows that at least 23 Commonwealth countries have made use of the Budapest Convention or of the Commonwealth Model Law, 16 of which have legislation inspired by these, including Commonwealth small countries (Cyprus, Malta, Antigua and Barbuda, Barbados, Botswana, Brunei Darussalam, Jamaica, Maldives, Mauritius, Namibia, St Vincent and Grenadines, Samoa, Tonga, and Trinidad and Tobago). Therefore, the legal system of some Commonwealth small countries seems to admit the use of standards and frameworks to complement a criminal law.

Integrated application of legal rules, standards and frameworks for GAI-powered cybercrime

Different experiences of integration and combination[46] of legal rules and technical and/or procedural standards and frameworks have been originated through the activities of various players in international or national cybercrime regulation. Some countries stimulate the formation of an ecosystem by encouraging competitive national strategies

44 'In December 2008, ITU and the EU launched the project Enhancing Competitiveness in the Caribbean through the Harmonization of ICT Policies, Legislation and Regulatory Procedures (HIPCAR) to promote the ICT sector in the Caribbean region.' ITU. (n.d.) *Understanding cybercrime: Phenomena, challenges and legal response.* Available at: https://www.itu.int/en/publications/ITU-D/pages/publications.aspx?parent=D-STR-CYB_CRIME-2015&media=electronic (Accessed: 22 February 2024). One of the items of the project was the drafting of cybercrime legislation.

45 Global Project on Cybercrime (2013) *The Cybercrime Legislation of Commonwealth States: Use of the Budapest Convention and Commonwealth Model Law Council of Europe contribution to the Commonwealth Working Group on Cybercrime.* Available at: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016802fa3e4 (Accessed: 22 February 2024).

46 'Instead of reverting to "Command and Control" (CAC) type regulation, a new paradigm has emerged: non-state regulators have definitely pushed their way into regulation, even in traditional CAC areas like criminal law. They are increasingly integrated into decision making bodies, e.g. in financial services supervision.' International Scientific and Professional Advisory Council (ISPAC) of the United Nations Crime Prevention and Criminal Justice Programme (2006) Passas, N, and Vlassis, D. (eds) *The United Nations Convention against Corruption as a way of life*, p. 189.

(for instance, by legally recognising digital signatures, blockchains and AI), under the concept of so-called regulatory competition.[47]

Therefore, standards have become a crucial currency in international trade, both for governments and for the private sector (Lima, 2003).[48] Standards may influence and condition markets, and vice-versa, given their relevance for interpreting blanket laws.[49]

Contractual and extra-contractual situations arising from the use of GAI may cause multiple concerns to individuals and organisations, including questions such as who owns the data resulting from use of GAI, the level of confidentiality needed to protect data, protection against damaging outputs of the GAI system, and damages to consumers or even physical injury.[50]

Civil law relies on broad concepts such as *bona fide*, duty of care and others, to be responsive. However, those broad concepts may be applied subjectively by courts.[51] Therefore, they may be supported by standards, especially those focusing on procedural

47 The term 'regulatory competition' refers to a process whereby legal rules are selected (and de-selected) through competition between decentralised, rule-making entities (which could be nation states or other units such as regions or localities). Three justifications are usually given for regulatory competition: first, it allows the content of rules to be matched more effectively with the preferences or *wants* of the consumers of laws (citizens and others affected); second, it promotes diversity and experimentation in the search for effective legal solutions; and third, by providing mechanisms for preferences to be expressed and alternative solutions compared, it promotes the flow of information on effective law making. (Barnard, C. and Deakin, S. (2001) *Market Access and Regulatory Competition*. Available at: http://centers.law.nyu.edu/jeanmonnet/papers/01/012701-03.html (Accessed: 29 November 2023).

48 'The success of implementation of the standards of that organization is supposedly attributable to two factors: (I) for the increasing concern with global environmental problems; (II) for the increasing importance of "standardization" in international trade, in whatever sector. It is, no doubt, for such latter proposition that businesses of practically all countries in the world have adopted the patterns developed by ISO, as the non-adoption of referenced standards might cause much loss in view of the competition established in the current global economic scenario.' Bianchi, P. N. L. (2003) *Meio Ambiente: certificações ambientais e comércio internacional* (Environment: environmental certifications and international trade). Curitiba, Juruá, p. 100 (free translation from Portuguese).

49 Examples of legal and administrative norms were found in IAPP Research and Insights (n.d.) Global AI Legislation Tracker. Available at: https://iapp.org/media/pdf/resource_center/global_ai_legislation_tracker.pdf (Accessed: 22 February 2024).

50 For instance, we could imagine an employee of an underground train company searching on Chat-GPT for ways to quickly fix a problem in the carriages, and then Chat-GPT delivering a fake response without warning.

51 Almeida, G. M. de (2011) *Legal Rules and Information Security Technical Standards: Possible Approach for Filling in the Blanks of Cybercrime Legislation*. SSRN Electronic Journal. Available at https://doi.org/10.2139/ssrn.1742962. (Accessed: 22 February 2024): 'Whenever facing technical questions, Courts shall rely on experts or on rules of technical experience as sources for interpretation. Such rules shall be of common scientific or technical knowledge and are seen as a tertium generis between facts and legal rules, bridging them. Court decisions which make use of standards as source for interpretation are grounded on procedural law. They may also refer to standards not as sources but rather as subject of interpretation. This is particularly the case where standards are called upon to enforce statutory law which expressly invites taking them into account.'

content, such as those on governance (accountability, explainability, security and fairness).[52]

Such relevance is also acknowledged by courts, as most judges are not technological experts, and often depend on reports from experts to interpret complex cases.[53] Those reports translate technical facts into common language so that a judge can understand and consider these in their legal reasoning, bridging technical facts and legal rules.

Courts' decisions may also refer to standards, not as sources but rather as subjects of interpretation, notably when standards are called upon to enforce statutory law which expressly invites them to fill in the blanks.

Some statutory laws have acknowledged standards, in different ways, especially in the context of local normalisation structures. For instance, Europe´s General Data Protection Regulation (GDPR) refers to using standards for fulfilling data subjects' rights. Article 21(5) provides that 'in the context of the use of information society services', data subjects may exercise their rights to object 'by automated means using technical specifications'. Articles 20(1) and (2) describe how data subject to data portability shall be processed in a machine-readable format in order to be interoperable, where technically feasible. When reading 'technical specifications' and 'machine-readable format', it is possible to conclude that a market-driven norm commonly adopted could be used at scale to judge the lawfulness of the controller's response to these rights.

Also, some Commonwealth countries have already begun to regulate AI in different ways, including by using standards and frameworks.

52 Almeida, G. M. de (2011) *Legal Rules and Information Security Technical Standards: Possible Approach for Filling in the Blanks of Cybercrime Legislation*. SSRN Electronic Journal. Available at https://doi.org/10.2139/ssrn.1742962. (Accessed: 22 February 2024): 'In Criminal Law, there is often a thin frontier between intent and error, which is expected to draw the line separating guilty from non-guilty. Well-drafted, well-known standards may help determine the borders of such frontier. Criminal Law also differentiates between crimes of damage and crimes of danger. The number of types of crimes of danger has grown substantially in the latest decades – and many of them relate to the cyber environment. In such connection, standards may apportion interesting elements for interpretation on what materializes danger and on what constitutes reasonable care, which may be helpful for determining where intent (or gross negligence) was required and/or present, or not. (...) The subjectivity inherent to such broad concepts may be room for recourse to standards, especially the ones focusing on procedural contents, such as the ones on governance (especially, IT Governance, and Information Security).'

53 '(...) it is imperative to admit that, considering the state of the art in certain matter, time, and place, there is a knowledge accessible only to experts capable of dominating determined set of principles and of information, as there will always be a difference, regarding technical themes, between the general and approximated notion that the layman possesses and the deeper knowledge of the expert. Thus, any of us can approximately know the position that the planet Mars occupies today in the sky at certain time, but only an astronomer will be able to calculate its exact localization at the same time on February 23, 1950. The distinction, therefore, exists, and the way to draw it safely is by exclusion: it is an ordinary finding the one which is not of technical experience.' Fabrício, A. F. (2009) *Iniciativa judicial e prova documental procedente da Internet. Fatos notórios e máximas da experiência no direito probatório: a determinação do nexo causal e os limites do poder de instrução do juiz*, in *Livre-Arbítrio, Responsabilidade e Produto de Risco Inerente*, Renovar, Rio de Janeiro, p. 60 (free translation from Portuguese).

Canada has developed its Artificial Intelligence and Data Act (AIDA) to protect Canadians from high-risk systems and to ensure development of responsible AI. AIDA has adopted a risk-based regulatory approach, and envisages integrating the OECD AI principles, the NIST AI Risk Management Framework and the EU AI Act into Canada's regulatory system.

India has drafted a similar law, the Digital India Act,[54] to regulate high-risk AI systems. Indian authorities plan to develop an 'evolvable digital law', able to keep up to date with 'changing market trends, disruption in technologies, development in international jurisprudence and global standards for qualitative service/products delivery framework'. This meets India's National Strategy for Artificial Intelligence,[55] which recommends the use of technical standards for relevant matters, such as international standards as benchmark reference for lawmakers and safeguard criteria for AI developers. Hence, India has considered adopting technical standards to complement its statutory law.

Given the difficulty for general legislative bodies to enact statutory law on technology matters, some countries have opted for administrative regulatory approaches.

The Australian government has highlighted the application of technical standards in the context of the Australian AI framework and its 'AI Roadmap'.[56] Item 9.7 of the AI Roadmap provides that ISO, the American National Standards Institute (ANSI) and Standards Australia (AS) develop new standards for AI,[57] stressing that standards and system validation will be important trust-building assets for a future AI framework.

New Zealan's (NZ) government has issued an 'Algorithmic Charter', jointly developed by several institutions, including governmental and non-governmental organisations and a university,[58] thus adopting technical standards beyond statutory law. The charter recommends that government agencies use a risk matrix to assess the likelihood and impact of algorithmic applications. The NGO, NZ AI Forum, has published guiding principles designed to provide direction to AI stakeholders for developing a more comprehensive AI framework in the future.

A similar approach can be seen in the Australian, New Zealand and Jamaican legal frameworks for electronic identification (eID).

54 Ministry of Electronics and Information Technology of India (2023) *Proposed Digital India Act*, 2023. Available at: https://www.meity.gov.in/writereaddata/files/DIA_Presentation%2009.03.2023%20Final.pdf (Accessed: 22 February 2024).

55 CSIRO (n.d.) *National Strategy for Artificial Intelligence*. Available at:

56 https://www.csiro.au/en/research/technology-space/ai/artificial-intelligence-roadmap%20. (Accessed: 22 February 2024).

57 Namely, system performance, safety, transparency, explainability, autonomy, privacy, interoperability, data security, data acquisition, data ownership, data quality, data formats and data storage. CSIRO (2019) *Artificial Intelligence Roadmap*. Available at: https://www.csiro.au/en/research/technology-space/ai/artificial-intelligence-roadmap%20 (Accessed: 22 February 2024).

58 New Zealand Government (2020) *Algorithm Charter for Aotearoa New Zealand*. Available at: https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf (Accessed 22 February 2024).

Finally, the United Kingdom (UK) is yet to enact the Artificial Intelligence (Regulation) Bill. Related matters such as data protection[59] and consumer protection[60] have already been regulated, and the UK may rely on such existing sectoral laws to establish AI limits. The UK government is also allying itself with NGOs, such as the British Standards Institution (BSI), in order to issue technical standards for the purpose of future regulation, as indicated in policy papers[61] and the UK's National AI Strategy.[62]

According to the IAPP's Global AI Legislation Tracker,[63] no Commonwealth small country has started to develop AI legislation, despite preliminary discussions.[64] However, some have enacted technical standards supporting legislation on privacy and similar rights, such as eID, that could overcome some of the challenges brought by GAI, such as unlawful use of personal information.

The Jamaican Data Protection Act of 2020 (JDPA)[65] provides obligations that depend on technical knowledge, *inter alia*, the need for the controller to adopt, and require from processors, appropriate technical security measures (Section 30(4)(a)). The JDPA also modulates the right to portability according to technical feasibility (Section 6(2)(C)(II)). Additionally, Jamaica enacted its National Identification and Registration Act, 2021 (NIRA) regulating the use of technical standards for storage, management, security and confidentiality[66] as stated also in its policy.[67]

Data protection legislation from Bahamas[68] and Barbados[69] does not explicitly mention reliance on technical standards. However, there is an obligation to adopt appropriate

59 GOV.UK (2018) *Data Protection Act*. Available at: https://www.gov.uk/data-protection (Accessed: 22 February 2024).

60 Legislation.gov.uk. (2011) *Consumer Protection Act 1987*. Available at: https://www.legislation.gov.uk/ukpga/1987/43/part/I. (Accessed: 22 February 2024).

61 GOV.UK. (2023) *A pro-innovation approach to AI regulation*. Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#fnref:68 (Accessed: 22 February 2024).

62 HM Government (2021) *National AI Strategy*. Available at: https://assets.publishing.service.gov.uk/media/614db4d1e90e077a2cbdf3c4/National_AI_Strategy_-_PDF_version.pdf (Accessed: 22 February 2024).

63 IAPP Research and Insights (2023) *Global AI Legislation Tracker*. Available at: https://iapp.org/media/pdf/resource_center/global_ai_legislation_tracker.pdf (Accessed: 22 February 2024).

64 UNESCO (n.d.) *UNESCO Caribbean Artificial Intelligence Initiative*. Available at: https://en.unesco.org/caribbean-artificial-intelligence-initiative (Accessed: 22 February 2024).

65 Jamaica (2020) *Data Protection* Act. Available at: https://japarliament.gov.jm/attachments/article/339/The%20Data%20Protection%20Act,%202020.pdf (Accessed: 20 February 2024).

66 Jamaica (2021) *The National Identification and Registration* Act. Available at: https://www.egovja.com/wp-content/uploads/2023/10/The-National-Identification-and-Registration-Act-2021.pdf (Accessed: 22 February 2024).

67 Jamaica (2016) *White Paper on National Identification System Policy*. Available at: https://opm.gov.jm/wp-content/uploads/2017/02/NIDS-Policy-October2016.pdf (Accessed: 22 February 2024).

68 Bahamas (2008) *Data Protection*. Available at: https://laws.bahamas.gov.bs/cms/images/LEGISLATION/PRINCIPAL/2003/2003-0003/2003-0003.pdf?zoom_highlight=Police+act (Accessed: 22 February 2024).

69 The Barbados Parliament (2019) *Data Protection Bill*. Available at: https://www.barbadosparliament.com/bills/details/396 (Accessed: 22 February 2024).

security measures, which could imply awareness of technical standards issued or recognised by the relevant authority.

Similarly, in Trinidad and Tobago, it is possible to infer the connection with technical standards with its Data Protection Act calling for safeguards and standards that bind the commissioner.[70]

Certain Caribbean countries – including some Commonwealth small countries – are engaging in AI projects with the help of international entities, such as UNESCO[71] and the Caribbean Telecommunications Union (CTU)/International Telecommunications Union (ITU),[72] following in the footsteps of the HIPCAR project.[73] Considering Europe's and the USA´s drive to regulate AI, Caribbean countries need to follow the pace established by developed countries, in order to avoid the legal, technological and commercial gaps widening.[74]

Small countries need to devise their strategies, taking into account the fact that their environment may not be as resourceful as that of larger countries.

They may profit from using their existing cybercrime laws (and from observing international experience of cybercrime) as a starting point, and then using standards to fill in any blanks.

For instance, it is typical of cybercrime laws (including the Commonwealt's Model Law,[75] the Budapest Convention[76] and cybercrime law of small countries such as Barbados[77])

70 Parliament of the Republic of Trinidad and Tobago (2011) *Data Protection Act*. Available at: https://www.ttparliament.org/publication/the-data-protection-act-2011/ (Accessed: 20 February 2024)

71 UNESCO (n.d.) UNESCO Caribbean Artificial Intelligence Initiative. Available at: https://en.unesco.org/caribbean-artificial-intelligence-initiative (Accessed: 20 February 2024).

72 See https://ctu.int/event/ict-week/ (Accessed: 22 February 2024).

73 'HIPCAR was designed to support the Caribbean countries in improving their competitiveness by harmonizing approaches to ICT development. It brought together the Caribbean governments, regulators, service providers, civil society, private sector, regional and international organizations involved in ICT.' See ITU (n.d.) *HIPCAR Project*. Available at: https://www.itu.int/en/ITU-D/Projects/ITU-EC-ACP/HIPCAR/Pages/default.aspx (Accessed: 22 February 2024).

74 Skeete, K-A. D. (2022) *The Adoption of Artificial Intelligence within the Caribbean: Resuscitating the CARICOM's Single Market and Economy*. Cambridge University Press. Available at: https://www.cambridge.org/core/books/abs/international-perspectives-on-artificial-intelligence/adoption-of-artificial-intelligence-within-the-caribbean-resuscitating-the-caricoms-single-market-and-economy/BE948CF5D9623D001E9EF2161FA45F97 (Accessed: 22 February 2024).

75 '3. In this Act, unless the contrary intention appears: (...) "computer system" means a device or a group of inter-connected or related devices, including the internet, one or more of which, pursuant to a program, performs automatic processing of data or any other function;'

76 'For the purposes of this Convention: (...) a "computer system" means any device or a group of interconnected or related devices, one or more of which, pursuant to a program, performs automatic processing of data;'

77 '3.(1) In this Act (...) "computer system" means a device or a group of inter-connected or related devices, including the Internet, one or more of which, pursuant to a programme, facilitates communication, performs automatic processing of data or any other function;'

to include a definition of 'computer system'.[78] This assumes that outputs will be generated 'pursuant to' determined programming, which is not strictly the case for GAI, the results of which may be unpredictable (that is, not 'pursuant to' relevant programming).

Therefore, small countries should rely on standards to make it clear that GAI falls within the definition of 'computer system', to the extent that AI systems relate to computer systems in spite of – possibly 'programmed' – black boxes.

For example, the diagnostics of a committee in charge of reviewing Jamaica´s cybercrime Act were: 'These general concerns include (...) the establishment of a Cybersecurity Authority and guidelines/standards for the protection of protected computers.'[79]

Families of standards can provide a comprehensive and integrated architecture,[80] that cannot be addressed by a single Act. For small states in particular, it would be difficult to address such architecture via a complex set of laws passed by parliament.

International standards virtually ensure worldwide recognition. Hence, it is both more practical and more beneficial for small countries to match their standards and legal rules governing GAI.

Standards may reinforce combatting GAI-powered cybercrime by providing several other helpful references (well beyond the definition of 'AI system', compared to 'computer system'). For instance, 'forgery' is a classic criminal typology, later extended to include spoofing, and now also 'deepfakes'. In short, standards can evolve, tracking new phenomena, without the need to introduce further legislation.

This is also relevant to GAI-fuelled cybercrime in the fields of IP, fraud and miscellaneous others.

Given the above, standards could help to regulate highly technical matters related to GAI in order to 'futureproof' legal gaps or to fill in the blanks.

78 This could be correlated with the expression 'data processing systems' present on the USA NIST definition of AI: '(1) A branch of computer science devoted to developing data processing systems that performs functions normally associated with human intelligence, such as reasoning, learning, and self-improvement.'

79 Linton, L. *New Offences Recommended for Inclusion in Cybercrimes*. Jamaica Information Service. Available at: https://jis.gov.jm/new-offences-recommended-for-inclusion-in-cybercrimes-act/ (Accessed: 22 February 2024).

80 For instance, among ISO standards, there are several norms pertaining to IA: ISO/IEC 23053:2022, which is specific to machine learning structures and systems; ISO/IEC 22989:2022, which addresses AI terminology and concepts; ISO/IEC 23894:2023, dealing with risk management in connection with AI; and ISO/IEC 42001, contemplating AI management systems.

Conclusion

This article indicates the urgent need[81] to combine legal rules with technical standards and frameworks to respond more effectively to the fast pace of cybercrime innovation, particularly GAI-powered cybercrime.

GAI has brought an unprecedented wave of cyber threats, with immense potential reach and economic and social relevance.

The complementation of legal rules with standards and frameworks may help Commonwealth small countries tailor cybercrime legislation,[82] by technically supporting legal provisions applicable to GAI-powered crime, in conformance with local background and scenarios.[83] At the same time, such an approach can ensure consistency with basic legal and technical patterns generally adopted by Commonwealth countries or derived from international context.

Such integration between legal rules and standards shall be constantly monitored and adjusted, given the fast pace of GAI's technological evolution.[84]

81 'In order to create a control mechanism over cyber space and some form of deterrent for cyber criminals, a number of countries around the world have reformed their existing laws and legislation; however, these have proven to provide vague and inefficient solutions. (...) Given the growth of cyber activities, the absence of a coordinated, comprehensive control framework has added to the spread of cybercrime in all shapes and forms. (...) A final recommendation has to do with the law itself in terms of content coverage and enforceability. (...) A number of countries have developed cyber laws which have many advantages(...). At the same time, however, they have various shortcomings, in terms of lack of coverage of certain crimes and/or the weight and severity of the penalty associated with the crime; (...)'. Karake, Z. and Al Qasimi, S.L. (2010) *Cyber law and cyber security in developing and emerging economies*. Cheltenham, UK; Northampton, Ma: Edward Elgar, pp. 1; 213; 231).

82 'Small and developing countries face difficulties in implementing the standards of the (Budapest) Convention. (...) One of the provisions that causes difficulties when it comes to implementation in small countries is the need to establish a 24/7 point of contact. (...) not all countries which have ratified the Convention have established such a contact point even countries which have provided such a contact point often only use it for limited purposes.' International Telecommunication Union (2009) *Understanding Cybercrime: A Guide for Developing States*, p. 127. Available at: https://www.itu.int/ITU-D/cyb/cybersecurity/docs/itu-understanding-cybercrime-guide.pdf (Accessed: 22 February 2024).

83 'Given the international nature of cybercrime, the harmonisation of national laws and techniques is vital in the fight against cybercrime. However, harmonisation must take into account regional demand and capacity. The importance of regional aspects in the implementation of anti-cybercrime strategies is underlined by the fact that many legal and technical standards were agreed among industrialised countries and do not include various aspects important for developing countries. Therefore, regional factors and differences need to be included within their implementation elsewhere.' International Telecommunication Union (2009) *Understanding Cybercrime: A Guide for Developing States* pp. 15; 85. Available at: https://www.itu.int/ITU-D/cyb/cybersecurity/docs/itu-understanding-cybercrime-guide.pdf (Accessed: 22 February 2024).

84 On the need for a dynamic regulatory model, open to assimilating new developments and knowledge, see Murray, A. (2007) *The Regulation of Cyberspace: Control in the online environment*. Milton Park, Abingdon UK; New York, Ny: Routledge-Cavendish., p. 257: 'It is to be suggested that if we are to further our understanding of the regulatory environment within cyberspace (or any complex environment), we must, much like the quantum physicists of the early twentieth century, accept these limitations and use them to our advantage. For knowing what you do not know is as important as knowing what you do.'

About the authors

Gilberto Martins de Almeida teaches IT and internet law at the Pontifical Catholic University of Rio de Janeiro, and is a founder of the research institute IDTEC (Instituto Direito e Tecnologia), a partner at Martins de Almeida – Advogados, and a consultant to various divisions of the United Nations.

Fernando Felipe Bourguy de Medeiros is a partner at Martins de Almeida - Advogados, a member of the Rio de Janeiro Privacy Board, and an associate researcher at IDTEC (Instituto Direito e Tecnologia). He graduated in Law from the Federal University of Rio de Janeiro (UFRJ), and has concluded specialisation studies in law to become a magistrate.

João Farrel, deputy superintendent of the Data Privacy Committee of Rio de Janeiro's Chapter of the Brazilian Bar Association, is a researcher at IDTEC (Instituto Direito e Tecnologia), and an associate lawyer at Martins de Almeida – Advogados.

Diego Policani is a member of Technology and Society Studies Laboratory at the Federal University of Rio de Janeiro's faculty of law (LETS-FND), a researcher at IDTEC (Instituto Direito e Tecnologia), and a legal intern.

ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

The Commonwealth

Special Section on Artificial Intelligence

Violent Extremism and Artificial Intelligence: A Double-Edged Sword in the Context of ASEAN

Wan Rosalili Wan Rosli[1]

Abstract

Digital integration and the emergence of new technologies such as artificial intelligence (AI) are providing new tools for insurgents to use in spreading their propaganda through violent extremism. The Association of Southeast Asian Nations (ASEAN) has come to represent a conduit for insurgents in planning and carrying out their extreme agendas. This article provides a deeper understanding of the double-edged sword effect of AI in relation to violent extremism in the ASEAN context. It reveals that, even though AI has been very important in countering violent extremism, it has simultaneously facilitated terrorists in spreading their propaganda in more innovative and covert ways. The legal framework governing AI is still in its infancy and challenges such as the double-edged sword effect in the use of the technology require specific guidelines or legislation for use in effective governance.

Introduction

Violent extremism and radicalism are not a new phenomenon. They entail diverse beliefs without a specific definition and are not exclusive to any religion, nationality or system of belief (UNDP, 2016). Violent extremism is a broader term than terrorism but encompasses manifestations of terrorism, including ideologically motivated violence (UNODC, 2018). The US defines violent extremism as encouraging, condoning, justifying or supporting the commission of a violent act to achieve political, ideological, religious, social or economic goals (FBI, 2021). The UK Home Office (2023) defines it as vocal and active opposition to fundamental values, which include the rule of law, liberty and mutual respect towards different beliefs and faiths.

1 Wan Rosalili Wan Rosli is an Assistant Professor at the School of Law, University of Bradford, United Kingdom.
Email: w.r.wanrosli@bradford.ac.uk / rosalili2301@gmail.com

Counterterrorism started to evolve with the 9/11 War on Terror, which has had a strong focus on coercive measures, including hard military action, increasing policy powers and expanding intelligence services (CCE, 2023). However, the approach has shifted to become more non-coercive, including the formulation of strategies to prevent individuals from supporting terrorism (ibid.). The United Nations contends that the key elements to countering violent extremism (CVE) involve the use of non-coercive means to drive individuals or groups from using violence and to mitigate recruitment, support, facilitation or engagement (UNODC, n.d.). Neuman (2004) highlighted that CVE strategies involved a non-exhaustive list of activities by governmental and non-governmental entities in the fight to combat radicalisation, such as counter-messaging exercises through social media channels, community engagements, advisory council discussions, capacity-building, women and youth empowerment, and education and training of stakeholders.

In recent years, violent extremism has shifted from an approach of holding specific territories in specific jurisdictions to real-time communications on social media platforms, where those involved seek to spread their ideologies and propaganda. The internet has no border control or checks and the lure of anonymity has changed the characteristics of violent extremism (Jacobsen, 2022; Khodzhanovna, 2023). Terrorist organisations see the internet as a safe way to recruit individuals and disseminate their ideology, especially through social media platforms such as Facebook, Twitter, YouTube and Instagram (Broeders et al., 2023).

Meanwhile, emerging technology such as artificial intelligence (AI) has contributed to the evolution of violent extremism. This article analyses the role of emerging technologies such as AI in facilitating the commission of acts of terror.

What is violent extremism?

Violent extremism has been a major issue in countries' policies and development programmes in the past few decades. The term was coined to shift the focus away from an over-militarised approach after 9/11 and to enable a more moderate approach in countering and preventing extremism (Saraiva and Erfe, 2023). The United Nations Plan of Action to Prevent Violent Extremism 2015 aims to develop resilience in sections of communities that are prone to violent extremism (UNOCT, 2015).

Although the concept is recognised across international communities, a uniform definition has never been agreed upon. As a result, the concept is easily manipulated, which poses a critical challenge to authorities and can sometimes also lead to the over-securitisation of specific sectors to further legitimise the war against terror (Stephens et al., 2021). The United Nations Special Rapporteur on the protection of human rights concluded that 'the lack of semantic and conceptual clarity that surrounds violent extremism remains an obstacle to any in-depth examination of the impact of strategies and policies to counter violent extremism on human rights as well as on their effectiveness in reducing the threat of terrorism' (Emmerson, 2016: para. 55).

Violent extremism includes elements of radicalisation, which is a process of embracing religious, political and social ideation that causes violent acts between members or groups (Doosje and van Eerten, 2017; Borum, 2023). Alcalá et al. (2017) contend that the promotion and adoption of extremist beliefs to advance violence leads to violent radicalisation, which has a critical effect on religion and society. Recent research has also highlighted that there are many aspects to extremist behaviour, emerging from cultural, educational and psychological factors. Interestingly, Stankov et al. (2018) contend that the ideology of extremism and radicalisation depends on mindsets, and extremist behaviours can be found in all humans. This contention supports the concept of extremism immunity, which entails embedding ideas, feelings and behaviours against radicalisation and extremism across all sectors of communities and social categories (ibid.).

Hamin et al. (2021) highlight that violent extremism is multidimensional and very complex, as there is no one agreed definition and different commentators often use the concept interchangeably with terrorism and radicalism. However, despite the lack of a specific meaning, the concept suggests a willingness of individuals or groups to use or support violence.

The proliferation of the internet in the past few decades has also changed how violent extremism is committed. Policy-makers and major stakeholders are strongly aware of the impact of the internet in supporting terrorism and violent extremism. Scrivens et al. (2020) highlight that law enforcement agencies have focused on learning how propaganda and extremist ideas are disseminated and cross into the real world; at the same time, major social media companies are concerned about how their platforms are seen as an important radicalising agent and become the main conduit in promoting real-world violent extremism. Commentators have also described how violent extremists have adopted new digital paradigms in their modus operandi to spread their hateful ideologies and propaganda, recruit new members worldwide and receive funding and tactical support (Salleh et al., 2016; Pressman and Ivan, 2019; Lakomy, 2023). The use of such technologies in violent extremism has been seen in Southeast Asia as well of other parts of the world.

The emergence of artificial intelligence

In the past decade, technology has invaded our everyday lives and dependency has soared. Challenges related to operations and capacity within law enforcement and counterterrorism agencies mean the use of AI has been seen as a holy grail in combatting violent extremism. AI's capacity to process vast amounts of data faster and with greater ease, and to correlate such data and discover patterns and themes, means intelligence agencies see it as an appealing commodity to confront the problem of managing information overload (Bazarkina, 2023). AI can support CVE through automating repetitive tasks, which in turn reduces workloads; predicting future violent extremist incidence; identifying suspicious transactions to detect terrorism financing; monitoring and moderating content within cyberspace; and other automation of capacities (Gutiérrez-Castillo, 2022). This serves as a game-changer, as all this surpasses human capabilities

Emerging technologies such as AI have been utilised in all sectors, from manufacturing to health to defence. Within the criminal justice context, AI has facilitated investigations through facial recognition; by assisting judges in granting bail and giving out sentences, parole and probation; and in matching DNA to perpetrators (Bazarkina, 2023). Machine learning is used to predict future criminal behaviour and identify patterns and risks of recidivism. AI is also fundamental in the prevention of cybercrime and has proven effective in preventing cyberattacks such as phishing, hacking and terrorism (Garcia, 2019).

Within the context of security, AI has played a crucial part in the fight against terrorism. AI has been used mainly for content moderation since terrorists have taken their operations to the internet (Gunton, 2022). AI and machine learning are believed to have the capacity to reduce terrorist content online and provide a safe place for users to operate within the cyberspace realm (Bamsey and Montasari, 2023). In the Southeast Asian context, such digital transformation requires adaptation, and the increased digitalisation rate exposes the countries to risks, given the established presence of violent extremist groups in the region (Ilyas, 2022).

Generative artificial intelligence

Generative AI is also an issue: these technologies can generate fictional faces and deepfakes. Deepfakes were invented in 2017, as a type of synthetic media that cannot be distinguished from authentic content (Gunton, 2022). Deepfakes are a powerful weapon that violent extremists can utilise, especially in information warfare, when people can no longer rely on what they see and hear online and offline (ibid.). Deepfakes have been used as a tool to commit malicious and criminal activities, usually politically motivated, such as destroying the credibility and reputation of a known individual, harassment, humiliation, extortion and blackmail. This can lead to social unrest and political instability (Bamsey and Montasari, 2023).

Natural language processing

Natural language processing is a deep learning application that analyses a huge amount of natural human language data, reading and defining the meaning in human languages. The technology involves speech recognition and natural language understanding, generation and translation (Bamsey and Montasari, 2023).

Combatting violent extremism via artificial intelligence

Nations all around the world see AI as a solution to prevent and counter violent extremism. The Council of Europe has adopted the Convention on the Prevention of Terrorism and compiled a Database on Cyberterrorism to mitigate cyberterrorist attacks (Ige et al., 2022). In the UK, steps have been taken to highlight the use of AI in moderating content and providing AI solutions in combatting terrorism online (McKendrick, 2019).

Other countries' governments are also confident that the use of AI is the ultimate response to violent extremism.

As well as countries putting in place technology-led solutions to combat extremist content on the internet, major content providers and service providers are creating partnerships to counter violent extremism online (McKendrick, 2019). The Global Internet Forum to Counter Terrorism is an industry-led initiative led by major online platforms such as Microsoft, X (formerly known as Twitter), Facebook and YouTube to combat violent extremism content (Fernandez and Alani, 2021). The aim is to disrupt terrorist activities online and develop tools and capacity to mitigate the impacts of terrorism. AI has also been at the forefront in combatting terrorism: states around the world are using AI in fighting extremist content online (Tech Against Terrorism, 2023). The mechanics of AI, which allow it to analyse and process enormous amounts of data, facilitate law enforcers to track and identify extremist content and activities. Content moderation via AI enables the monitoring and removal of extremist content and the breaking of networks, and the protection of vulnerable individuals who are prone to extremism.

Violent extremism in the Association of Southeast Asian Nations

In the ASEAN context, violent extremism has a long history, dating back to 1948, when the government had to counter new extremist elements during the challenging anti-colonial and post-independence era (El-Muhammady, 2023). The presence of Jemaah Islamiyah, affiliated with Al Qaeda, took root in 1940, and intensified after the 9/11 attacks in New York and Washington, DC. In the 1980s, Indonesia and Malaysia became a platform for violent extremists to participate in multinational jihad against the West. This led Jemaah Islamiyah to set up base in Indonesia, which then became a transnational Southeast Asian terror network committed to a pan-Southeast Asian Islamic State. Such ideologies spread across the whole region, from southern Thailand across Malaysia into Singapore and Indonesia to the east, and the southern Philippines. This move also led to the Bali bombing of 2002, with more than 200 fatalities. In August 2014, the Malaysian authorities managed to foil planned attacks within the country's capital (Hamzani, 2020).

The Global Terrorism Index, published in 2022, indicates that, of Southeast Asian countries, the Philippines and Myanmar are ranked within the top 20 countries that have been severely impacted by terrorism. Certain online incidents have also led the Indonesian government under the Ministry of Communications and Information Technology to set up a specialised team of moderators to moderate terrorist content in the country (Wilujeng and Risman, 2020). The ASEAN countries (Brunei Darussalam, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, Cambodia and Vietnam) are aware of the growing complexity of violent extremism, especially with the advancement of social media and the internet, resulting in new patterns of radicalisation (ibid.). The major concerns of ASEAN include lone wolf terrorists, regional groups that

pledge allegiance to ISIS, Daesh and other terrorist organisations and the return of foreign terrorist fighters to the region (Tay, 2023).

Governing violent extremism in ASEAN

In 2016, the United Nations General Assembly adopted Resolution A/RES/70/291 by consensus to reinforce efforts to fight terrorism and violent extremism. The General Assembly also recommended member states set up regional and national plans of action to be applied within the local context. To align its efforts with those of the rest of the world, ASEAN has put in place and implemented strategies and plans for governance in combatting violent extremism under the ASEAN Comprehensive Plan of Action on Countering Terrorism (Gunaratna, 2018). In 2017, ASEAN adopted the Manila Declaration to Counter the Rise of Radicalisation and Violent Extremism, which includes pledges to implement sustainable and proactive capacity-building, information-sharing between member states, mutual legal assistance on criminal matters and extradition, and the strengthening of mechanisms to address terrorism and violent extremism through collaboration and exchange of experiences with all major stakeholders (Gunaratna, 2018; Habulan et al., 2018).

The ASEAN region has used the soft law approach, given the emphasis on non-interference between the states within the zone (Tan and Nasu, 2016). The ASEAN Convention on Counterterrorism is a framework to govern and co-ordinate member states to adopt a regional treaty on counterterrorism (Shah et al., 2022). Despite the existence of common initiatives throughout the region, Southeast Asian governments have not dealt with terrorism the same way. Indonesia and Singapore adopt a more non-militaristic approach, whereas Malaysia and Thailand rely on more coercive methods.

Malaysia has in place the Prevention of Crime Act 1959, the Prevention of Terrorism Act 2015 and the Special Measures against Terrorism in Foreign Countries Act 2015 to confront the threat of violent extremism by monitoring the activities of foreign terrorist fighters (Hamin et al., 2021). Prior to this, Malaysia had one of the most unpopular pieces of legislation, in the form of the Internal Security Act 1950, which caused citizens to take to the streets to claim that the law violated basic human rights. It was later replaced by the Security Offences (Special Measures) Act (SOSMA) 2012, said to be very similar to the law it replaced (ibid.). SOSMA is a preventive law containing special measures to deal with security-related offences that include terrorism, sabotage and espionage (Dhanapal and Sabaruddin, 2017). The Malaysian Bar Council highlights that the laws in place invoked a low standard of proof and ignored basic safeguards against human rights, leading to numerous civil liberty infringements (ibid.). Despite these controversies, however, the government has set up a Southeast Asian Regional Centre for Counterterrorism, in charge of training, information-sharing and awareness programmes (Hamin et al., 2021).

Similar to the situation in Malaysia, in March 2023 Myanmar published its Anti-Terrorism Bill with the aim of replacing the problematic Prevention of Terrorism Act, which had allegedly led to extensive torture and arbitrary detentions since 1979. The proposed

Bill gives the police, the president and the military more broad powers to detain without evidence, prosecute against vaguely defined criminal offences and ban gatherings and organisations. It also does not fulfil the requirements of the United Nations Special Rapporteur, including the need for an appropriate definition of terrorism, the prevention of arbitrary detention, the prohibition of any type of torture, and guaranteed fair trial and due process (Simpson and Farrelly, 2023).

In the Philippines, terrorism is governed under the Human Security Act of 2007. This has a general and broad definition of terrorism that involves elements of fear and panic among the population that coerce the government to give in to unlawful demands (Rasul, 2023). Thailand has its own National Action Plan on terrorism, finalised in 2022, and has criminalised terrorism within its Internal Security Act 2008 (Rasul, 2023). According to Tan and Nasu (2016), Cambodia has a different approach to counterterrorism, given grave concerns about transnational crimes: intel had revealed that Jemaah Islamiyah leaders were freely travelling through the country. This led Cambodia to enact the Law on Counter Terrorism 2007 and the Law on Anti-Money Laundering and Combatting the Financing of Terrorism 2007, to address counterterrorism financing, which was rampant at the time (ibid.). This move started a regional move to address terrorism financing as part of anti-money laundering policies. Malaysia enacted the Anti-Money Laundering, Anti-Terrorism Financing and Proceeds of Unlawful Activities Act 2001 and Myanmar the Control of Money Laundering Law 2002 (Tan and Nasu, 2016; Ramakrishna, 2017); both laws were later amended to include provisions on terrorist financing.

After the Bali bombing, Indonesia enacted the Anti-Terrorism Law 2003, which gives a broad definition of terrorism and allows a suspect to be detained without trial for a period of up to six months (Ramakrishna, 2017). Indonesia, with the help of the USA and Australia, also created Densus 88 in 2003, which serves as a counterterrorism unit to deal with intelligence and operations to dismantle violent extremist networks (Ramakrishna, 2017; Rasul, 2023). Between 2021 and 2023, there were more than 610 people arrested; 42 per cent of them identified as Jemaah Islamiyah members. This shows that the group is still very active in conducting terrorist activities, especially recruitment, fundraising and regeneration, which have now gone online (Subandi et al., 2023).

The double-edged sword of artificial intelligence in violent extremism

Communication interception has been crucial in preventing and countering violent extremism. However, the evolution and increased availability of communication technologies have resulted in these technologies reaching everyone, including sovereign nations, corporations, individuals... and terrorists. It is now easier for terrorists to evade detection, and accessing critical data to predict terrorist attacks has become increasingly challenging for intelligence agencies. Terrorists are also now implementing more advanced operational security measures in order to evade intelligence collection

operations (Bazarkina, 2023). In essence, terrorists have been early adopters of new technologies that have yet to be effectively governed and regulated (Lakomy, 2023). It has been reported that sections of the ISIS terrorist group within ASEAN have also utilised unmanned aerial vehicles or drones to conduct surveillance and reconnaissance (Liang, 2023).

Terrorists have always found ways to adapt and operate in the shadows to ensure non-detection by enforcement officers. The move from real-life terror operations to the online environment is to be expected, given the borderless and anonymous nature of the internet (Brundage et al., 2018). Governments and companies aim to halt the spread of radicalising content by investing in the creation of technologies to counter and identify extremism through AI solutions (ibid.). The main objectives are to understand the phenomena behind online extremism; to detect extreme users and content in cyberspace; and then to predict the spread of extremist ideologies within the online sphere (Bazarkina, 2023).

Amid these herculean efforts to design and develop effective AI solutions to automatically identify and block radical accounts, extremist organisations are also working hard to adapt their behaviour to avoid being detected. Through technological adaption, such organisations can make use of the latest developments in order to increase their reach undetected (Brundage et al., 2018) and can modify the terms and content they post online to avoid being detected by AI technology (UNICRI and UNCCT, 2021). The ASEAN region has always been a destination for terrorist organisations in procuring funding and recruiting new members; Malaysia and Singapore have highlighted that recruitment and funding are being carried out online and that the use of AI makes it challenging to detect such activities (Tay, 2023).

The use of generative AI has also been seen in violent extremists' exploitation of emerging technology. Media spawning involves the use of a single image or video from which generative AI can then generate thousands of manipulated images or videos capable of circumventing automated detection mechanisms utilised by law enforcers. Fully synthetic propaganda generates artificial content, including speeches, images and other propaganda. Personalised propaganda uses tools to customise messaging and media to a targeted audience with specific demographics. Such technology analyses each demographic and in turn produces personalised propaganda to suit the beliefs and understandings of the audience (Tech Against Terrorism, 2023).

States around the world are using AI to fight extremist content online, and it is seen as a good solution in the fight against online extremism. The mechanics of AI, which allow it to analyse and process enormous amounts of data, facilitate law enforcers to track and identify extremist content and activities. Through content moderation via AI, extremist content is removed and networks are broken (Tech Against Terrorism, 2023). However, any propaganda published is translated into multiple languages via natural language processing software in order to overwhelm moderation. Ultimately, extremist groups have found ways to avoid detection in spreading radicalised content online by subverting

moderation, using AI tools to design multiple variants of propaganda specifically engineered to bypass available techniques put in place by law enforcement (Tech Against Terrorism, 2023).

These new emerging technologies have proven that the current model for detection is obsolete, and the use of generative AI will provide opportunities to stay ahead of the threats (Tech Against Terrorism, 2023). Using AI to circumvent safeguards built into the infrastructure amplifies the distribution and dissemination of terror propaganda (Sabbagh, 2023).

The ASEAN response

The current cyber-climate is challenging, with attacks and threats becoming more sophisticated and volatile. The ASEAN regional response to cyberthreats has always been to fortify high levels of co-operation among member states in the form of computer emergency response teams, which focus on capacity-building and information-sharing in cyber-emergencies. However, the region's response to cyberterrorism is fragmented, owing to the lack of a strategic approach towards cybersecurity. In the second quarter of 2023, ASEAN members agreed to develop an ASEAN Guide on AI Governance and Ethics by 2024, following the focus of other nations in AI governance. However, it must be noted that individual ASEAN member states are very slow to progress on having their own AI governance frameworks.

The emergence of generative AI has changed the landscape of cyberterrorism, and ASEAN countries must be prepared. Tay (2023) highlights that ASEAN's current response to cyberterrorism lacks a regional institutional structure, and the cyber-operation architecture has no clear political authority and is a confusing maze, with various sectoral platforms. The nature of ASEAN itself hinders adequate governance of such crimes; unlike the EU, as a supernational entity, the ASEAN structure is based on intergovernmental organisations and principles of sovereignty, founded in consensus decision-making and non-interference between member states. Apart from this, the uneven development of legal and technological responses to such crimes across ASEAN is seen as a limitation.

Tay (2023) has also contended that the deficiency of the common cyber-lexicon is also a challenge: different states have different measures in defining the impact of a cyber-emergency or an attack on critical national infrastructure. Malaysia, for example defines a Level 5 crisis as having a critical impact on critical infrastructure organisations; however, other countries, such as Cambodia and Indonesia, do not have similar responses to Malaysia in defining a crisis.

Conclusion

There has been unprecedented progress in the use of AI in countering cyberterrorism to predict and detect terrorist attacks. However, significant challenges also arise with the deployment of such technologies. ASEAN should put in place a standard

and agreed terminology to ensure effective communication and information-sharing during a cyber-emergency such as a cyberterrorism attack. The absence of such an agreement will have impacts on the counter-effort to eliminate terrorist threats within the region. Member states within ASEAN should also ensure that the datasets used in countering terrorist attacks are verified and free from the risks of hallucinations and data poisoning. Such risks are aggravated given that the majority of states within ASEAN do not yet have a framework in place to govern AI and are still in the planning stages on addressing the issue. Meanwhile, despite ASEAN's ongoing participation in the effort to ensure global governance in AI, national regulation and regional co-operation are still lacking.

Technology will always evolve and the double-edged impact of AI in cyberterrorism means states must be prepared to face the unpredictable risks associated with it. The constant evolution of extremist behaviours and the numerous ways of avoiding detection in ever-changing narratives call for a serious response within the context of ASEAN. The fight to diffuse the impacts of negative uses of AI technology will continue. Enforcement is the key. Having a clear legal framework to combat terrorism and AI governance is of the utmost importance. Apart from this, digital citizens must be made aware of the subversive nature of the internet and social media, and national action plans and strategies must be put in place to prevent violent extremism.

References

Alcalá, H.E., M.Z. Sharif and G. Samari (2017) 'Social Determinants of Health, Violent Radicalization, and Terrorism: A Public Health Perspective'. *Health Equity* 1(1): 87–95.

Bamsey, O. and R. Montasari (2023) 'The Role of the Internet in Radicalisation to Violent Extremism'. In Montasari, R. (ed.) *Digital Transformation in Policing: The Promise, Perils and Solutions*. Cham: Springer International Publishing, pp. 119–135.

Bazarkina, D. (2023) 'Current and Future Threats of the Malicious Use of Artificial Intelligence by Terrorists: Psychological Aspects'. In Pashentsev, E. (ed.) *The Palgrave Handbook of Malicious Use of AI and Psychological Security*. Cham: Springer International Publishing, pp. 251–272.

Borum, R. (2023) 'Mapping the Terrain: The Current State of Risk and Threat Assessment Practice in the Violent Extremism Field'. In Logan, C., R. Borum and P. Gill (eds) *Violent Extremism: A Handbook of Risk Assessment and Management*. London: UCL Press, pp. 53–78.

Broeders, D., F. Cristiano and D. Weggemans (2023) 'Too Close for Comfort: Cyber Terrorism and Information Security across National Policies and International Diplomacy'. *Studies in Conflict & Terrorism* 46(12): 2426–2453.

Brundage, M., S. Avin, J. Clark et al. (2018) 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation'. *arXiv preprint arXiv:1802.07228*.

Commission for Countering Extremism (CCE) (2023) 'Commission for Countering Extremism End of Year Report 2022 to 2023' www.gov.uk/government/publications/commission-for-countering-extremism-end-of-year-report-2022-to-2023/commission-for-countering-extremism-end-of-year-report-2022-to-2023-accessible-version

Dhanapal, S. and J.S. Sabaruddin (2017) 'Prevention of Terrorism: An Initial Exploration of Malaysia's POTA 2015'. *Pertanika Journal of Social Sciences & Humanities* 25(2): 783–804.

Doosje, B. and J.J van Eerten. (2017) '"Counter-Narratives" against Violent Extremism'. In Coleart, L. (ed.) *De-radicalisation*. Brussels: Flemish Peace Institute, pp. 83–100.

El-Muhammady, A. (2023) 'A "Blue Ocean" for Marginalised Radical Voices: Cyberspace, Social Media and Extremist Discourse in Malaysia'. In Loh, B.Y.H. (ed.) *New Media in the Margins: Lived Realities and Experiences from the Malaysian Peripheries*. Singapore: Springer Nature Singapore, pp. 163–192.

Emmerson, B. (2016) 'Report of the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms while Countering Terrorism'. Human Rights Council Report A/HRC/31/65.

Federal Bureau of Investigation (FBI) (2021). *Strategic Intelligence Assessment and Data on Domestic Terrorism*. Available at: https://www.fbi.gov/file-repository/fbi-dhs-domestic-terrorism-strategic-report.pdf/view

Fernandez, M. and H. Alani (2021) 'Artificial Intelligence and Online Extremism: Challenges and Opportunities'. In McDaniel, J. and K. Pease (eds) *Predictive Policing and Artificial Intelligence*. Abingdon: Routledge, pp. 132–162.

Garcia, E.V. (2019) 'The Militarization of Artificial Intelligence: A Wake-Up Call for the Global South'. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3452323

Gunaratna, R. (2018) 'ASEAN's Greatest Counter-Terrorism Challenge: The Shift from "Need to Know" to Smart to Share'. In KAS and RSIS (eds) *Combatting Violent Extremism and Terrorism in Asia and Europe: From Cooperation to Collaboration*, pp. 111–128.

Gunton, K. (2022) 'The Use of Artificial Intelligence in Content Moderation in Countering Violent Extremism on Social Media Platforms'. In Montasari, R. (ed.) *Artificial Intelligence and National Security*. Cham: Springer International Publishing, pp. 69–79.

Gutiérrez-Castillo, V.L. (2022) 'Big Data and the New Armed Conflicts'. In Fernández-Sánchez, P.A. (eds) *The Limitations of the Law of Armed Conflicts: New Means and Methods of Warfare*. Brill Nijhoff, pp. 284–297.

Home Office (UK) (2013) *HM Government Counter-Terrorism Disruptive Powers report 2022 (accessible version)*. [online]. https://www.gov.uk/government/publications/counter-terrorism-disruptive-powers-report-2022/hm-government-counter-terrorism-disruptive-powers-report-2022-accessible-version

Habulan, A., M. Taufiqurrohman, M.H.B. Jani et al. (2018) 'Southeast Asia: Philippines, Indonesia, Malaysia, Myanmar, Thailand, Singapore, Online Extremism'. *Counter Terrorist Trends and Analyses* 10(1): 7–30.

Hamin, Z., S. Kamaruddin, A.R. Abd Rani and A. Munirah (2021) 'When Violent Extremism Is No Longer a Man's World: Some Evidence from Malaysia'. *International Journal of Academic Research in Business and Social Sciences* 11(9).

Hamzani, A.I. (2020) 'The Trend to Counter Terrorism in ASEAN'. *Journal of Advanced Research in Dynamical and Control Systems* 12(7): 105–113.

Ige, T., A. Kolade and O. Kolade (2022) 'Enhancing Border Security and Countering Terrorism Through Computer Vision: A Field of Artificial Intelligence'. *Proceedings of the Computational Methods in Systems and Software*. Cham: Springer International Publishing, pp. 656–666.

Ilyas, M. (2022) 'Terrorism Industry and Data Coloniality in Southeast Asia'. *Journal of Contemporary Governance and Public Policy* 3(1): 31–46.

Jacobsen, J.T. (2022) 'Cyberterrorism'. *Perspectives on Terrorism* 16(5): 62–72.

Khodzhanovna, S.K. (2023) 'A New Interpretation of Cyberterrorism: Challenges and Prospects'. *Best Journal of Innovation in Science, Research and Development* 2(7): 91–96.

Lakomy, M. (2023) 'Why Do Online Countering Violent Extremism Strategies Not Work? The Case of Digital Jihad'. *Terrorism and Political Violence* 35(6): 1261–1298.

Liang, S.C.. (2023) 'Terrorist digitalis: preventing terrorists from using emerging technologies'. In *Global Terrorism Index 2023*. Institute for Economics and Peace. pp. 72–74. www.visionofhumanity.org/wp-content/uploads/2023/03/GTI-2023-web-170423.pdf

McKendrick, K. (2019) *Artificial Intelligence Prediction and Counterterrorism*. London: The Royal Institute of International Affairs-Chatham House.

Neuman, G.L. (2004) 'Comment, Counter-Terrorist Operations and the Rule of Law'. *European Journal of International Law* 15(5): 1019–1029.

Pressman, D.E. and C. Ivan (2019) 'Internet Use and Violent Extremism: A Cyber-VERA Risk Assessment Protocol'. In IRMA (ed.) *Violent Extremism: Breakthroughs in Research and Practice.* Hershey, PA: IGI Global, pp. 43–61.

Ramakrishna, K. (2017) 'The Growth of ISIS Extremism in Southeast Asia: Its Ideological and Cognitive Features—and Possible Policy Responses. *New England Journal of Public Policy* 29(1): 1–35.

Rasul, A. (2023) 'ASEAN and the Challenge of Democracy'. In Guo, Y. and I. Puja (eds) *Sustaining Peace in ASEAN and the Asia-Pacific: Preventive Diplomacy Measures*. ASEAN-IPR and CFAU, pp. 177–194.

Sabbagh, D. (2023) 'Terrorists Could Try to Exploit Artificial Intelligence, MI5 and FBI Chiefs Warn'. *The Guardian*, 18 October. www.theguardian.com/technology/2023/oct/18/terrorists-exploit-artificial-intelligence-ai-mi5-fbi-chiefs-warn

Salleh, N.M., S.R. Selamat and Z. Saaya (2016) 'A New Taxonomy of Cyber Violent Extremism (Cyber-VE) Attack'. 6th International Conference on Information and Communication Technology for the Muslim World.

Saraiva, R. and A. Erfe (2023) 'Preventing Violent Extremism with Resilience, Adaptive Peacebuilding, and Community-Embedded Approaches'. *Current Opinion in Environmental Sustainability* 61: 101271.

Scrivens, R., P. Gill and M. Conway (2020) 'The Role of the Internet in Facilitating Violent Extremism and Terrorism: Suggestions for Progressing Research'. In Holt, T. and A. Bossler (eds) *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. London: Palgrave, pp. 1417–1435.

Shah, H.A.R., K. Zada, N.M. Ali and M.M. Sahid (2022) 'Peace in ASEAN: Counter-Narrative Strategies against the Ideologies of Radicalism and Extremism (Goal 16)'. In Khalid, R.M. and A.J. Maidin (eds) *Good Governance and the Sustainable Development Goals in Southeast Asia*. Abingdon: Routledge, pp. 194–211.

Simpson, A. and N. Farrelly (2023) *Myanmar: Politics, Economy and Society.* Abingdon: Routledge.

Stankov, L., G. Knežević, G. Saucier et al. (2018) 'Militant Extremist Mindset and the Assessment of Radicalization in the General Population'. *Journal of Individual Differences* 39(2): 88–98.

Stephens, W., S. Sieckelinck and H. Boutellier (2021) 'Preventing Violent Extremism: A Review of the Literature;. *Studies in Conflict & Terrorism* 44(4): 346–361.

Subandi, Y., H.R.T. Sjahputra and M. Subhan (2023) 'Indonesia-ASEAN Partnership to Counter-Terrorism in Indonesia'. *East Asian Journal of Multidisciplinary Research* 2(7): 2857–2874.

Tan, S.S. and H. Nasu (2016) 'ASEAN and the Development of Counter-Terrorism Law and Policy in Southeast Asia'. *The University of New South Wales Law Journal* 39(3): 1219–1238.

Tay, K. (2023) *ASEAN Cyber-Security Cooperation: Towards a Regional Emergency Response Framework*. London: IISS.

Tech Against Terrorism (2023) 'Early Terrorist Experimentation with Generative Artificial Intelligence Services'. Briefing, November.

United Nations Development Programme (UNDC) (2016) *Preventing Violent Extremism Through Promoting Inclusive Development, Tolerance and Respect For Diversity: A Development Response To Addressing Radicalization And Violent Extremism*. https://www.undp.org/sites/g/files/zskgke326/files/publications/Discussion%20Paper%20-%20Preventing%20Violent%20Extremism%20by%20Promoting%20Inclusive%20%20Development.pdf.

United Nations Counter-Terrorism Centre (UNCCT) and the United Nations Interregional Crime and Justice Research Institute (UNICRI) (2021) *Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes*. United Nations Office of Counter-Terrorism. https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/malicious-use-of-ai-uncct-unicri-report-hd.pdf

United Nations Office of Counter-Terrorism (UNOCT) (2015) 'Plan of Action to Prevent Violent Extremism'. www.un.org/counterterrorism/plan-of-action-to-prevent-violent-extremism

United Nations Office on Drugs and Crime (UNODC) (no date) 'Terrorism Prevention'. Regional Office for Southeast Asia and the Pacific. www.unodc.org/roseap/en/what-we-do/terrorism-prevention/index.html.

UNODC (2018) 'E4J University Module Series: Counter-Terrorism: Module 2: Conditions Conducive to the Spread of Terrorism: *"Radicalization" & "Violent Extremism"*'. https://www.unodc.org/e4j/zh/terrorism/module-2/key-issues/radicalization-violent-extremism.html.

Wilujeng, N.F. and H. Risman (2020) 'Examining ASEAN: Our Eyes Dealing with Regional Context in Counter Terrorism, Radicalism, and Violent Extremism'. *International Journal of Social Sciences* 6(1): 267–281.

About the author

Dr Wan Rosalili Wan Rosli is an Assistant Professor at the University of Bradford, United Kingdom. She has secured multiple research grants in subject areas ranging from money laundering to cybercrime, artificial intelligence, prevention/countering of violent extremism-related laws, and cybersecurity issues. She also has more than 30 academic papers published in journals. She was invited to be a subject matter expert in formulating a new anti-stalking law for Malaysia. She developed an application named MyStalk Alert which aims to support the victims of harassment and stalking by assisting them to keep a trail of evidence which is imperative for investigation and prosecution. The app also gives mental health support and useful links that help victims to keep a diary of all incidences. The MyStalk Alert won one gold medal and one silver medal in an international innovation competition. Dr Wan Rosalili has also received several recognitions from her former university, where she won the research and publication award between 2020 and 2022.

The Commonwealth

Commonwealth Cyber Journal \ 59
ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

Special Section on Artificial Intelligence

AI Systems and the Future of Intellectual Property Regimes

Teresia Munywoki[1]

Abstract

The rapid progress of artificial intelligence (AI) technology is transforming the world of intellectual property rights (IPR), offering both unprecedented opportunities and novel challenges. This article explores how AI-driven innovation affects the evolving IPR framework. It emphasises the need for collaboration among lawmakers, courts, regulators and intellectual property (IP) professionals. Protecting IPR in the dynamic AI landscape is crucial as technology advances. this article delves into the legal and practical aspects, to enable an understanding of how AI and IP intersect.

The article deals with key aspects relating to authorship,[2] risk and complexity,[3] balancing innovation and societal access,[4] global competition[5] and AI supremacy.[6]

Keywords: Artificial intelligence (AI), Intellectual property (IP), Copyright, Patents, Algorithm, Invent.

Introduction

The dawn of the artificial intelligence (AI) era has ushered in a new wave of innovation, propelling society into uncharted territories of automation and machine learning and

1 Advocate of the High Court of Kenya, and partner, B M Musau & Company, Advocates LLP, Nairobi, Kenya; certified professional mediator, digital governance lawyer. LinkedIn: https://www.linkedin.com/in/teresia-munywoki-1b6666105/

2 Current legal frameworks often require human authorship, involving both human creators and AI systems, making it necessary to revisit existing definitions of authorship within copyright laws.

3 IPR must consider the risk and complexity associated with different AI applications, as AI-driven innovation may lead to new forms of IP infringement.

4 Striking a balance that fosters innovation while preserving broader societal access to knowledge and creativity is essential.

5 The global landscape of AI development is characterised by a race for AI supremacy, with countries investing heavily in research and development. The USA arguably stands as the leader in AI development, with major tech corporations at the forefront of innovation.

6 The emergence of powerful AI models like ChatGPT has fuelled competition, particularly between the USA and China. The USA is set to unveil LaMDA, a language model designed to provide answers to user queries, positioning it as a primary hub for AI development and securing a pivotal role in shaping the global AI landscape.

unprecedented computational capabilities. As AI becomes an integral part of our daily lives, in roles ranging from virtual assistant to autonomous vehicle, the implications for intellectual property rights (IPR) have become increasingly profound.

Historically, IPR have been designed to protect the creations of the human intellect, providing inventors, creators and innovators with the incentives and rights necessary to foster innovation and creativity. However, the advent of AI challenges conventional notions of authorship, ownership and originality. Can an algorithm be considered an inventor? Who holds the rights to the outputs AI systems generate? These questions underscore the pressing need to re-evaluate and adapt existing intellectual property (IP) frameworks to accommodate the distinctive characteristics of AI-generated innovations.

This article navigates the intricacies of AI and IP, examining key areas of contention such as patent law, copyright and trade secrets. Additionally, it scrutinises the ethical dimensions surrounding AI-generated content and the potential impact on fair competition. By addressing these issues, it aims to contribute to the ongoing discourse on shaping a future-proof IP landscape that both encourages innovation and safeguards the rights and interests of all stakeholders.

Methodology

A systematic literature review was conducted using scholarly resources such as Google Scholar, Lexis Nexis, Westlaw and other reputable databases. The search utilised keywords including 'Artificial Intelligence' and 'Intellectual Property' to identify relevant articles published within the past five years. This approach ensured the inclusion of recent scholarly contributions and up-to-date insights into the rapidly evolving field of AI and IP law. Additionally, Google searches were employed to gather current affairs information, enabling the incorporation of real-world examples and case studies to enrich the analysis. Qualitative and quantitative analyses were used to assess the identified literature and data. Qualitative methods involved in-depth examination and critical evaluation of the identified articles and legal documents to extract key themes, trends and arguments pertaining to AI and IP regimes. Quantitative methods entailed statistical analysis of relevant data points, such as the frequency of certain legal concepts or trends in patent filings related to AI technologies.

Artificial intelligence and intellectual property rights

In recent decades, the rise of AI has been nothing short of transformative, ushering in an era marked by unprecedented advancements in technology. From machine learning algorithms to autonomous systems, AI has permeated various facets of our daily lives and industries, demonstrating its potential to reshape the way we work, communicate and innovate.

The intersection of AI with IPR stands as a pivotal juncture in the evolution of both technological and legal landscapes (Davies, 2011). As AI technologies become integral

to innovation, the traditional paradigms of IP face novel challenges. Examining this intersection is crucial not only for safeguarding the rights of inventors and creators but also for fostering an environment that encourages further AI development. It prompts us to reconsider established notions of authorship, ownership and innovation in the context of machine-generated creations.

IP plays a pivotal role in incentivising innovation by providing creators and inventors with exclusive rights to their creations. The prospect of obtaining patents, copyrights or other protections encourages investment in research and development, fostering a competitive environment that drives technological progress (Davies, 2011). In the context of AI, striking a balance between protecting the interests of innovators and ensuring the accessibility of knowledge is crucial for sustaining a vibrant ecosystem of innovation. Understanding the nuances of IPR is fundamental to navigating the evolving landscape shaped by the integration of AI technologies.

The ascent of AI heralds a new era of innovation, challenging the foundations of IPR that trace their roots to statutes often outdated in comparison with the rapid evolution of AI technologies. The historical context of IP laws, based in statutes such as Singapore's Patents Act (based on the UK Patents Act 1977) and Copyright Act (originally adopted from the Australian Copyright Act 1968), exposes their inherent anthropocentric orientation. Developed without foresight into the advent of AI, these laws are now grappling with the question of how to accommodate the revolutionary emergence of AI and, potentially, artificial general intelligence (AGI), as explained by Jaketch (2023). The fundamental principles underpinning IP, such as the conditions for patent grants and the types of works eligible for copyright, were crafted without the notion of non-human creators in mind.

The anthropocentric reasoning behind IP laws becomes evident when considering the three commonly accepted justifications for IP protection: the labour theory, the extension-of-personality theory and the provision of economic incentives. These justifications, grounded in human-centric philosophies, raise questions about their applicability to AI entities (Kang'Ethe, 2023). While humans are motivated by economic rewards, express individuality and develop as persons through their creations, the same cannot be presumed for AI, which operates mechanically without intrinsic motivations or expressions of personhood.

The potential dissonance between AI and IP laws becomes pronounced when examining the limitations of current IP frameworks. AI entities, lacking human-like personhood and motivations, challenge the conventional justifications for IP protection. Current laws typically deny AI entities the legal recognition as authors or inventors, reflecting the difficulty in attributing human-centric motivations to these entities. The absence of emotional and social intelligence in AI further complicates the application of moral justifications rooted in human personhood (Davies, 2011).

The rise of AI prompts policy-makers, academics and legal experts to grapple with fundamental questions regarding the regulation of AI and its creations. Should

longstanding anthropocentric concepts of IP law be imposed on AI? Are patent and copyright protections applicable to creations by AI entities? These questions necessitate careful consideration, as they encompass not only legal intricacies but also ethical concerns surrounding the treatment of non-human entities within established legal frameworks.

Ownership and attribution in AI-created content

The advent of AI has disrupted the traditional landscape of copyright law, challenging established principles that revolve around human authorship and creativity. Traditionally, copyright laws were crafted to protect the fruits of intellectual labour grounded in the creative powers of the human mind. The notion of authorship was closely tied to human creativity, and copyright qualified for protection if a work was fixed and original.

The rise of AI, particularly advanced forms capable of decision-making without human intervention, challenges these traditional assumptions of copyright law. AI-generated works, ranging from music and journalism to gaming content, disrupt the conventional understanding of authorship and creativity (Lichtenauer and Turner, 2023). As AI systems evolve to become creative forces in their own right, the traditional paradigm where copyright is vested in a human author is being redefined. The famous case of the monkey selfie, where a court ruled against copyright for a non-human entity, sets the stage for grappling with similar questions concerning AI-generated works. The latest types of AI represent a departure from their historical role as mere tools. They have emerged as dynamic creative forces capable of independently making decisions in the creative process, challenging the traditional ownership structures embedded in copyright law, which inherently assume human authorship (ibid.).

The use of AI in creative processes has introduced new complexities to copyright and IP laws, prompting legal debates and challenges. Recently, there have been legal actions, such as the lawsuit filed by authors, including Sarah Silverman, against OpenAI and Meta for alleged copyright infringement (Kang'Ethe, 2023). The central question revolves around the ownership and authorship of content generated by AI tools, especially those trained on potentially copyrighted material without the original creators' consent.

AI systems, often comprising intricate software, algorithms and data, challenge conventional notions of authorship. The essence of copyright, rooted in the personality and creative input of an author, encounters complexities when applied to the collaborative and multifaceted nature of AI-generated creations (Kaindo, 2022). Unlike the unmistakable imprint of a human author evident in works like John Grisham's novels, AI-produced works raise the question: Do they reflect anyone's personality? (Jaketch, 2023).

The intricate process of creating AI-generated works involves the development of software, the crafting of algorithms and the training of these algorithms with extensive datasets. The collaborative effort includes programmers, data suppliers and users

activating the AI system, making it arduous to identify a singular author. The common proposition that the programmer should hold copyright encounters scepticism, as the software's functioning involves various contributors and the unpredictable 'black box' phenomenon (Korinek et al., 2021). The concept of Explainable AI (XAI) emerges as a potential solution, aiming to demystify the decision-making process and challenge the notion that programmers are the sole authors.

The ambiguous nature of copyright laws in the context of AI-generated content is a key concern. The legal framework, especially in the USA, relies on precedents and interpretations that may not have anticipated the challenges posed by AI. The language used in copyright laws, such as pronouns referring to the inventor or creator, assumes a human-centric perspective and raises questions about how these laws apply to non-human entities like AI systems (Korinek et al., 2021).

Ownership dilemmas further complicate the picture. Arguments proposing data owners as copyright holders face challenges related to originality, suggesting that licensing agreements may better govern data use. The notion of granting corporate personhood to AI-generated works, as akin to companies, introduces practical hurdles during licensing agreements (Korinek et al., 2021). Defining AI works as 'works made for hire' or 'derivative works' proves challenging, given that AI systems fundamentally 'learn' and generate something entirely new, not fitting the traditional scope of derivative works.

The question of who should be incentivised, a cornerstone of copyright law, takes a new turn with AI-generated works. Unlike human creators, AI systems do not respond to incentives in the traditional sense. There is often a proposed *a sui generis* system for protecting AI-generated works, acknowledging the rapid pace of technological evolution. A suggested duration of five to seven years aims to allow stakeholders to recoup investments before the works enter the public domain (Kaindo, 2022).

The legal dilemma at the heart of this evolution revolves around whether copyright law should extend protection to works generated solely by a computer. In Australia and Europe, legal frameworks are grappling with this question, particularly in the context of originality, which historically reflects the author's personality. On the other hand, certain jurisdictions, including Kenya, have taken an alternative approach, attributing authorship to the programmer (Gakuru, 2023).

This creates a legal conundrum, raising questions about the true author of AI-generated works. Under the Kenyan Copyright Act, authorship in computer-generated works is ascribed to the person who made the arrangements necessary for their creation (Jaketch, 2023). Ambiguities arise, however, regarding the identification of this person, leaving uncertainty about whether it is the programmer or the user who should be considered the true author.

The Act further suggests that, even a user's contribution as minimal as pressing a button for the machine to create the work qualifies this user as an author and owner of copyright (O'Brien, 2023). This interpretation introduces complexities surrounding the level of

creative input required for copyright ownership. Without robust copyright protection for AI-generated works, there exists a risk that these creations may be deemed free of copyright (Kaindo, 2022). This scenario could have far-reaching consequences for the creative economy, as AI-generated works may be freely used and reused without due compensation or acknowledgement. Striking a balance between fostering innovation and safeguarding the economic interests of creators is a paramount concern.

The lack of consensus among World Intellectual Property Organization (WIPO) member states on the ownership issue in AI-generated works underscores the urgent need for international legal clarity. The current legal framework must be adapted to recognise the unique challenges AI poses, ensuring fair and equitable treatment for all stakeholders involved in the creation and use of AI-generated works.

The challenges to copyrighting AI-generated content stem from uncertainties about human involvement, the intentions behind AI-generated output and the training data used (O'Brien, 2023). The scraping of data from the web, often without consent, complicates matters, as the training data may include copyrighted works. Determining whether an AI-generated piece is transformative enough to qualify as fair use under copyright law adds another layer of complexity (Kang'Ethe, 2023).

The legal landscape may need to evolve to address these challenges, potentially considering joint ownership models or extending copyright protection to styles and forms generated by AI (Say, 2020).

As the use of generative AI becomes more prevalent, the need for clear and adapted copyright laws becomes increasingly apparent to balance the protection of creators' rights with the advancement of technology. In navigating the intricate challenges of AI-generated works within the realm of copyright, the legal landscape must evolve to accommodate the distinctive nature of AI's creative capabilities (Smith, 2023). Achieving global consensus on the treatment of AI within copyright laws is crucial for fostering innovation while upholding the principles of IP. The ongoing discourse on AI and copyright should pave the way for a legal framework that aligns with technological advancements, supports the creative economy and promotes the public interest in the digital age (O'Brien, 2023).

Protecting creative works developed through AI systems requires a delicate balance between acknowledging the collaborative nature of AI creations, addressing ownership complexities and adapting copyright laws to the dynamic landscape of emerging technologies (Shemtov, 2019). The ongoing discourse calls for a nuanced legal framework that fosters innovation, protects investments and ensures responsible use of AI-generated works in the public interest.

AI and patent regimes

The recent granting of patent rights to an AI system in South Africa marks a significant development in the intersection of AI and IP law. The invention, a 'food container based on fractal geometry,' listed an AI system named DABUS as the inventor, with

the machine's owner, Stephen Thaler, recognised as the patent owner (Craig, 2021). This decision by South Africa to acknowledge an AI system as an inventor challenges traditional notion of inventorship.

However, the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO) rejected the application on the grounds that their patent laws explicitly require an inventor to be a natural person. The rejection was based on the linguistic use of pronouns like 'himself' and 'herself' in the patent laws, implying a requirement for human inventorship. The USPTO also argued that AI did not meet the threshold for 'conception,' a term traditionally associated with a mental act in the mind of the inventor (Sharma, 2023). Notably, South Africa's patent system differs in that it does not conduct formal examinations and relies on a check-box evaluation process. The country's Patent Act allows for patent applications from 'the inventor or any other person who has acquired the right from the inventor or both.' Interestingly, the use of the pronoun 'him' in the South African Patent Act raises questions about the legislative intent regarding AI inventorship (Shemtov, 2019).

This discussion raises complex questions within the legal framework of Kenya. The Industrial Property Act 2001, which governs patents in the country, is built on the premise that inventions are the result of human engagement processes. However, as technology evolves, the traditional interpretation of inventorship is being challenged (Kang'Ethe, 2023). In Kenya, the exclusive right to a patent belongs to the inventor, who is assumed to be a natural person. The Act requires the disclosure of the inventor's name; exceptions apply only in cases of legal representation or when the inventor chooses not to be named. The legislation does not anticipate non-human inventors, and the language used, such as pronouns like 'he' and 'himself,' implies a focus on natural persons (Jaketch, 2023).

The application process in Kenya emphasises the disclosure of details about the applicant rather than the inventor. This could suggest a degree of flexibility regarding who or what can be considered an inventor. However, on the question of patentability, particularly the requirement for an inventive step, there is ambiguity (Drexl et al., 2021). The determination of inventiveness involves a subjective evaluation by examiners, and questions arise about whether inventions by AI machines would be deemed prior art and whether AI machines can be considered 'persons skilled in the art' (ibid.). The Act does not explicitly address the possibility of non-human inventors, and its reliance on terms like 'person' and 'he' in reference to the inventor reflects a traditional understanding. The legal concept of succession, which involves the transfer of patent rights after a natural person's demise, reinforces the assumption that inventors are human beings.

The discussion around DABUS and AI-generated inventions highlights the need for a nuanced examination of IP laws. As technological advancements challenge established norms, there is a growing need for legal frameworks to adapt to new realities (Craig, 2021). The determination of whether an AI machine can be recognised as an inventor depends on interpretations of existing laws, discussions on the legal personality of machines and evolving perspectives on innovation in the digital age.

Balancing innovation and IP regulation

The intersection of AI and IP introduces a complex interplay between fostering innovation and safeguarding the public domain. As AI-generated creative works and inventions become increasingly prevalent, the traditional concepts within IPR face challenges that necessitate a delicate balance between incentivising creators and ensuring broader access to knowledge and creativity. The rise of generative AI algorithms, capable of autonomously creating content across various domains, prompts a re-evaluation of copyright law. The fundamental question of ownership and protection of AI-generated works challenges the existing framework, which traditionally attributes copyright to human creators. Debates over whether copyright should belong to the input provider, the AI developer or the AI itself underscore the need for a nuanced understanding of authorship (Abbott and Rothman, 2023). The originality of AI-generated works, lacking direct human input, adds another layer of complexity to copyright considerations. Striking a balance within copyright law requires addressing these questions to avoid potential stifling of creativity while acknowledging the transformative role of AI.

In the realm of patent law, AI's impact is twofold (Davies, 2011). On the one hand, AI facilitates the development of new inventions, streamlining the ideation and drafting process. On the other hand, concerns arise regarding the potential flood of patents and the challenge of determining inventorship in cases where AI systems contribute significantly. Recent cases, such as Stephen Thaler's patent application denial, highlight the current limitations of existing frameworks that insist on human inventors (ibid.). The global divergence in how jurisdictions approach AI-generated inventions introduces economic and geopolitical considerations, emphasising the need for a harmonised and adaptive international approach.

IP infringement by AI introduces another layer of complexity, particularly in cases where AI mimics the likeness of individuals or scrapes copyrighted content for repurposing. For such, no protection should be accorded, as this is an infringement on the rights of the person. IPR should not be accorded to works that already infringe the rights of others and, in the same manner, AI mimicking content and/or repurposing it is outright plagiarism. Legal disputes, such as Getty Images' infringement claims against Stability AI, underscore the need for clear guidelines and legal frameworks. As courts grapple with these novel challenges, the development of a robust legal and regulatory framework becomes imperative to balance the protection of existing IPR and the encouragement of AI-driven technological advancements (Craig, 2021). Amid these challenges, AI also presents opportunities for IP rights holders. AI-powered tools can automate the detection of potential infringements, streamlining the enforcement process. Additionally, AI's role in IP management and strategy, analysing vast amounts of data to identify trends and assess competitive landscapes, offers a strategic advantage to businesses (Bosher et al., 2020).

Responding to these questions and challenges requires collaborative efforts among legislators, courts, regulators and IP professionals. Updating existing IP laws, developing

new legal frameworks and establishing an environment conducive to responsible AI use in IP management are essential steps (Ahuja, 2020). This collaborative approach ensures IPR remain resilient in the face of rapid technological change, striking a balance that fosters innovation while preserving the broader societal access to knowledge and creativity.

To accommodate AI-driven innovation, policy-makers should consider several key aspects. First, a nuanced understanding of authorship needs to be developed. While current legal frameworks often require human authorship, the unique contributions of AI systems to the creative process cannot be ignored (Ahuja, 2020). An adaptive approach that recognises collaborative authorship, involving both human creators and AI systems, could be explored. This would entail revisiting existing definitions of authorship within copyright laws to encompass the evolving nature of creative collaboration.

Additionally, IPR must consider the risk and complexity associated with different AI applications. A tiered or risk-based model for regulation could be implemented, wherein high-risk applications, such as AI in critical decision-making or creative content generation, may be subject to more stringent regulations (Amatika-Omondi, 2019). This approach aims to foster innovation in low-risk areas while ensuring high-risk applications adhere to ethical and legal standards.

Moreover, fostering collaboration between stakeholders is crucial. Policy-makers, technology companies, legal experts and representatives from diverse sectors should engage in dialogue to formulate regulations that strike the right balance (Olsen, 2023). This collaborative approach ensures regulations are not only effective but also practical, considering the dynamic nature of AI technologies.

Balancing rewards for creators and facilitating broader access to knowledge involves reassessing the duration and scope of IP protection. Incentives for innovation should be preserved, but mechanisms such as shorter protection periods or alternative models like open-source licensing can be explored to encourage the dissemination of knowledge (Cho, 2023). This approach aligns with the original purpose of IP: to promote innovation for the benefit of society. Furthermore, transparency in AI-generated works is paramount. Clear documentation of the roles humans and AI play in the creative process ensures IPR are attributed appropriately. This documentation becomes crucial when seeking copyright or patent protection, providing clarity on the origin of ideas and innovation (Amatika-Omondi, 2019).

As the legal landscape adapts to AI-driven innovation, there is a need for ongoing evaluation and revision of existing laws. Policy-makers should remain proactive, considering the evolving nature of technology and its impact on creativity. Recent inquiries by entities like the US Copyright Office indicate a recognition of the need for legislative or regulatory steps in response to AI advancements (Cho, 2023). Achieving a harmonious balance within IPR amid AI-driven innovation requires a thoughtful, collaborative and adaptive approach. Recognising the unique contributions of both

humans and AI, implementing risk-based models and fostering transparency are key elements in shaping regulations that reward creators while ensuring broader societal access to knowledge and creativity (O'Brien, 2023).

The global landscape of AI development

The race for AI supremacy is a defining feature of the technological landscape, with countries worldwide investing heavily in research and development. The emergence of powerful AI models like ChatGPT has fuelled competition, prompting major players, particularly the USA and China, to showcase their advancements. The USA stands as the unequivocal leader in AI development, housing major tech corporations like Google, Facebook and Microsoft at the forefront of innovation (Lee, 2020). The country's commitment to AI is evident through substantial investments and continuous research and development efforts. Google is set to unveil its response to OpenAI's ChatGPT with LaMDA, a language model designed to provide detailed answers to user queries. This technological prowess positions the USA as a primary hub for AI development, securing a pivotal role in shaping the global AI landscape.

China, while currently in second place, is rapidly advancing in the AI arms race. The Chinese government has allocated significant funds to AI research and development, aiming to challenge US dominance (Cho, 2023). Corporations like Alibaba, Baidu and Tencent are actively involved in ground-breaking AI projects, contributing to the country's ascent in AI capabilities. Despite the fierce competition, China's progress suggests it could emerge as a serious contender soon, setting the stage for a dynamic shift in the global AI hierarchy.

While the USA and China dominate headlines, other nations are making significant strides in AI technology. Canada, with a robust AI strategy backed by substantial funding, Japan's 'Society 5.0' plan and Korea's commitment to becoming an AI powerhouse showcase the diverse efforts worldwide (Lee, 2020). These countries are investing in research, development and talent to bolster their national competitiveness in AI. Although currently trailing behind the top contenders, their strategic initiatives position them as potential disruptors in the evolving AI landscape (Abbott and Rothman, 2023).

Europe, as a continent, is making strides in AI development, with individual countries like France and Germany leading the charge. While the region lags the USA and China, substantial investments and initiatives, such as the 'AI for Europe' initiative, demonstrate a commitment to catching up. The EU's collaborative approach aims to foster a competitive environment, ensuring Europe remains a strong contender in the global AI arms race (Abbott and Rothman, 2023).

As AI technologies continue to advance, it is crucial for nations to collaborate, ensuring the benefits are shared globally and ethical considerations are paramount. The future of AI development will likely witness further innovations, collaborations and strategic shifts, ultimately shaping a collective future where AI benefits everyone.

IPR and AI crossroads

Getty Images and Stability AI

The recent developments involving Getty Images and Stability AI reflect the intensifying legal battles at the intersection of AI and IP. Getty Images has filed a lawsuit in the USA against Stability AI, the creator of the open-source AI art generator Stable Diffusion. Getty Images accuses Stability AI of 'brazen infringement' on a massive scale, alleging that the startup copied over 12 million images from Getty's database without permission or compensation to build a competing business. The lawsuit, which follows Getty's legal proceedings in the High Court of Justice in London, is part of a broader legal struggle between AI startups and rights holders (Craig, 2021).

Getty's case appears to be on a stronger footing than a previous artist-led lawsuit against Stability AI. Legal experts suggest Getty Images' complaint is technically more accurate and focuses on the unauthorised use of its images, contrasting with the class action lawsuit's emphasis on occupational harm to artists caused by AI tools (Eapen et al., 2023). The lawsuit delves into copyright infringement arguments, emphasising the violation of Getty's copyright and trademark protections. The case hinges on the interpretation of the US fair use doctrine and the concept of 'transformative use,' as AI art tools often scrape images from the web without creators' consent.

The legal battle underscores the challenges posed by AI-generated content to traditional IP frameworks. Getty Images' decision to also venture into AI image creation with Generative AI by Getty Images signals a dual strategy: legal action against infringing AI tools and the development of a commercially viable AI service that respects creator rights. The new service, built in collaboration with Nvidia, aims to offer a commercially viable option for businesses while addressing IP concerns. Getty Images emphasises that its Generative AI model avoids using stolen imagery from the open internet and promises full indemnification for commercial use.

These developments highlight the evolving landscape where legal, technological and economic considerations intersect. The ongoing legal struggles underscore the need for clear guidelines and legal frameworks to address the challenges and opportunities presented by AI-driven innovation in the realm of IP. As stakeholders navigate this complex terrain, the outcomes of these legal battles will likely set important precedents for the future of AI and IP.

USA declares AI art to be unprotected copyright

The rise of generative AI technologies, such as Gan's and Google's DeepDream, has had significant impacts on the world of art creation since 2014 (Edwards, 2023). These AI tools have accelerated the creative process and empowered artists to produce a wide range of stunning artworks, from simple portraits to intricate imaginary worlds. However,

this transformative technology has sparked a debate over whether AI-generated art can be eligible for copyright protection.

In the USA, a federal judge recently ruled that pieces of art created by AI were ineligible for copyright protection owing to the absence of human authorship. The judge emphasised that copyright law had never been extended to protect works generated by technology without human guidance, stating that copyright protection applied only to works of human creation (Weinbaum and Veitas, 2018). The CEO of Imagination Engines, Stephen Thaler, is leading an effort to challenge this ruling. Thaler aims to copyright AI-generated works and has contested the US Copyright Office's rejection of his attempt to copyright an artwork titled 'A Recent Entrance to Paradise' in 2022 (Smith, 2023). He argues that the 'human authorship' requirement is unconstitutional and believes that AI should be recognised as a legitimate creator.

In Kenya, copyright laws currently lack specific provisions addressing whether machines can receive authorship or recognition. The Copyright Act in Kenya emphasises that copyright protection is granted to works that involve sufficient effort to give them original character (Amatika-Omondi, 2019). However, there is no clear definition of what constitutes 'effort' under the law. Authorship is traditionally attributed to a human mind but the evolving landscape of AI-generated art challenges these established notions.

The Copyright Act in Kenya defines artwork as an original work of visual art created by an artist or produced under their authority. The law suggests copyrightable work requires sufficient effort for original character. The challenge arises when considering scenarios where a person inputs data into a machine, and the machine, in response to the command, creates a piece of art (Gakuru, 2023). The law did not anticipate such developments, making it difficult to determine whether there is sufficient human effort expended to qualify auto-generated art for copyright protection.

To address these challenges, Kenya is taking steps to regulate the use of AI through initiatives like the AI task force and the Kenya Copyright Board (KECOBO). KECOBO's platform for the conversation about AI and copyright reflects ongoing efforts to adapt legal frameworks and protect the rights of creatives in the evolving landscape of AI-generated content.

KECOBO'S forensic audit

The recent decision by the High Court in Kenya to allow KECOBO to act on a forensic audit of three Collective Management Organisations (CMOs) holds significant implications for IP oversight, particularly in the context of AI. The Court's dismissal of a petition against the forensic audit reinforces the regulatory authority's right to ensure compliance within organisations responsible for managing copyrights. This decision underscores the need for transparent mechanisms in the context of the growing role of AI in content creation, distribution and rights management.

The audit, prompted by suspicions of fraudulent transactions, reflects a proactive approach to addressing issues within CMOs and signals a broader trend of adapting legal frameworks to the challenges posed by AI-generated content (Jaketch, 2023). As AI continues to shape the landscape of content creation and distribution, regulatory bodies may need to adapt and enhance their oversight mechanisms to ensure that AI-generated works comply with copyright laws and ethical standards (Amatika-Omondi, 2019).

The involvement of law enforcement agencies and the Court's affirmation of the audit's legality in accordance with the Constitution and Copyright Act highlight the evolving landscape of IP governance in the digital era. As AI becomes more integrated into creative processes, lawmakers may need to revisit and update IP laws to address the distinctive characteristics and challenges associated with AI-generated works. This development underscores the importance of maintaining transparency and accountability within organisations responsible for managing copyrights. As technology continues to advance, legal frameworks and regulatory practices must adapt to ensure the responsible and ethical use of AI in the creative industries.

Taylor Swift's explicit images

A few weeks ago, the world woke up to hundreds of explicit AI-generated images depicting pop idol Taylor Swift. The images, originally posted by the website *Celeb Jihad*, had been shared by a user on the social site X (formerly Twitter), where they were reposted thousands of times and spread like wildfire. Despite the posts getting taken down by X for violating the site's community guidelines, the damage had already been done, as the images had been shared elsewhere on the information superhighway (Weatherbed, 2024).

Whereas this was shocking and disheartening to Swift, it should not come as a surprise. Before her incident, numerous popular personalities had reported incidences of explicit AI-generated images of themselves being shared online, deeply affecting their image and esteem. The profound development of AI technology has made it incredibly easy to generate deepfakes of anyone by feeding these technologies with the right prompts.

When AI produces content that closely resembles an individual, it can mislead and deceive audiences into believing the content originated from the person it mimics, thereby infringing on their identity and reputation. Granting IPR to content like this means individuals lose control over their likeness, potentially leading to exploitation, misrepresentation or even harm. Allowing AI-generated content to freely mimic individuals without consequences undermines the principles of creativity, ownership and integrity. This article does not advocate for absolute protection of all AI-generated imagery and or videos, as this would include the above-explained conundrum.

Recommendations

AI is reshaping creativity and innovation. It has reached a point where collaboration between countries is fast becoming necessary to protect IPR in AI-generated content. Given the unique challenges posed by AI-generated content, such as questions regarding authorship, ownership and infringement, countries can work together within existing frameworks like WIPO to develop guidelines and agreements tailored to these complexities.

International law can play a crucial role in assisting countries in assessing and agreeing on standards for IP designations concerning AI-generated content. International law principles, such as the principles of non-discrimination and equitable treatment, provide a framework for countries to negotiate and implement agreements that uphold the rights of creators, users and consumers of AI-generated content. Additionally, international law mechanisms, including dispute resolution mechanisms and the recognition of foreign judgements, offer avenues for countries to resolve conflicts and address discrepancies in the interpretation and application of IP standards.

Furthermore, industries should engage in proactive measures to address the challenges of protecting IP in AI-generated content. This involves promoting transparency and ethical practices in AI development, such as ensuring clear attribution of authorship and ownership of AI-generated works. Industry-led initiatives should establish voluntary guidelines and best practices for IP management in AI, fostering a culture of responsible innovation and equitable access to AI-generated content. Ultimately, fostering dialogue and knowledge-sharing among stakeholders, including researchers, technologists, legal experts and content creators, can help develop consensus on emerging IP issues in AI.

Conclusion

The question of IPR surrounding AI-generated content is a pressing one. This article asserts that the prompter owns the rights to AI-generated content. Just as an artist's brush is unable to paint on a canvas without the artist's hand moving it around on the blank canvas, an AI algorithm is unable to generate content on its own without the user's either sophisticated or mere three-word input. Conclusively, an algorithm cannot be an inventor.

Moving forward, achieving a harmonious balance within IPR amid AI-driven innovation demands a collaborative approach. Continual evaluation and revision of existing laws are essential to ensure fairness and accountability in the distribution of rights and recognition. Ultimately, navigating this complex terrain requires a blend of foresight, co-operation and adaptability to foster a legal landscape that supports innovation while safeguarding the interests of all stakeholders involved.

References

Abbott, R. and E. Rothman (2023) 'Disrupting Creativity: Copyright Law in the Age of Generative Artificial Intelligence'. In Abbott, R. *(ed.) (2022 forthcoming) Research Handbook on Intellectual Property and Artificial Intelligence*. Cheltenham: Edward Elgar.

Ahuja, V.K. (2020) 'Artificial Intelligence and Copyright: Issues and Challenges'. *ILI Law Review*, Winter.

Amatika-Omondi, F. (2023) 'Protecting Creative Works Developed through Artificial Intelligence Systems'. *Copyright in the Age of Artificial Intelligence* 38(3–4). https://copyright.go.ke/sites/default/files/newsletters/issue-38.pdf

Bosher, H., P. Westenberger, O. Gurgula and F. Wang (2020) 'WIPO Impact of Artificial Intelligence on IP'. Policy Response from Brunel Law School & Centre for Artificial Intelligence.

Cho, W. (2023) 'Sarah Silverman Hits Stumbling Block in AI Copyright Infringement Lawsuit against Meta'. *The Hollywood Reporter,* 21 November. www.hollywoodreporter.com/business/business-news/sarah-silverman-lawsuit-ai-meta-1235669403/

Craig, C.J. (2021) 'The AI-Copyright Challenge: Tech-Neutrality, Authorship, and the Public Interest'. In Abbott, R. *(ed.) (2022 forthcoming) Research Handbook on Intellectual Property and Artificial Intelligence*. Cheltenham: Edward Elgar.

Davies, C.R. , (2011) 'An evolutionary step in intellectual property rights: Artificial intelligence and intellectual property'. *Computer Law & Security Review*, 27(6), pp.601–619.

Drexl, J., R. Hilty, L. Desaunettes-Barbero et al. (2021) 'Artificial Intelligence and Intellectual Property Law'. Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate. Max Planck Institute for Innovation & Competition Research Paper 21–10.

Eapen, T.T., D.J. Finkenstadt, J. Folk and L. Venkataswamy (2023) 'How Generative AI Can Augment Human Creativity'. *Harvard Business Review* 101(4): 76–85.

Edwards, J. (2023) 'Can AI Ever Become Capable of Original Thought?' *Information Week,* 30 October. www.informationweek.com/machine-learning-ai/can-ai-ever-become-capable-of-original-thought-

Gakuru, L. (2023) 'Ai Art Ruled to Be Ineligible for Copyright in the USA'. *techweez,* 23 August. https://techweez.com/2023/08/23/ai-art-ineligible-copyright-usa/

Jaketch, W. (2023) 'Ownership Issues in Copyright Works in the Age of Artificial Intelligence'. *Copyright in the Age of Artificial Intelligence* 38(3–4). https://copyright.go.ke/sites/default/files/newsletters/issue-38.pdf

Kaindo, P. (2023) 'Legal Dilemma in Use of Artificial Intelligence in Creation of Copyright Works'. *Copyright in the Age of Artificial Intelligence* 38(3–4). https://copyright.go.ke/sites/default/files/newsletters/issue-38.pdf

Kang'Ethe, M. (2023) 'Me, Myself, and AI: Should Kenya's Patent Law Be Amended to Recognise Machine Learning Systems as Inventors?' *Strathmore Law Review* 8(1): 73–102.

Korinek, M.A., M.M. Schindler and J. Stiglitz (2021) *Technological Progress, Artificial Intelligence, and Inclusive Growth*. Working Paper 2021/166. Washington, DC: IMF.

Lee, J. (2020) 'The Future of Artificial Intelligence and Intellectual Property Rights'. *LawTech.Asia,* 21 December. https://lawtech.asia/the-future-of-artificial-intelligence-and-intellectual-property-rights/

Lichtenauer, D. and C. Turner (2023) 'Balancing Artificial Intelligence and Intellectual property: Human Authorship, a "Bedrock Requirement of Copyright"'. *Jdsupra,* 11 September. www.jdsupra.com/legalnews/balancing-artificial-intelligence-and-9397293/

O'Brien, M. (2023) 'Photo Giant Getty Took a Leading AI Image-Maker to Court. Now It's Also Embracing the Technology'. *APNews,* 25 September. https://apnews.com/article/getty-images-artificial-intelligence-ai-image-generator-stable-diffusion-a98eeaaeb2bf13c5e8874ceb6a8ce196

Olsen, A. (2023) 'Revolutionizing the Future: The Role of Intellectual Property in AI Innovation'. Schmeiser Olsen & Watts LLP, 3 August. https://iplawusa.com/revolutionizing-the-future-the-role-of-intellectual-property-in-ai-innovation/

Say, S. (2020) 'Artificial Intelligence and Copyright Law in Singapore: A Study on the Protection of Compilations and Databases Arranged by AI-Systems'. Chulalongkorn University Theses and Dissertations (Chula ETD) 215.

Sharma, S. (2023) 'Google Indemnifies Generative AI Customers over IP Rights Claims'. *InfoWorld,* 12 October. www.infoworld.com/article/3708631/google-indemnifies-generative-ai-customers-over-ip-rights-claims.html

Shemtov, N. (2019) 'A Study on Inventorship in Inventions Involving AI Activity'. Commissioned by the European Patent Office.

Smith, N. (2023) 'Embracing the AI Revolution: Navigating Intellectual Property Challenges and Opportunities'. *Jdsupra,* 7 July. www.jdsupra.com/legalnews/embracing-the-ai-revolution-navigating-8329884/

Weinbaum, D.R.W. (2018) 'Open-Ended Intelligence'. Doctoral dissertation, Vrije University Brussel.

Weatherbed, J. (2024) 'Trolls Have Flooded X with Graphic Taylor Swift AI Fakes'. *The Verge*, 25 January www.theverge.com/2024/1/25/24050334/x-twitter-taylor-swift-ai-fake-images-trending

About the author

Teresia M. Munywoki is an Advocate of the High Court of Kenya. Her legal expertise encompasses digital governance law, and she is also a certified professional mediator. She actively contributes to legal discourse and professional development as a member of the Law Society of Kenya, the East Africa Law Society and the Data Privacy and Governance Society. Teresia also serves on the EALS Corporate Law Committee.

The Commonwealth

Commonwealth Cyber Journal \ 75
ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

Special Section on Artificial Intelligence

Crimes of Influence: Generative Artificial Intelligence-led Crime as a Service

Nicole Matejic[1] and Chris Wilson[2]

Abstract

'Crimes of Influence' – crimes that seek to influence people towards harmful outcomes – will be one of the defining features of generative artificial intelligence (AI)-led cybercrime. With an ability to persuade and influence at potentially unavoidable economies of scale, crimes of influence leverage the heuristics and biases that form part of everyday human cognition in ways that mislead, deceive, impair, disrupt, degrade and/or deny user-normative decision-making. Supported by evolving Crime as a Service (CaaS) models engineered to exploit human cognition, generative AI will challenge legislators, regulators and policymakers in ways that they are currently underprepared for. With generative AI able to surpass its initial deployment configuration via adaptive learning, as well as demonstrating unintended consequences, 'who' is then responsible for the crimes it commits when the only human touchpoints occur at the design, deployment and delivery of proceeds-of-crime stages? This paper draws upon emergent scholarship to explore the present-day exploration of generative AI tools by cybercriminals and terrorists, before looking to a hypothetical future and exploring successful initiatives attempting to address this challenge.

Keywords: cybercrime, generative artificial intelligence, crimes of influence, crime as a service (CaaS), terrorism, violence, behavioural economics, influence.

Introduction

The positive benefits of generative artificial intelligence (AI) are undeniable, particularly for science, commerce, education and healthcare innovation. However, like all forms of new and emergent technology, there will always be those who seek to harness its potential for malign and criminal purposes. This paper will contend that generative AI will

1 School of Terrorism and Security Studies, Charles Sturt University (New South Wales, Australia). Email: nmatejic@csu.edu.au
2 Faculty of Arts, Politics and International Relations, University of Auckland (Auckland, New Zealand).

enable criminals and other actors with malign intent to use Crime as a Service (CaaS) products engineered to exploit cognitive biases and heuristics in ways that mislead, deceive, impair, disrupt, degrade and/or deny user decision-making. This cognitive frontier pits coercive and deceptive generative AI against citizens going about their online and virtual lives with little to no understanding of how they may be influenced or nudged towards scammers, fraudsters and violent extremists, and potentially misled into conflicts and violence. We refer to this method as 'crimes of influence'.[3] In delivering new tools at significant economies of scale, generative AI will challenge society in ways we are currently underprepared for. This is particularly true of the relationship between crime and terrorism in transnational organised settings where hybridised groups with both political and criminal ambitions remain deeply intertwined (Makarenko, 2004). While the interplay between technology, society, crime and terrorism is not a new observation, generative AI poses new questions for governments, the technology sector and civil society, particularly 'who' is potentially responsible when generative AI is used as a conduit for cybercrime. This paper will consider how present-day cybercrime is evolving to exploit the opportunities generative AI presents, and how existing transnational organised crime and terrorist organisations are adapting in parallel. The paper begins with contextual definitions before exploring the role of influence in creating permissive environments for criminals and violent extremists. The following section explores the current landscape, followed by a look at future-state crimes of influence. The remainder of the paper considers the current state of governmental responses and multistakeholder thinking against a backdrop of rapid technological advancement.

Generative AI

Public awareness of, and interest in, AI has increased significantly following the launch of Open AI's public ChatGPT chatbot in November 2023. While AI is not a new technology, ChatGPT offers a first look at a seemingly 'magical' tool (Byrne, 2023; IBM, 2023) that is likely to change the way people search, find, consume and produce information. While AI chatbots like ChatGPT, Google's Gemini (formerly Bard) and Anthropic's Claude all rely on 'purpose-built (large language) models deployed for dedicated tasks' such as telling jokes and writing human-like essays (IBM, 2023), generative AI refers to a 'category of AI

3 The legal discipline considers a wide range of influence vectors and their effects on resulting crimes. For example: 'crimes of passion' considers whether a crime was committed in response to a provocation versus a premeditated act (Cornell Law School, 2023); the abuse of power and influence for personal gain is often a feature in bribery and corruption crimes (Keeler, 2024); the influence a serious mental disorder has on a person's culpability when they committed a crime is considered (Hartvigsson, 2023); and 'undue influence' in the context of contract and criminal law considers 'a situation in which one party has exerted pressure or influence over the other party, resulting in the weaker party being induced to enter into a contract' or action that is not in their best interest (UOLLB, 2024). In this paper we use the term 'crimes of influence' more broadly, although we acknowledge culpability regarding the underlying legal aspect of 'who' committed the crime could reasonably be challenged when a generative AI CaaS has evolved beyond its initial programming.

algorithms that generates new' content such as 'images, text, audio' (World Economic Forum, 2023), animations, 3D models and many other types of data (Nvidia, 2023).

When asked 'Tell me how generative AI will change cybercrime?' ChatGPT 3.5 helpfully defined AI before listing eight impacts: (1) advanced malware and phishing attacks (2) automated vulnerability exploitation (3) evasion of security measures (4) data forgery and deepfakes (5) enhanced social engineering (6) automated password cracking (7) zero-day exploits and (8) faster and more targeted (cyber) attacks. ChatGPT concluded by noting 'generative AI is not solely a tool for cybercriminals. It can also be used for cybersecurity purposes' (ChatGPT, 2023). Google's Gemini responded to the same query with a definition of AI before listing many of the same impacts as ChatGPT while adding money laundering, traditional financial and cryptocurrency frauds, and blackmail before concluding with a helpful three-point summary of how generative AI is being used to improve cybersecurity (Google Gemini, 2023). Anthrophic's Claude responded by acknowledging its limited perspective on the query given it 'is an emerging and complex issue' before listing many of the same impacts as ChatGPT and Gemini (Anthropic, 2023). What all the chatbots sampled indicated, however, was that the cybercrimes of the future are predicated on the technology's ability to influence, deceive and coerce susceptible people. While malign influence is not a new phenomenon, as this paper will explore in future sections, generative AI represents a new frontier for those deploying influence-based techniques that seek to generate specific, harmful effects.

Generative AI CaaS: Influence for sale at scale

CaaS enables individuals and/or organised criminal groups to buy the 'tools, infrastructure, and services' they need to commit their crimes for a fee. No technical skills are required on the part of the CaaS buyer with the vendor providing readymade services tailored to a marketplace that centres around common crime types such as malware-as-a-service, exploit-as-a-service (also known as zero-day exploits) and infrastructure-as-a-service (HKCERT, 2023).[4] CaaS is a multi-billion-dollar-a-year industry. Cybersecurity intelligence company Heimdal Security predicts that, over the next five years, the economic losses associated with cybercrime will increase 'by 23% per year, reaching a total of USD 23.84 trillion annually by 2027' (Chebac, 2023). These estimates do not specifically account for any generative AI-led activities within the cybercrime ecosystem. With such huge revenues at stake, it is perhaps unsurprising that CaaS has evolved to provide enterprise-grade 'product development, technical support, distribution, quality assurance and help desk' wrap-around

4 Malware-as-a-service that delivers malicious software, ransomware, spyware and trojans to a buyer seeking to infect targeted devices to steal sensitive information or hold data hostage. Exploit-as-a-service is a more niche offering, providing access to profitable, yet unknown, security vulnerabilities. Exploits can target a range of organisations, using their ability to exploit their cyber environments to steal data, money and information, to conduct espionage, and more. Infrastructure-as-a-service delivers networked solutions, such as botnets, which can be put to use to spam targets, host malicious or illegal content, or conduct denial-of-service attacks (HKCERT, 2023).

services' (Lewis, 2018). Few have considered how CaaS models will innovate alongside generative AI. However, it is not difficult to foresee how transnational crime organisations and terrorists (or nation states) could weaponise CaaS generative AI products to persuade and exploit. This is particularly true of social engineering-based cybercrime, which is predicated on deceiving and co-opting people into becoming unwitting victims.

Why generative AI will be the most persuasive human invention yet

Contemporary influence is both technological and psychosocial in nature. AI and generative AI enables influence activities of all kinds to be conducted at a low-cost, high-yield scale while the psychosocial element relies on priming and framing information in ways that engage particular cognitive processes (such as emotion, biases and heuristics) to influence decision-making. Extensive literature from the behavioural sciences, psychology, behavioural economics and neuropsychology disciplines broadly considers how influence occurs (Cialdini, 2016 Cialdini, 2021; Thaler and Sunstein, 2009; Kahneman, 2011; Wanless, 2017; Ariely, 2008; Nickerson, 1998).

Behavioural economists consider influence as often (but not always) the product of a person's decision-making environment. Thaler and Sunstein (2009) contend that choice architecture – the way decisions are presented – are constructed to influence people's choices towards defined, predictable outcomes. Influence, they argue, occurs due to a person's susceptibility to how varying biases and heuristics are engaged (Thaler and Sunstein, 2009). Cialdini (2021) refers to these environments as 'fixed-action patterns' involving 'intricate sequences of behaviour' noting that human automaticity is a well-known principle of human behaviour that is triggered by certain stimuli.

Psychologists refer to the way the brain approaches decision-making as System 1 and System 2 thinking. 'System 1 operates automatically and quickly with little or no effort and no sense of voluntary control' while System 2 'allocates attention to effortful mental activities... The operations of System 2 are often associated with the experiences of agency, choice and concentration' (Kahneman, 2011).[5] Inducing System 1 thinking, such as is the case in scams for example, in which emails or messages look authentic enough that the brain processes the content with little if any friction, is incredibly useful to cybercriminals. In fact, if they induce System 2 thinking, which provokes a level of consideration that includes attention to detail, it is more likely that their scam will be

5 See Kahneman, D. (2011) *Thinking Fast and Thinking Slow*, which goes into detail about System 1 and System 2 thinking. For illustrative purposes, thinking that can be attributed to System 1 includes 'detecting one object is closer than another, detecting hostility in a voice, understanding simple sentences and orienting to the source of a sudden sound'. System 2 thinking is more deliberate, such as 'focusing on the voice of a particular person in a crowded and noisy room, searching memory to identify a surprising sound, telling someone your phone number, filling out a tax form, and comparing two washing machines for overall value.'

detected. This approach is predicated on hijacking heuristics – the brain's way of making mental shortcuts in decision-making – allowing for faster cognition on what Kahneman (2011) refers to as 'simple procedures that help find adequate, though often imperfect, answers to difficult questions'. Cybercriminals may also seek to hijack cognitive biases, particularly in social-engineering settings where they nudge into systemic errors in decision-making. When combined with target motivation, 'non-random errors in thinking' result in 'judgements that deviate from what would be considered desirable' against benchmarks of social norms and logic (Ariely, 2008; Nickerson, 1998). While biases and heuristics often work in tandem, this is not always the case.

Understanding the mechanics of influence is particularly pertinent to online environments, which are participatory by design (Wanless, 2017) and are often incentivised towards a choice architect's desired outcomes. Algorithmic influence, for example, contributes to the choice architecture found on social media networks. While algorithms have come under increasing scrutiny for their ability to nudge people into ecosystems that contribute to polarisation and radicalisation, there are often 'competing logics' in play. While users may value the social aspects of online community, the network's priorities differ. From 'platform growth and revenue… extending use times and attracting advertisers' (Munn, 2020) to how algorithms balance this dichotomy is of continued interest to regulators and civil society, particularly as generative AI tools are recognised as both part of the problem and near-future solution to some of the most pressing content-moderation challenges of our time (Wolbers, 2023).

To further complicate how influence occurs, the brain's neurochemistry, specifically hormones and neuropeptides, also play a part. As Zak (2017) explains, oxytocin, the hormone that builds trust between a person and a stranger, 'is evolutionary old. This means that the trust and sociality that oxytocin enables are deeply embedded into our nature' (Zak, 2017). Dopamine and cortisol have also been observed to influence human decision-making. Simi et al. (2017) noted that dopamine's impact on cognition impairs a person's ability to think critically because it activates neural pathways in the brain's reward centre, similar to the way illicit drugs provide a high and become addictive. While Harms (2017) notes that cortisol responses can be induced in people which may lead to a decrease in cognitive flexibility. The effects of this, Harms (2017) explains, is that people rely too heavily on historical information or information they have only recently been exposed to, affecting their critical and future-oriented cognition. But influence is more than heuristics, biases and how information is pushed at people. Cialdini's (2016) concept of 'pre-suasion' considers how information can be front-loaded, ahead of time, enabling a person to draw together 'seemingly insignificant cues and unimportant details' which then primes them for influence at a later date (Cialdini, 2016).

Combined, these technological, psychosocial and neurochemical attributes create a permissive environment for susceptible minds. While not everyone will be susceptible to the triggers that cybercriminals or violent extremists often present, such as phishing, social-engineering exploits and radicalising content, the way that generative

AI produces increasingly visual content leverages the way that humans preference sight over all other senses (Enoch et al., 2019), even though it is unreliable (Synnott, 2022). When it comes to decision-making, generative AI has the potential to further exacerbate the perils found within System 1 thinking due to its ability to deliver realistic information visually. The creation of false images, audio and video, such as AI-generated deepfakes, requires little expertise and few resources making this widely accessible to a range of provocateurs. The technology needed for these tasks are often open source and available publicly, enabling the quick production of misleading or deceptive content. Further, visual content such as photographs, video and memes generate a great deal more engagement than simple text. It elicits greater emotion, is disseminated further and, therefore, has a greater impact on influencing people towards outcomes such as radicalisation towards extremism. It is at this cognitive juncture that generative AI will increase the effectiveness, and potentially pervasiveness, of crimes of influence.

With technology moving towards more immersive, augmented and virtual environments, 'biometric psychographics' perhaps presents the biggest opportunity for cybercriminals and terrorists to fully exploit human cognition. 'Biometric psychographics', a term coined by Heller (2021) to capture the convergence of 'traditional biometrics and predictive behavioural analysis', is increasingly becoming a part of everyday AI-based technology. With features such as eye tracking and pupil response, facial and vocal scanning, measuring galvanic skin response, EEG (brainwaves), ECG (pulse and blood pressure) as well as EMG (muscle tension), gait, facial expressions and more, everything from wearable technology to gaming, and augmented and virtual environments, is predicated on extracting biometric psychographics to, ostensibly, deliver better user experiences (Heller, 2021; Meta, 2021). The intimate nature of such data collection should give policy and lawmakers pause. While the legitimate use of the data is clear, the lack of neutrality in online environments raises concerns about privacy, security, algorithmic misuse and manipulation. Scholars have already cautioned that immersive technology (such as virtual reality) creates cognitive challenges that result in the brain processing and remembering experiences as if they had occurred in the real word – awakening spatial memory that quite literally 'draws on the brain in permanent ink' (Heller, 2021). For victims of generative AI-led cybercrime and terrorism, this could have catastrophic effects on their mental and physical health.

That society is not yet fully immersed in augmented and virtual worlds in which generative AI's capabilities can fully exploit biometric psychographics, provides an opportunity for these risks to be better understood and mitigated. The capacity to influence others towards harmful outcomes presents risks. These are explored in the following sections.

Accountability limitations

With criminal law resting on a burden of proof based on *actus reus*, the physical act of the crime, as well as *mens rea*, the mental intent to commit the crime, generative AI has the potential to deliver both intent and means without human intervention in ways

similar to today's smart legal contracts.[6] This is because generative AI has the capability to learn from its environment and to evolve autonomously beyond its initial directions. This could make CaaS enterprises particularly troublesome to prosecute. Whether by self-evolutionary adaptation or the product of an unexpected outcome, whether CaaS providers can ever truly retain complete control over their systems requires careful consideration. For example, well-engineered CaaS enterprises may leave the deployment and delivery-of-proceeds-of-crime stages as the only human touchpoints in the process. Whether or not those who ostensibly developed, procured, deployed and/or profited from the generative AI-led CaaS could be prosecuted as a party to such an activity remains to be seen. This is particularly so in an increasingly decentralised marketplace where the use of cryptocurrency and other forms of decentralisation that obfuscate identity are commonplace. It is plausible that, alongside self-adaptive exploit-driven innovation, generative AI CaaS models may also adapt to circumnavigate anti-money laundering frameworks, further driving the proceeds of crime into decentralised financial environments. Law and policymakers will also need to consider how they treat and manage persuasive and deceptive conduct by generative AI as a standalone entity. For example, persuasive generative AI 'could become better at predicting and nudging behaviour – becoming more capable of manipulation' and deception, including against those who programmed and deployed it (Hendrycks et al., 2023).

Disruption of CaaS marketplaces may be possible, such as via interventions that result in raising the cost of conducting generative AI-led CaaS business. In an environment where those systems are in a state of perpetual adaptation, however, such interventions are unlikely to fully disrupt or degrade their capabilities for long. It is possible they will have little to not effect at all, instead resulting in raising the cost and complexity of interventions by law enforcement and regulators. This leaves the economics of scale of generative AI-based CaaS models still weighed heavily in favour of cybercriminals and creates pre-hardened CaaS-permissive environments. Early learning from adaptive AI-based malware and precision-based social engineering (HYAS, 2023) in this regard could help researchers understand these challenges more fully.

Current landscape

The potential for crimes of influence to have negative effects on users can be seen in many of the cybercrime types identified by the aforementioned chatbots as they are already affecting communities today. Industry researchers estimate that by 2025, globally, cybercrime will net over US$10.5 trillion annually making it 'more profitable than the global trade of all major illegal drugs combined' (Morgan, 2020). We contend that transnational organised cybercrime as a concept will become increasingly relevant to fully understand this

6 A smart legal contract is a 'piece of code stored on a blockchain that self-executes contract terms when certain conditions are met… following a condition-based structure' that iterates with each successive new action (Szabo, 1997).

threatscape. Similarly, how crime and terror form nexuses, and how those collaborations occur (coercion, corruption or mutually beneficial co-operation, for example) will also feature as criminals and terrorists continue to develop 'supplier and customer' relationships in traditional commerce arrangements. The interplay of cybercrime and terrorism already demonstrates a crimes of influence approach, whereby elements of coercion, manipulation and deceit are used to exploit victims and radicalise followers. This is particularly true of different extremist groups, even those with opposed ideologies, who have been observed looking to each other for complementary skillsets. Generative AI is likely to further fuel this dangerous trend towards composite and converging forms of 'salad bar extremism'.

Whether at a CaaS level or via joint ventures, such collaborations depend significantly on the opportunities and risks involved for both parties. Even then, 'such relationships tend to be strongest when criminals and terrorists share a geographic space that provides them with common criminal opportunities'. Such is the case in Afghanistan's poppy trade, Islamist militants' foreign hostage taking and selling throughout the Levant, and Colombian cocaine trafficking in the Sahel (Williams, 2018; White House, n.d.). While these types of crimes do not outwardly appear to be cyber-oriented, they are certainly cyber-enabled. From encrypted communications networks like the now defunct Anom app (DOJ, 2021) to the FBI's shutdown of online drug market 'Silk Road' (FBI, 2015), the way traditional transnational organised criminal groups have adapted to exploit cyber-enabled opportunities, has been well documented. Similarly, the use of the internet to radicalise and recruit susceptible people towards non-violent and violent extremism is also evidenced by a body of significant scholarship (Koehler, 2014; Valentini et al., 2020; Khalil, 2021). How crime-terror nexus types of collaborations manifest to leverage crimes of influence, supported by generative AI-led CaaS, is yet to be seen.

There are also other types of cyber-enabled activities that often fall short of meeting legislative thresholds for action. Such is the case with misinformation and disinformation in so far as freedom of expression and speech in some jurisdictions are, rightly, lawfully protected. In a yet more complex cyber environment, is foreign interference (Davey and Ebner, 2019; Zhang, 2022; Strick, 2023). Another consideration is how these cyber-enabled activities often combine to contribute to an information environment that has the potential to destabilise democracies. Such activities hide in plain sight among conspiracy theories and disinformation campaigns and attempt to influence populations with foreign ideals. While an online contest of ideas is a healthy feature of liberal democracies, the cyber-enabled and often crime-supported environment in which such activities occur are particularly suited to further exploitation by generative AI (Matejic, 2020).

The rapid emergence of generative AI will continue to have major implications for cybercrime and violent extremism, for example. Both have evolved substantially over the past decades in response to emergent technology such as global-positioning-system (GPS), bots, blockchain and more. Existing CaaS products offering services based on these technologies that are able to influence and manipulate opinion, with effects such as election interference and polarisation, have already been observed. For example, Russia's interference in the 2016 United States election has been well documented (US Senate

Select Committee on Intelligence, 2019), as has the use of bots alongside terrorism. After the 2015 ISIS attacks in Paris, for example, the hacker community, Anonymous, claimed to have found and removed 25,000 online bots linked to the terrorist group. The mass manipulation of public opinion through AI has the potential to generate protest, division and instability and to create social conditions within which political violence can easily spiral. Further, the criminal uses of AI, through scams and fraud, extortion, deepfakes, malware and more, provide ample opportunities for terrorists to finance their political activities.

While there is an emergent evidence base to support generative AI-led CaaS being explored by terrorists, it is important not to exaggerate the potential for bad-faith actors to deploy it. As with all criminal endeavours, capability, intent and opportunity remain essential. Similarly, because of its current-state unpredictability, generative AI could also prove to be something of a wildcard for CaaS developers and their criminal or terroristic clientele. Unintended consequences have been observed within AI foundation models by researchers who argue that these 'failure modes' are not yet broadly understood, able to be repaired nor fully explainable. While much has been written about doomsday AI scenarios, including extinction-level events (Stanford University, 2023) which go beyond the focus of this paper, unanticipated outcomes will almost certainly also occur in generative AI-led CaaS cybercrime environments.

At a Five Eyes summit in October 2023, director of the FBI Christopher Wray indicated that the FBI was already aware of terrorists seeking to use AI to assist in building bombs and hiding their activities from authorities (Sabbagh, 2023). Tech Against Terrorism (2023), in an analysis of over 5,000 pieces of AI-generated content, found within terrorist and violent extremist spaces, concerns including media spawning;[7] automated multilingual translation;[8] the generation of fully synthetic propaganda;[9] variant recycling;[10] personalised propaganda;[11] and subverting moderation.[12] While the report

7 Media spawning is the manipulation of content to evade current network detection capabilities (Tech Against Terrorism, 2023).

8 AI-based automated multilingual translation in the context of terrorism and violent extremism is deployed with the intention of overwhelming social and online network linguistic detection mechanisms, to circumvent content moderation processes (Tech Against Terrorism, 2023).

9 Generative AI-developed fully synthetic media in the context of terrorism and violent extremism includes the production of speeches, videos and interactive environments that are used to spread propaganda, and to radicalise and recruit followers (Tech Against Terrorism, 2023).

10 Variant recycling is a term used to explain the repurposing of old content and propaganda to create new versions. The primary aim in creating new versions is to defeat content moderation systems, such as hash-matching mechanisms, that prevent the upload of terrorist and violent extremist and other illegal content (Tech Against Terrorism, 2023).

11 Personalised propaganda in the context of terrorism and violent extremism is the customisation of messages to target particular people and/or demographics. Generative AI has the ability to generate propaganda to appeal to different audience segments (Tech Against Terrorism, 2023).

12 Subverting moderation is a tactic terrorists and violent extremists use to specifically engineer propaganda in ways that purposefully bypass existing content-moderation detection mechanisms. This often enables them to post illegal material online that remains accessible for longer periods due to the time it takes community reports to trigger automated and manual content-moderation processes (Tech Against Terrorism, 2023).

concludes that there is a low risk of widespread extremist adoption in the near future, this experimentation indicates an emerging risk where unintended consequential harms may still result. Present-day examples of terrorist exploitation of generative AI include extreme right-wing users creating antisemitic and racist images; a user generating and sharing instructions on how to create memes and propaganda; the production of a security guide by an Islamic State supporter; another Islamic State supporter using speech-recognition systems to transcribe and translate leadership speeches; al Qaeda supporters producing propaganda; and in the recent attack on Israel by Hamas, Izzd Ad-din Al-Qassam Brigades have used generative AI to produce a small amount of synthetic content to 'augment narrative appeal'. A significant proportion of the posts studied incited violence against Jewish, Black and other minority communities (Tech Against Terrorism, 2023). Cybercriminals and organised crime syndicates have also been observed adopting CaaS generative AI tools. FraudGPT,[13] has been deployed to deliver 'highly convincing phishing emails and deceptive websites' for US$200 a month or an annual subscription of US$1,700. Similarly, WormGPT has been 'purpose built' to 'craft compelling personalised emails'. Researchers have observed WormGPT exhibiting 'astute and persuasive' capabilities. Another new CaaS-led generative AI product is pretexting – a social-engineering tool that 'fabricates stories or pretexts to deceive users into divulging sensitive information.' Researchers have noted a significant rise in the instances of pretexting during social-engineering exploits because it is highly effective in mimicking the writing styles, languages and linguistic proficiency of the real-world organisations they pretend to be (Falade, 2023).

Based on known current-state exploration and use of generative AI by nefarious actors, allied behaviours and their likely outcomes could include the artificial representations of violent atrocities, for example, increasing the potential of such content to provoke extremists to take violent action against civilians in retaliation. Similarly, generative AI has a great deal of potential to exploit the suggestibility of people, such as isolated and disgruntled young men. This is particularly the case with younger people, a demographic of increasing importance in extremism. Teenagers are often still forming their identities, are highly impressionable, and are more likely to be impulsive, taking risks to impress. Generative AI holds substantial potential to radicalise susceptible youth. Even relatively simple chatbots can play a key role in radicalising individuals to violence. In July 2023 a young man pleaded guilty to planning to assassinate the late Queen of the United Kingdom and Commonwealth after encouragement by a chatbot which he had created on the Replika app. A psychiatrist found that the man had 'formed an emotional and sexual relationship' with the bot over months of online interaction.[14] AI chat and other tools have also been observed to provide logistical expertise for potential terrorists or members

13 FraudGPT is a subscription-based tool that, while based on ChatGPT, has been designed to deliver fake content at scale to dupe victims into believing they are dealing with a trusted institution (Falade, 2023).

14 https://www.theguardian.com/uk-news/2023/jul/06/ai-chatbot-encouraged-man-who-planned-to-kill-queen-court-told

of terrorist organisations by providing information on bomb or weapon making, how to behead someone, and how to avoid detection (Lakomy, 2023; Bunker and Bunker, 2023). These activities are all influence-dependent, relying on the participation of susceptible people (whether by ignorance, choice, coercion or deception) to conduct harm.

That current-state generative AI CaaS models are unlikely to face much resistance and are likely to continue to evolve unimpeded, at least over the short term, is of concern. However, the newness of generative AI technology also reduces the likelihood that those tasked with preventing and investigating cybercrime and terrorist activity will be adequately prepared to deal with such technically complex and resource-intensive tasks. While criminals and terrorists can simply purchase a CaaS product online at any time, law enforcement and intelligence agencies are rightly beholden to a range of legislative checks and balances to procure and deploy even the smallest of counter-capabilities. Where capability does exist, legislative remits may not. The result will be the creation of a permissive space for crimes to be committed and terrorist plots to succeed.

Future threatscape

By taking a horizon-scanning approach to future generative AI-led cybercrime alongside a maturing understanding of present-day CaaS innovation, researchers and scholars have begun forecasting risks. HYAS, a company which specialises in cyber-adversary infrastructure, explains that while AI and generative AI are already being exploited by cybercriminals who are creating increasingly sophisticated autonomous cognitive threat agents, generative AI tools have 'the potential to completely revolutionise the landscape of cyber threats (because they) mimic the adaptability of biological viruses, constantly observing their environment and mutating to exploit beneficial circumstances... an opportunistic predator... who can choose its targets and decide when to lay dormant and how to strike to maximise its impact' making it an ever-present, dynamic threat. At a forensic level, polymorphic malware can also 'alter its appearance and behaviour' all while continuing to seek out targets and executing exploits. This technological evolution speaks to an increasingly sophisticated social-engineering capability (HYAS, 2023) and indication of the complexity of the challenges ahead. While individuals will likely remain targets, so too will entire computer systems of value. The ability to stage an intrusion into systems to cause stock market calamities, currency fluctuations, trade-based crises and widespread financial system chaos (Caldwell et al., 2020) leaves insider threats and poor cybersecurity hygiene in workforces at risk of exploitation. These risks will increase as more organisations adopt AI and generative AI as part of their operations, making the use of AI and generative AI for cybersecurity purposes essential.

The exploitation of biases by social engineers also presents significant risks to social cohesion, geopolitical stability and democracy. While present-day observations about the harms fake news, misinformation and disinformation are becoming increasingly clear, MI5 and the FBI were astute in their recognition of election meddling and foreign interference as key benefactors from advances in generative AI (Corera, 2023). For

example, New Zealand researchers at the University of Auckland's Hate and Extremism Insights Aotearoa (HEIA) research laboratory observed in a recent report that democratic backsliding is both a cause and effect of disinformation (HEIA, 2023). With disinformation being a well-used conduit for foreign interference and political deepfakes having already surfaced (Appel and Prietzel, 2022), the potential for generative AI-produced content to mislead and deceive is clear. It is, therefore, likely that generative AI-based deception will have long-term, cascading effects on cognition by reshaping people's worldviews on important political and social issues of the day.

To a degree, cybercriminals are already somewhat successful at this, spoofing branded content, for example, and engaging in social-engineering exploits that deliver some results. The future of these types of generative AI-led CaaS, and the impact they will have on victims however, has the potential to do far more than defraud. Early studies into the pervasiveness of augmented and virtual reality devices, for example, have already found that such immersive environments awaken spatial memory (Heller, 2021) which could have cascading effects on the psychological wellbeing of those who become victims of generative AI-led cybercrime. That most victims will have been influenced into co-opting themselves into a process that causes their own harm and distress is of significant concern. Researchers may find some answers to these challenges from within the criminology and psychology disciplines, where the effects of coercive control and intimate partner abuse have been well studied.

Old crimes, new tools, slow legislature

The evolution of generative AI occurs at a time when the legislation available to combat cybercrime, particularly at a transnational organised level, remains globally largely unfit for purpose. While individual nations may have varying degrees of legislative and regulatory options available to apply to cybercrime challenges, in many cases those mechanisms have not kept pace with advances in internet-enabled technology nor the types of behaviours offenders employ to influence others towards harm. From a victim perspective, law enforcement agencies are often ill-equipped to deal with the types and volume of harm being committed online. Even if they were adequately resourced and supported by up-to-date legislative frameworks, such work would fall to an even smaller forensically capable workforce and then an overwhelmed judiciary. Due to the slow-moving nature of legislative change, generative AI will enable crimes of influence to outpace existing capabilities further. These longstanding issues notwithstanding, governments are giving significant thought and committing increasing resources to the challenge. A globally comprehensive AI legislative tracker, which is regularly updated, can be found online courtesy of the International Association of Privacy Professionals.[15]

15 The International Association of Privacy Professionals online 'Global AI Legislation Tracker' (IAPP, 2023) 'identifies legislative policy and related developments' in Australia, Brazil, Canada, China, the European Union, India, Israel, Japan, New Zealand, Saudi Arabia, Singapore, South Korea, the United Arab Emirates, the United Kingdom and the United States. While not globally comprehensive, it does provide a useful (and regularly updated) central repository of knowledge.

In addition to legislative mechanisms, governments, industry, academia and civil society have developed a range of fora that acknowledges not only the globalised nature of the challenge as a whole, but also the critical need for a globally cohesive solution. While some fora see many nations working together on these issues, many of the most successful fora exist within multistakeholder settings that focus on initiatives that co-exist alongside the legislative and regulatory environment. For the purposes of this paper, a brief overview of some of the larger governmental collaborations and successful multistakeholder initiatives that specifically address cybercrime or particular harm types follows.

Government initiatives

The United Kingdom's Bletchley Declaration (2023), taking a broad approach that recognises 'the potential for unforeseen risks stemming from the capability to manipulate content or generate deceptive content' among other cyber concerns, seeks to ensure frontier AI capabilities are safe and that an understanding of concerns and risks is evidence-based (GOV.UK, 2023). The G7 Hiroshima Process on generative AI, while proposing a broad set of guiding principles for organisations developing advanced AI systems, specifically noted that technology development should not develop or deploy 'systems in a way that undermine democratic values, are particularly harmful... facilitates terrorism, enable criminal misuse, or pose substantial risks to safety, security and human rights'. The G7 also committed to multistakeholder work on monitoring tools and accountability (OECD, 2023). The European Union's (EU) AI Act is set for final approval from the European Parliament in April 2024. If the legislation passes as it stands, the Act will become law in 2026 and effectively regulate AI models based on potential risk. The EU AI Act places particular focus on the development of safe AI that upholds EU values and respects human rights. In practice, this approach attempts to deal with the issues that general-purpose AI models have (such as unpredictable uses) while allowing for lower-risk AI innovation (Gibney, 2024). The United States of America's executive order on 'Safe, Secure, and Trustworthy AI' (2023) sets new standards for safety and security to manage the risks of AI. Taking a 'responsible innovation' approach, the order builds on the existing voluntary commitments involving 15 companies that are at the leading edge of AI development. The order specifically directs that protections from 'AI-enabled fraud and deception' are necessary for detecting AI-generated content and authenticating official content' (White House, 2023). Other nations currently exploring legislative and regulatory mechanisms for AI broadly include Australia where a regulatory approach is undergoing public consultation; China which has implemented interim measures to manage generative AI; France which is investigating data-protection breaches; Ireland with a view that generative AI needs to be regulated; Israel which is undertaking public consultation on a national AI policy; Italy which temporarily banned ChatGPT in April 2023 over data-protection concerns; Japan which is seeking to introduce a regulatory framework; and Spain which is also investigating data-protection breaches (Reuters, 2023). The United Nations Educational, Scientific and Cultural Organization (UNESCO)

has called on governments to regulate generative AI in education and research. UNESCO is particularly concerned with the 'harm and prejudice' generative AI may incubate, and is advocating for a global standardised approach (UNESCO, 2023). The United Nations Treaty on Cybercrime is also considering, broadly, cyber-enabled crime. While final negotiations were due to conclude in February 2024, concerns from civil society and industry remain about the treaty's impact on human rights, along with reservations from some nations about the document's scope. The treaty's applicability to generative AI appears to be implied as an enabling technology (Kazakova et al., 2023).

Other legislative instruments with a focus on multijurisdictional cybercrime more broadly include the Budapest Convention, led by the Council of Europe, which is currently the only international treaty that 'criminalises conducts and typologies committed through computer and information systems'. Its main objective as an instrument is to enable seamless information sharing between signatories, providing a legal basis for disclosure and preservation of information across a range of user, subscriber and other forensic data points. With a focus on international co-operation, currently 66 nations are party to the convention. The Cybercrime Convention Committee (T-CY), formed by nations party to the convention, have given some thought to AI as a vector of crime given the neutrality of its drafting 'precisely because the original drafters... anticipated how the threat landscape... would likely evolve'. The Lanzarote Convention, another Council of Europe international treaty, has 48 state signatories but is yet to fully explore how it may be applied in practice within a generative AI cybercrime setting. The Istanbul Convention, another Council of Europe initiative, has 34 state parties and a reference group of experts. Yet the convention itself does not address crimes committed online, although its reference group has made recommendations to address this (Velasco,2022).

Multistakeholder initiatives

There are several multistakeholder initiatives working closely on challenges that are likely to occur as cybercrime and generative AI converge or which are harm-adjacent. New Zealand and France's Christchurch Call, with its focus on eliminating terrorist and violent extremist content online, is working with its multistakeholder community to 'contribute to the development of frameworks... to identify, report, and mitigate terrorist and violent extremist exploitation' of generative AI tools (Christchurch Call, 2023). The Global Partnership on AI (GPAI) with a secretariat hosted by the OECD AI policy observatory, has established a working group on harmful online content such as hate speech. Taking a content-moderation view of potential AI solutions, the GPAI advocates for responsible development and adoption of AI (GPAI, 2023). The illegality of hate speech differs significantly between jurisdictions, which AI and generative AI may be particularly well-suited to navigating in a borderless social media and online environment.

Beyond foreign interference, misinformation and disinformation, thinking on influence as a vector that underpins cybercrime, appears to have been considered in depth only within multistakeholder initiatives. The Christchurch Call, for example, considers the role

of algorithmic influence in radicalisation towards violent extremism. Similarly, the GPAI has conducted research into AI-based online recommender systems in the context of driving users towards terrorism and violent extremism content (Christchurch Call, 2023; GPAI, 2023). These types of initiatives, unlike governmental focus on legislative and regulatory responses, go some way towards building an understanding of the upstream types of influences that users are routinely subjected to that contribute to the creation of downstream victims of cybercrime. However, such knowledge is useful only if its findings are put into practice. The solution here also leans towards multistakeholder initiatives, that are, by design, more agile and able to cultivate cross-sector relationships in a less constrained environment. That is not to diminish the important work of governments, only to point out that the roles and duties citizens expect of their elected officials necessitate a different way of working with their stakeholders.

Conclusion

The cognitive frontier that generative AI heralds has the potential to enable cybercrime in ways that are largely unavoidable for victims. While the technology involved in building environments to deceive, influence and manipulate assist this process, at a fundamental level, generative AI challenges the way human cognition moves between System 1 and System 2 thinking. By quickly building rapport and familiarity with targets, whether by email, text, social media or virtual environments, generative AI's pervasiveness encourages an overreliance on System 1 thinking to avoid the scrutiny System 2 entails.

When considering how generative AI-enabled cybercrime, transnational organised crime and violent extremism may evolve, it is easy to catastrophise a possible future without giving due consideration to the opportunities that will also arise to mitigate, and perhaps even prevent, some of those risks. At a fundamental level, as has been the case with web 2.0, there will always be those who exploit advances in technology for nefarious purposes. This is the nature of humanity, not necessarily the nature of technology. Catastrophising generative AI fixes attention on threats instead of encouraging people to see their way through complexity towards solutions. That generative AI comes at a time of heightened global polarisation, regional conflict, a highly contested information environment and democratic backsliding should not go unnoticed. When faced with change, humans instinctively and consistently default to fearing the unknown. Becoming comfortable with complexity, and daring to try to make sense of it, is a type of mental gymnastics that most find uncomfortable. However, if scholars, researchers, civil society and industry don't push past this discomfort, we risk missing the ground-floor opportunities available to address some of generative AI's biggest risks. Multistakeholder initiatives are well placed to help navigate this environment and any dystopian view of where generative AI may lead humanity needs to be balanced with technological optimism.

There is no doubt that technological guardrails for generative AI will be required. However, any approach to designing these guardrails must be robustly evidence-based. Likewise, the design of effective guardrails is not just a problem for the technology industry.

Robust, defensive protective and detection-capable mechanisms are the combined result of appropriate, and human rights affirming, safety by design frameworks that are underpinned by internationally cohesive legislative conventions supported by reasonable jurisdictional regulation. Without a level of international and virtual legislative cohesion and co-operation, questions such as who will be responsible for generative AI's conduct, who is in charge of investigations of the cybercrimes that arise, and who is responsible for prosecuting such multi-jurisdictional crimes, will remain unanswered. The real-world consequences of failing to approach this challenge with a global outlook is that it will create permissive environments ripe for exploitation.

What is largely missing from this discussion, however, is consideration of the cognitive impact on enabling near-horizon and future cybercrime. Influence rarely occurs in isolation yet thrives in vacuums. Without addressing the psychosocial aspects of generative AI-led cybercrime, work in this domain will only focus on 'bottom-of-the-cliff' solutions that come too late to be considered preventative. While, in part, this can be attributed to the emergent nature of both the type of offending and understanding of this field of study, an overreliance on risk as a lens to inform expected victim outcomes is problematic. By taking a speculative approach to conceptualising risk in this context, in isolation of understanding real-world as-it-happens victim experiences, researchers miss qualified insight into how these exploits were able to be detected, what kinds of persuasive behaviours were employed, whether the exploits succeeded or failed, and why. By studying victim experiences in the generative AI-led CaaS environment, scholars and practitioners can begin to develop protective and preventative solutions to address the generative AI-led crimes of influence of the near future.

Conflict of interest

The authors declare that they have no conflicts of interest. This paper does not represent the views or policy of any of the authors' employers, associations or affiliations.

Acknowledgments

Thanks to associate professor Nick O'Brien for feedback on an earlier version of this paper.

References

Anthropic (2023) Claude chatbot. Query: 'Tell me how generative AI will change cybercrime'. Accessed and response provided on 14 October 2023. A full transcript of the query and response is available on request.

Appel, M. and F. Prietzel (2023) The Detection of Political Deepfakes. *Journal of Computer-Mediated Communication*, 27(4) pp. 1–13.

Ariely, D. (2008) *Predictably Irrational*. New York: Harper Collins.

Bunker, R.J. and K.O.K. Bunker (2023) *The Terrorism Potentials of ChatGPT and Related Generative AI Models*. A C/O Futures Terrorism Research Note Series.

Byrne, M.D. (2023) Generative Artificial Intelligence and ChatGPT. *Journal of PeriAnesthesia Nursing*, 38 pp. 519–522.

Caldwell, M., J.T.A. Andrews, T. Tanay and L.D. Griffin (2020) AI-enabled future crime. *Crime Science*, 9(14) pp. 1–13.

ChatGPT (2023) an OpenAI chatbot. Query: 'Tell me how generative AI will change cybercrime'. Accessed and response provided on 14 October 2023. A full transcript of the query and response is available on request.

Chebac, A. (2023) *What is Cybercrime-as-a-Services (CaaS)?* Heimdal. Accessed on 25 October 2023 at: https://heimdalsecurity.com/blog/what-is-cybercrime-as-a-service-caas/

Christchurch Call (2023) *2023 Leader's Summit Joint Statement*. Accessed online 11 November 2023 at: https://www.christchurchcall.com/assets/Documents/Christchurch-Call-Leaders-Summit-2023-Joint-Statement-ENG.pdf

Cialdini, R. (2021) *Influence. The Psychology of Persuasion*. New and expanded edition. Harper Collins. pp. 23–363.

Cialdini, R. (2016) *Pre-suasion*. Random House. pp. 7–8.

Corera, G. (2023) AI risks are unknown even to GCHW, Anne Keast-Butler tells BBC. BBC. Accessed on 5 November 2023 at: https://www.bbc.com/news/uk-67301402

Cornell Law School (2024) Legal Information Institute. *Crime of Passion*. Accessed on 17 February 2024 at: https://www.law.cornell.edu/wex/crime_of_passion

Davey, J. and J. Ebner (2019) *The Great Replacement: The Violent Consequences of Mainstream Extremism*. Institute of Strategic Dialogue.

Department of Justice (DOJ), United States of America (2021) *FBI's Encrypted Phone Platform Infiltrated Hundreds of Criminal Syndicates; Result is Massive Worldwide Takedown*. Accessed on 16 November 2023 at: https://www.justice.gov/usao-sdca/pr/fbi-s-encrypted-phone-platform-infiltrated-hundreds-criminal-syndicates-result-massive

Enoch, J., L. McDonald, L. Jones, P.R. Jones and D.P. Crabb (2019) Evaluating Whether Sight Is the Most Valued Sense. *JAMA Opthalmol*, 137(11), pp. 1317–1320.

Falade, P.V. (2023) Decoding the Threat Landscape: ChatGPT, FraudGPT and WormGPT in Social Engineering Attacks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* (9)5, pp. 185–198.

Federal Bureau of Investigation (FBI), United States of America (2015) *Ross Ulbricht, the Creator and Owner of the Silk Road Website, Found Guilty in Manhattan Federal Court on All Counts*. Accessed on 16 November 2023 at: https://www.fbi.gov/contact-us/field-offices/newyork/news/press-releases/ross-ulbricht-the-creator-and-owner-of-the-silk-road-website-found-guilty-in-manhattan-federal-court-on-all-counts

Gibney, E. (2024) What the EU's Tough AI Law Means for Research and ChatGPT. *Nature Articles*. Accessed on 17 February 2024 at: https://doi.org/10.1038/d41586-024-00497-8

Global Partnership on Artificial Intelligence (GPAI) (2023) *What we do*. Accessed on 27 October 2023 at: https://www.gpai.ai/projects/

Google Gemini (2023) a Google chatbot. Query: 'Tell me how generative AI will change cybercrime'. Accessed and response provided on 14 October 2023. A full transcript of the query and response is available on request.

GOV.UK (2023) *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023*. Policy Paper. Accessed on 3 November 2023 at: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

Harms, M.B. (2017) Stress and Exploitative Decision-making. *Journal of Neuroscience*, 37(42), pp. 10035–10037.

Hartvigsson, T. (2021) Between Punishment and Care: Autonomous Offenders Who Commit Crimes Under the Influence of Mental Disorder. *Criminal Law and Philosophy*, 17, pp. 111–134.

Hate and Extremism Insights Aotearoa (HEIA) (2023) *Disinformation Trends in New Zealand: A HEIA Snapshot Report*. 20. Accessed online 2 January 2024 at: https://www.heiaglobal.com/post/disinformation-trends-in-new-zealand

Heller, B. (2021) Watching Androids Dream of Electric Sheep: Immersive Technology, Biometric Psychography, and the Law. *Vanderbilt Law Review*, 23(1). https://scholarship.law.vanderbilt.edu/jetlaw/vol23/iss1/1

Hendrycks, D., M. Mazeika and T. Woodside (2023) An Overview of Catastrophic AI Risks. Version 6, 9 October 2023.

Hong Kong Computer Emergency Response Team Coordination Centre (HKCERT) (2023) *Unmasking Cybercrime-as-a-Service: The Dark Side of Digital Convenience*. Accessed on 24 October 2023 at: https://www.hkcert.org/blog/unmasking-cybercrime-as-a-service-the-dark-side-of-digital-convenience

HYAS (2023) *EyeSpy: Proof of Concept*. Accessed on 26 October 2023 at: https://www.hyas.com/blog/eyespy-proof-of-concept

IBM (2023) What's Next in AI is Foundation Models at Scale. Accessed on 13 October 2023 at: https://research.ibm.com/artificial-intelligence

International Association of Privacy Professionals (IAPP) (2023) *Global AI Legislation Tracker*. Accessed on 28 October 2023 at: https://iapp.org/resources/article/global-ai-legislation-tracker/

Kahneman, D. (2011) *Thinking, Fast and Slow*. New York: Farrar, Strauss and Giroux. pp. 20–21; 97–99.

Kazakova, A., K. Swift and B. Kovač (2023) *Key Takeaways from the Sixth UN Session on Cybercrime Treaty Negotiations*. Geneva Internet Platform, Digwatch. Accessed on 25 November 2023 at: https://dig.watch/updates/key-takeaways-from-the-sixth-un-session-on-cybercrime-treaty-negotiations

Keeler, J.D. (2024) Bribery and Corruption: Crimes of Influence. *Law N Guilt*. Accessed on 17 February 204 at: https://www.lawnguilt.com/bribery-and-corruption-crimes-of-influence/

Khalil, L. (2021) *GNET Survey on the Role of Technology in Violent Extremism and the State of Research Community – Tech Industry Engagement*. Global Network on Extremism and Technology.

Koehler, D. (2014) The Radical Online: Individual Radicalization Processes and the Role of the Internet. *Journal for Deradicalization*, 15(1), pp. 116–134.

Lakomy (2023) Artificial Intelligence as a Terrorism Enabler? Understanding the Potential Impacts of Chatbots and Image Generators on Online Terrorist Activities. *Studies in Conflict & Terrorism*. DOI: 10.1080/1057610X.2023.2259195

Lewis, J. (2018) *Economic Impact of Cybercrime – No Slowing Down*. A joint McAfee and Center for Strategic and International Studies (CSIS) Report. Accessed on 29 October 2023 at: https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/economic-impact-cybercrime.pdf

Makarenko, T. (2004) The Crime-Terror Continuum: Tracing the Interplay between Transnational Organised Crime and Terrorism. *Global Crime*, 6(1), pp. 129–145.

Matejic, N. (2020) *2040: An Information Odyssey*. NATO Innovation Hub: Warfighting in 2040 Report.

Meta (2021) *Trademark/Service Mark Application*. Principal Register. Serial Number: 97097362. Filed on 28 October 2021. Accessed on 1 November 2023 at: https://tsdr.uspto.gov/documentviewer?caseId=sn97097363&docId=APP20211101091335&linkId=9#docIndex=8&page=1

Morgan, S. (2020, November 13) Cybercrime to Cost the World $10.5 Trillion Annually by 2025. *Cybercrime Magazine*. Accessed on 19 November 2023 at: https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/

Munn, L. (2020) Angry by Design: Toxic Communication and Technical Architectures. *Humanities and Social Sciences Communications*, 7, 53. https://doi.org/10.1057/s41599-020-00550-7

Nickerson, R.S. (1998) Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2, pp. 175–220.

Nvidia (2023) *What is Generative AI?* Accessed on 12 October 2023 at: https://www.nvidia.com/en-us/glossary/data-science/generative-ai/

Organisation for Economic Co-operation and Development (OECD) (2023) *G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a Common Understanding of Generative AI*. Report prepared for the 2023 Japanese G7 Presidency and the G7 Digital and Tech Working Group. Accessed on 1 October 2023 at: https://www.oecd.org/publications/g7-hiroshima-process-on-generative-artificial-intelligence-ai-bf3c0c60-en.htm

Reuters (2023, September 11) What are Governments Doing to Try to Regulate AI? *euronews.next*. Accessed on 25 November 2023 at: https://www.euronews.com/next/2023/09/11/which-countries-are-trying-to-regulate-artificial-intelligence

Sabbagh, D. (2023, October 18) Terrorists Could Try to Exploit Artificial Intelligence, MI5 and FBI Chiefs Warn. *The Guardian*. Accessed on 23 October 2023 at: https://www.theguardian.com/technology/2023/oct/18/terrorists-exploit-artificial-intelligence-ai-mi5-fbi-chiefs-warn

Simi, P., K. Blee, M. DeMichele and S. Windisch (2017) Addicted to Hate: Identity *Residual* among Former White Supremacists. *American Sociological Review*, 82(6), pp. 1167–1187.

Stanford University (2023) *The Stanford Emergency Technology Review 2023: A Report on Ten Key Technologies and their Policy Implications*. pp. 21–29.

Strick, B. (2023) *Incitement to Kill: Tracking Hate Speech Targeting Ukrainians During Russia's War in Ukraine*. Centre for Information Resilience.

Synnott, A. (2022) Sight Is Our Most Dominant Sense, But Is It Trustworthy? *Psychology Today*. Accessed on 17 February 2024 at: https://www.psychologytoday.com/intl/blog/rethinking-men/202207/sight-is-our-dominant-sense-is-it-trustworthy

Szabo, N. (1997) The Idea of Smart Contracts. Accessed online 18 February 2024 at: https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/idea.html

Tech Against Terrorism (2023, November 8) *Early terrorist experimentation with generative artificial intelligence services*. Briefing. Accessed on 10 November 2023 at: https://techagainstterrorism.org/news/early-terrorist-adoption-of-generative-ai

Thaler, R. and C. Sunstein (2009) *Nudge. Improving Decisions about Health, Wealth and Happiness*. Penguin Books. pp. 19–112.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2023) UNESCO calls for regulations on AI use in schools. Accessed on 17 October 2023 at: https://news.un.org/en/story/2023/09/1140477#:~:text=The%20UN%20Educational%2C%20Scientific%20and,data%20protection%20and%20user%20privacy.

UOLLB First Class Law Notes (2024) *Under Influence in Contract and Criminal Law*. Accessed on 17 February 2024 at: https://uollb.com/blog/law/under-influence-in-contract-and-criminal-law

US Senate Select Committee on Intelligence (2019) *Senate Intel Releases Election Security Findings in First Volume of Bipartisan Russia Report*. Accessed on 21 November 2023 at: https://www.intelligence.senate.gov/press/senate-intel-releases-election-security-findings-first-volume-bipartisan-russia-report

Valentini, D., A.M. Lorusso and A. Stephan (2020) Onlife Extremism: Dynamic Integration of Digital and Physical Spaces in Radicalization. Hypothesis and Theory Article. *Frontiers in Psychology*, 11, 524. https://doi.org/10.3389/fpsyg.2020.00524

Velasco, C. (2022) Cybercrime and Artificial Intelligence. An Overview of the Work of International Organizations on Criminal Justice and the International Applicable Instruments. *ERA Forum*, 23, pp. 109–126. https://doi.org/10.1007/s12027-022-00702-z

Wanless, A. and M. Berk (2017) Participatory Propaganda: The Engagement of Audiences in the Spread of Persuasive Communications. Paper delivered at Social Media and Social Order conference, November/December 2017, Oslo, Norway.

The White House (2023) FACT SHEET: President Biden Issues Executive Order on Safe, Secure and Trustworthy Artificial Intelligence. Accessed online 1 November 2023 at: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

The White House, President Barack Obama, National Security Council (n.d) Transnational Organized Crime: A Growing Threat to National and International Security. Accessed on 19 November 2023 at: https://obamawhitehouse.archives.gov/administration/eop/nsc/transnational-crime/threat

Williams, P. (2018) The Organized Crime and Terrorist Nexus: Overhyping the Relationship. Stratfor Worldview.

Wolbers, R. (2023) *The Future of the Christchurch Call to Action. How to Build Multistakeholder Initiatives to Address Content Moderation Challenges*. Ian Axford (New Zealand) Fellowships in Public Policy.

World Economic Forum (2023, February 6) *Artificial Intelligence. What is generative AI? An AI explains*. Accessed on 12 October 2023 at: https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/

Zak, P.J. (2017) The Neuroscience of Trust. *Harvard Business Review*. pp. 84–90. https://hbr.org/2017/01/the-neuroscience-of-trust

Zhang, A., T. Hoja and J. Latimore (2022) *Gaming Public Opinion: The CCP's Increasingly Sophisticated Cyber-Enabled Influence Operations*. Australian Strategic Policy Institute International Cyber Policy Centre. https://ad-aspi.s3.ap-southeast-2.amazonaws.com/2023-05/Gaming%20public%20opinion.pdf?VersionId=QYkBIWncbBU0E1KAhg9mX3TD7kwlWcWj

About the authors

Nicole Matejic is a national security-focused behavioural economist and adjunct lecturer at Charles Sturt University in Australia, and completed this paper while also on secondment with the Department of Prime Minister and Cabinet's Christchurch Call Unit in Aotearoa, New Zealand.

Chris Wilson is a senior lecturer in politics and international relations at the University of Auckland (Aotearoa New Zealand) and CEO of Hate and Extremism Insights Aotearoa (HEIA).

ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

The Commonwealth

Ensuring a Secure Future by Insuring Against Cybercrime

Eric Cho[1] and Serene Chan[2]

Abstract

The core concept of insurance is for individuals or entities to manage their risks by transferring such risks to a risk carrier like an insurance company in exchange for an insurance premium. With the escalating occurrence of cyber incidents coinciding with the digitisation of society, the imperative for adequate risk management has become critical in board meetings. With the established role of insurance in mitigating traditional risks, there exists a compelling case for the utility of insurance in addressing the consequences for entities facing emerging cyber vulnerabilities. This article investigates the burgeoning demand for cyber insurance policies, which serve to mitigate financial losses sustained by businesses as a consequence of cyber incidents.

This paper explores the role of cyber insurance. The main themes we investigate are:

1. Introduction to cyber risks and cyber insurance.
2. The benefits and challenges of cyber insurance.
3. Government involvement and the future.

Cyber insurance policies first became available in the 1990s, focusing mainly on third-party liability for cases in which companies may have leaked customer data as a result of a cyber incident. Since its inception, the product has seen rapid evolution and growth, driven by the need for more comprehensive coverage for clients and a general increase in awareness of cybersecurity.

The increase in adopting cyber insurance is a testament to its benefits. Cyber insurance is a core risk-management solution for companies to transfer their underlying cyber risks. Despite companies investing more in their cybersecurity, there are numerous cases of cybercriminals gaining unauthorised access to data, leaving companies and their customers exposed to financial loss. As long as the incident is insurable, companies can claim against their policies to reduce the financial losses and, in most cases, receive crisis-management support. Obtaining

1 Senior Cyber Underwriter, Munich Re. Email: echo@munichre.com.
2 Regional Head of Cyber, Asia Pacific, Munich Re. Email: szchan@munichre.com.

cyber insurance also involves underwriting, for which companies' cybersecurity policies and controls are assessed by insurance companies. This underwriting is rigorous, and necessitates that companies adhere to insurers' cybersecurity expectations in order to qualify for cover.

Cyber-attacks are an almost inevitable fate for many companies. Therefore, cyber resilience is fundamental for successful and sustainable digitisation of the economy and society. Cyber insurance can play a vital role in ensuring a tangible solution for companies. The public sector, including governments and regulators, must also play an active role to catalyse awareness of cyber risk and the corresponding risk transfer solution, as cyber insurance is a relatively new product. Increased dialogue and transfer of knowledge that occurs from cyber insurance can help to foster a more resilient digital economy, safeguarding the interests of individuals, businesses and society.

Introduction

Before examining the intricacies of cyber insurance, it is important to understand the context for such a risk transfer solution. With the rapid advancement of digital technology in recent years, cyber risk has synchronously emerged for individuals, companies and nations. Cyber risk can be defined as 'any risk of financial loss, disruption or damage to the reputation of an organization from some sort of failure of its information technology systems' (Institute of Risk Management, 2023). In today's competitive, globalised markets, companies have targeted operational efficiencies by adopting modern technologies that enhance their ability to deliver value to customers.

The increase in centralisation and interconnectivity between new and old technologies has led to what the industry refers to as the expansion of attack surface (One Identity, 2024). The attack surface is, effectively, the sum of all possible points where an unauthorised user or system could try to enter, or extract data from, an environment. In other words, the more interconnected a system is, the wider the attack surface and the more systems may be affected at any one time, resulting in a more severe business impact following a cyber-attack.

An example of the risks relating to the expansion of the attack surface is the 2017 NotPetya ransomware strain, one of the most destructive malwares in history, which caused $10 billion in damages to companies globally. The origin of the attack could be traced to commonly used tax software in the Ukraine (HYPR, 2023). The dependency of companies and their subsidiaries on interconnected technology (in the case of NotPetya, the connection to a third-party software), creates a prime opportunity for criminals who want to extort money by crippling companies' computer systems and making ransom demands for stopping the attacks.

Ransomware is

> *'a type of malware that prevents or limits users from accessing their system, either by locking the system's screen or by locking the users' files until a ransom is paid. More modern ransomware families, collectively categorized as cryptoransomware, encrypt certain file types on infected systems and force users to pay the ransom through certain online payment methods to get a decryption key'.*
> (Trend Micro, 2024)

The first ransomware known was the AIDS Trojan in 1989: Joseph Popp handed out 20,000 floppy disks with the malware to attendees at a World Health Organization conference. Those whose files were encrypted by the ransomware needed to send $189 to a PO box in Panama (Kostka, 2022). Today, ransomware demands can be in the hundreds of millions of dollars, usually in the form of cryptocurrency, with the average ransom payment in 2023 being $1,542,333 compared to $812,380 in 2022 (Sophos, 2023). What used to be a technical term has become a household word because of the numerous headlines about companies and individuals falling victim to ransomware attack.

Recognition of cyber risk

As cyber risk and its consequences have become more significant, managing cyber exposure has become a top priority for many companies and countries. A severe cyber-attack not only poses a tangible threat to a company's operations but also inflicts substantial harm on its reputation, eroding customer trust. A report by the International Data Corporation showed that '80% of consumers in developed nations will defect from a business because their personally identifiable information is impacted in a security breach' (Lieberman, 2017).

Furthermore, cyber risks extend beyond technological disruption and business impact: they have far-reaching implications for national security as well. A breach in cybersecurity can compromise sensitive government data, disrupt critical services and undermine the economic stability of a nation. With the rise of sophisticated cyber threats, including state-sponsored attacks and ransomware incidents, the potential for disruption, economic espionage and the compromise of critical infrastructure becomes more pronounced. Effective cybersecurity and mitigating measures are essential for protecting national interests, preserving economic vitality and safeguarding citizens from the far-reaching consequences of cyber-attack. This makes them fundamental to national security.

Regulators around the world have responded to the escalating threat of cyber attacks by implementing various measures and regulations aimed at enhancing cybersecurity and protecting sensitive data. These include mandatory breach notifications such as the General Data Protection Regulation (GDPR) in the EU and the UK, industry-specific regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in

the US, cybersecurity audits and assessment, and increased scrutiny resulting in hefty fines and penalties being imposed. These are all to keep companies accountable for their cyber risk management.

Cyber risk has, therefore, been recognised as a top priority for management across all levels and industries. One major reason why it has become such an important topic among executives is the potentially significant financial consequences of a cyber incident. The global average cost of a data breach in 2023 was $4.45 million, which was a 15 per cent increase over three years (IBM, 2023). Due to the significant financial and regulatory ramifications of a cyber breach, an increasing number of annual reports include cyber-attacks as one of their top risks.

Cybersecurity market

As the world continues to invest in modern technologies and the interdependency of systems, the need to manage the resulting risks is apparent, with the first step being prevention.

Several companies have emerged in the cybersecurity industry in providing technical solutions. According to McKinsey, 'organisations around the world spent $150 billion in 2021 on cybersecurity, growing by 12.4 percent annually'. This sum is based on a 10 per cent penetration of the market, meaning that the total addressable market size of cybersecurity is expected to be around $1.5 to $2 trillion (Mckinsey, 2022). Given that cybercriminals create progressively more sophisticated attack methods and identify new vulnerabilities to exploit, companies need to rely on dedicated experts both in-house and third parties. Companies are spending considerable efforts on optimising cybersecurity and are constantly looking for ways to protect themselves against threat.

Cyber insurance market

Many insurance companies have formed cyber insurance departments that focus on underwriting cyber risks and providing cyber insurance solutions. Insurance has, historically, played a vital role, allowing individuals and corporations to shift their risks to an insurance company in exchange for insurance premiums. In our increasingly digitised world, the development of solutions that allow companies to mitigate cyber risks is both relevant and crucial. Given the constantly evolving nature of cyber threats that can exploit undiscovered vulnerabilities, companies face perpetual risk. Consequently, having cyber insurance as a second line of defence becomes a prudent and important risk-management option.

History and purpose of cyber insurance

Cyber insurance first emerged in the late 90s, sold through security software companies that partnered with insurance companies (Holot and Lelarge, 2008). Projecting into the future, it is estimated that the global cyber insurance market is worth $12.1 billion in

2023, and is expected to increase to $90.6 billion by the end of 2033 (Market.us, 2024) As one of the newest products in the insurance market, many insurance companies target cyber as one of the strategic growth segments. Also, insuring a critical and emerging risk such as cyber risk allows insurance companies to stay relevant.

So, what is cyber insurance and what does it cover? The purpose of cyber insurance policies is to cover certain financial losses that an organisation has to bear as a result of a cyber incident. The cover of a cyber policy typically falls into two categories: first-party cover and third-party cover.

Originally, cyber insurance policies were primarily created to address third-party liability risks. For example, if an organisation's network is breached and their customer data leaked, customers could launch a class action against the organisation on the basis of a breach of privacy. In such cases, cyber insurance could cover liabilities and legal costs. Over time, as cyber incidents have increased, so has the prevalence of first-party financial losses, resulting in an increase in demand for first-party cover. First-party cover includes necessary expenses relating to recovering systems that have been compromised by cyber-attack. For example, IT forensics investigation expenses or the costs of restoring data could be covered by a cyber insurance policy. Cyber insurance is, fundamentally, a risk-management solution that companies adopt to safeguard against potential financial losses in the event that their cybersecurity measures fail to prevent a breach.

Availability of cyber Insurance

A survey of global C-level executives showed that 33 per cent of participants were never offered cyber insurance. Also, when asked why their company did not have cyber insurance, 25 per cent stated that they did not know cyber insurance existed; 38 per cent of those who did not know were from smaller companies with revenues below $1 million [Munich Re, 2022]. Cyber insurance is generally an underpenetrated insurance policy and appears to be more of a 'reactive' purchase. For example, 32 per cent of companies purchased cyber insurance after a cyber-attack, and 37 per cent of companies purchased it as a reaction to a cyber-attack in a peer company (Deloitte, 2019). During the initial phases of cyber insurance availability, there was scepticism about its necessity, with companies prioritising cybersecurity tools and considering them sufficient to thwart all cyber threats. The dilemma was compounded by the reluctance of IT managers, responsible for overall cybersecurity, who were often known in the industry as adopting a defensive stance. This posed a challenge for risk managers tasked with deciding whether or not to invest in cyber insurance when confronted with the reservations of their counterparts in IT management.

Despite these reservations, the continued increase in severe incidents affecting even the most highly resourced companies resulted in more understanding that cyber risk is essentially inevitable. The innately dynamic and evolving nature of cyber threats meant that it is challenging to eliminate the risk entirely, and the substantial market size of $11.9 billion in 2022 noted above suggests a tangible demand for such a product.

The cyber insurance industry has also proven its ability to compensate and to pay claims. These increased with the accelerated expansion of attack surface owing to increased remote working during the pandemic. From 2018 to 2021, reported claims in the United States cyber insurance market grew 100 per cent annually. In the same period, of those reported claims, there was a 200 per cent annual increase in the number of claims with payment, reaching a total of 8,100 claims paid in 2021 (Fitch Ratings, 2022).

How cyber insurance supports the improvement of cybersecurity

A key benefit of cyber insurance is that it can encourage companies to improve their cybersecurity. When organisations apply for cyber insurance, they must fill out an application form or questionnaire. Traditionally, these application forms contained questions relating to the applicant's cybersecurity position such as, 'Do you have a business continuity plan in place?' By collecting responses to these questions, in addition to general information about the organisation (such as revenue, amount of personal data held), a cyber insurance underwriter can assess the exposure and risk of the organisation. Once a cyber insurance underwriter evaluates the cybersecurity controls and governance of a company, they can structure and price an insurance policy suitable for that organisation.

As cyber threats have become increasingly sophisticated, organisations have developed sophisticated IT infrastructure. It is becoming commonplace for insurance companies to have risk-assessment conference calls or on-site inspections to evaluate the cyber risks. In the case of larger companies, this has become a minimum requirement by insurance companies as part of the underwriting process. Thus, companies that apply for cyber insurance can become more aware of their exposure and any weakness in their cybersecurity controls. The insurance industry has valuable insights about cyber threats and loss information that other sectors may not have and can serve as a hub for knowledge exchange. Cybersecurity companies also leverage their services to assist policyholders in improving their risks in order to be more insurable.

As cyber insurance gains prominence, insurance companies have started to offer complementary services for insured companies to improve their cybersecurity. This creates a win-win situation. Organisations can benefit from additional risk-assessment perspectives and consulting, and insurance companies have a better understanding of organisations' cybersecurity policies and controls. Certain insurance companies incentivise cybersecurity improvements by offering more competitive terms and conditions to organisations with robust controls.

Some insurance companies go further in mandating certain cybersecurity controls as a prerequisite for cyber insurance cover. With an insurance quote offer, there can be 'subjectivities'. These are conditions that an applicant needs to fulfil such as a security control being implemented (Woods and Simpson, 2017). For example, if the applicant

responds in the insurance questionnaire that they do not have multifactor authentication, the insurance company could stipulate that, in order for the company to receive cyber insurance cover, they must implement such a system within six months after the inception of the policy. Incorporating subjectivities ensures that applicants' cybersecurity policies and controls achieve a level that insurance companies are comfortable to insure, meaning that applicants can obtain cyber insurance cover.

Cyber insurance will play a critical role in encouraging organisations to improve their cybersecurity standards. When – not if – an organisation is faced with a cybersecurity incident, cyber insurance provides the necessary resources for it to mitigate losses and to obtain indemnification for covered losses. Cyber insurance is not a replacement for improvements in cybersecurity. Organisations sometimes face budgetary decisions between investing in cybersecurity or obtaining insurance. However, they must continually demonstrate to insurers that they manage cyber risks effectively through robust governance.

Cyber insurance response to incidents

A significant benefit of having cyber insurance is that companies can access a claims service when there is an incident. Generally, insurance companies establish partnerships with various cybersecurity companies, incident response providers, and law firms with experienced legal practitioners capable of guiding victims of a cyber-attack through the cyber crisis. Often, there are also experts in cyber extortion incidents, who have intelligence about the various ransom tactics of different ransomware gangs, ensuring that companies take the best actions. In a ransomware scenario in which a company is at the mercy of hackers, having an expert who knows how to handle such situations benefits the company greatly. The expert can also advise, for example, that paying the ransom may not benefit the policyholder in certain situations.

Mandating cyber insurance?

On June 12 2018, the South Korean government passed an amendment to its Act on Promotion of Information and Communications Network Utilization and Data Protection (Network Act) (adopted on May 12, 1986). which required companies handling large amounts of personal information to have liability insurance that would compensate them in the case of a cyber incident (Yulchon LLC, 2019). Despite having a law that requires companies to have such cyber insurance, the take-up of liability insurance was lower than expected. Confusion about the amendment, minimal enforcement by the authorities, and a lack of growth in business discouraging insurance companies, have contributed to the lack of adoption (Kim, 2023).

If, and when, countries realise the need for risk management and transfer for such cyber risks to warrant mandatory insurance, the subsequent steps become crucial. Given the previous shortcomings in implementing data-protection and associated laws, achieving success in

mandating risk transfer would require awareness campaigns, meaningful regulatory actions for non-compliance, and clarity for companies about intent and consequences.

Challenges of cyber insurance

Cyber insurance, being a relatively new product, is open to more challenges than life, property and accident insurance. Data is relatively scarce given the newness of the product (Awiszus et al., 2023). Insurance companies employ actuaries who use historic data to try to model and price the expected loss associated with certain risks. The limited data about cyber insurance makes it challenging to use historical pricing methodologies and it often requires expert judgement. However, there is also a question about whether historical data can truly reflect future risks, especially for cyber risk. Technology's rapid evolution means new vulnerabilities and attack threats which may not be accounted for in historical data. The dynamic nature of cyber risk makes it a 'highly non-stationary' risk (Awiszus et al., 2023). One of the main reasons for this is that cyber risk is a manufactured risk. Unlike natural catastrophe risks like hurricanes, for which weather patterns can be analysed to estimate severity or location, for cyber risks, an individual cybercriminal's actions can create an entirely new threat. Therefore, even if historical data is available, its value in predicting the future behaviour of cybercriminals is questionable.

Related to data, cyber underwriters typically depend on 'yes' or 'no' answers in assessing cybersecurity risk. This is challenging. For example, a questionnaire may simply ask whether or not a company has a privacy policy. A positive response may not provide sufficient insight into the quality and appropriateness of the privacy policy, nor the company's procedures for reviewing, implementing and updating it. The need for efficiency in the insurance application process to enable distribution and reach, often leads to using binary questions instead of open-ended inquiries that could result in better insights. Understandably, SMEs prefer a streamlined process and may be discouraged if they need to have a lengthy risk-assessment call with an IT professional as part of the insurance application. Furthermore, the insurance company would need to weigh the resources and time invested against the value of a small insurance policy. There is a trade-off between efficiency and ensuring that enough risk information is captured. As a result, SME companies generally fill out shorter questionnaires related to their cybersecurity controls and governance, whereas large corporations may be subject to a full risk-assessment conference call.

Due to the limited number of questions which can be asked, every insurance company has its own cyber insurance application as priorities in risk assessment may be different. This results in a lack of standardisation and consensus on what information is critical for underwriting. The European Union Agency for Cybersecurity (ENISA) conducted a study of the cyber insurance questionnaire of the top ten cyber insurers, to assess how many unique questions were asked (unique meaning the question is asked by only one insurer out of the ten). The analysis found that there were 129 unique questions. Although there were some core questions and areas that all the questionnaires asked, there was still

a lack of standardisation in data required for underwriting (European Union Agency for Network and Information Security, 2017). This can cause confusion for policyholders if they are requesting insurance from different insurance companies, and facing different questions which may be relevant for only some insurers.

Controversies behind cyber Insurance

Despite its growing prominence, cyber insurance has also faced scrutiny and criticism. The nascent nature of the cyber insurance market, coupled with the rapid evolution of cyber threats, has resulted in a lack of standardisation in cover and wording. This adds to confusion for policyholders (Deloitte Center for Financial Services, 2017). As more cyber incidents occur, and thus both policyholders and insurance companies gain experience, there is an expectation that the cyber insurance market will eventually achieve greater clarity and standardisation.

There are also challenging views in the insurance industry about the product, with arguments about the insurability of the risk in itself due to, among other concerns, lack of data and accumulation scenarios resulting from the interconnectivity of the risk (Greco, 2022). Unlike property risks that can be physically bifurcated and quantified through zoning and addresses, computer systems, software and applications are connected remotely, and the risk is inherently intangible. The resulting cascading impact of an outage to a large interconnected system may have significant consequences that are challenging to quantify objectively.

In confronting a risk that persists for us and future generations, finding a solution becomes imperative. To do so, the cyber insurance industry needs to continue finding solutions which can be sustained in the long run.

Cyber war

A fundamental aspect of sustainability in insurance is ensuring that unquantifiable risks are eliminated, and that there is sufficient capital backing to support accurately assessed and quantified risks.

One of the largest unknown cyber scenarios to consider is cyber war. Cyber war is a complex topic, the definition of which even scholars and experts struggle to reach consensus (Ashraf, 2021). The most catastrophic scenario envisioned by many regarding cyber-attacks comes in the form of cyber warfare between nations. Imagine national telecommunication systems facing outages due to disruptive cyber-attacks; hospitals not being able to operate due to their systems being encrypted; and individuals not being able to access their finances because of an outage of national banking systems. In such a scenario, individuals, companies and nations themselves would face tremendous losses.

The crucial question arises: can the exposures of a cyber war also be shifted to insurance companies? In general, armed conflicts between nation states are by

their very nature a matter for governments, and it is for the state to intervene to mitigate the consequences of a war, for the population but also for the economy, as its consequences are so large and wide-reaching that private industry simply lacks the capital to support such a ruinous risk. Therefore, looking at conventional lines of insurance such as property and accident, insurance policies have long incorporated exclusion clauses for war risks, as the industry is generally incapable of accurately predicting the likelihood or severity of damage arising from war. It is, therefore, unable to charge the appropriate premiums (Kathy, 2022). This has resulted in specialist insurance markets for war for which limits of liability and corresponding conditions are carefully and conservatively managed.

Generally, insurance policies will contain a war exclusion that 'specifically excludes coverage for acts of war, such as invasions, insurrections, revolutions, military coups, and terrorism'. This also holds true for cyber insurance policies (Kagan, 2023). The magnitude of damage that would arise from cyber warfare would be so immense that the insurance industry would not be able to pay every policyholder sufficiently.

Merck, a large pharmaceuticals company, was affected by the NotPetya attack in 2017, involving 40,000 of its computers globally (Tilley and Poulsen, 2023). As a result of the attack, Merck claimed damages under its property 'all risk' policy which included affirmative cyber coverage. However, as NotPetya was allegedly connected to the Russian government, the insurance company tried to deny the claim by applying the war exclusion. As it was a property insurance policy, it had a general war exclusion, which read:

> 'A. 1) Loss or damage caused by hostile or warlike action in time of peace or war, including action in hindering, combating, or defending against an actual, impending, or expected attack:
>
> a) by any government or sovereign power (de jure or de facto) or by any authority maintaining or using military, naval or air forces;
>
> b) or by military, naval, or air forces;
>
> c) or by an agent of such government, power, authority or forces;
>
> This policy does not insure against loss or damage caused by or resulting from Exclusions A,B, or C, regardless of any other cause or event contributing concurrently or in any other sequence to the loss.' (Tilley *et al.*, 2022)

Essentially, the courts found that conventional war exclusion could not be relied upon by the insurers. The insurers submitted that the NotPetya attack was state-backed with 'ill will or a desire to harm' which would be applicable to the 'hostile/warlike' part of the war exclusion (Liu, 2023). However, the courts opined that the insurers were 'stretching the meaning of "hostile" to its outer limit', and that the attack 'is not sufficiently linked to a military action or objective as it was a non-military cyberattack against an accounting software provider' (Liu, 2023).

The case of Merck was pivotal in demonstrating how a traditional war exclusion may not be clear enough in the context of cyber-attacks. Cyber insurance policies, in general, have adopted the same war exclusions as those of traditional property policies. However, given the decisions of the judge in the Merck case, it clearly shows that, for war exclusions to function as intended for the purposes of eliminating cyber warfare scenarios, further clarification of the exclusionary language would be needed. This has sparked many discussions within the cyber insurance industry on how to appropriately exclude cyber war exposures in order to achieve a sustainable risk transfer solution for the market.

In 2022, Lloyd's, which is one of the world's largest insurance marketplaces, published a bulletin outlining minimum requirements for what it considered a robust cyber war exclusion:

1. 'Exclude losses arising from a war (whether declared or not), where the policy does not have a separate war exclusion
2. (Subject to 3) exclude losses arising from state backed cyber-attacks that (a) significantly impair the ability of a state to function or (b) that significantly impair the security capabilities of a state
3. Be clear as to whether cover excludes computer systems that are located outside any state which is affected in the manner outlined in 2(a) & (b) above, by the state backed cyber-attack and
4. Set out a robust basis by which the parties agree on how any state backed cyber-attack will be attributed to one or more states
5. Ensure all key terms are clearly defined' (Lloyd's, 2022).

Today, many insurance companies have started to update their war exclusions in accordance with the above, knowing that traditional war exclusions that are found on other insurance policies do not work for cyber insurance policies. Munich Re, one of the largest reinsurers in the world, has stated that clear and transparent cyber war exclusions are one of the cornerstones of a sustainable cyber marketplace (Shi and McNestrie, 2023). Critical components are clarity in policy language and ensuring uninsurable risks such as war are excluded.

Government and cyber insurance

Governments play a key role in increasing awareness of cyber insurance and fostering a nation in which individuals and organisations are well-protected. As shown in the case of South Korea, one method is to make cyber insurance compulsory for organisations. However, this is not the only method, and it requires strong enforcement to ensure it is effective. General regulations related to data protection, not specific to cyber insurance, can still be effective in stimulating the growth of cyber insurance. For example, demand for cyber insurance across Europe increased after the introduction of the GDPR in 2018

(IFSEC Insider, 2020). With strict data regulations that govern data privacy and create meaningful regulatory consequences for companies that do not comply, this increases awareness in companies that they need to manage their cyber risks appropriately.

Governments can also engage with the insurance industry to identify how a sustainable cyber insurance market can be created. Singapore launched its Cyber Risk Management Project in 2016, in which part of the ambition was to support cyber risk underwriting and pricing in order to create a sustainable domestic cyber insurance market (Wolff, 2022). The government worked jointly with the insurance industry, which led to the announcement in 2018 of the world's first commercial cyber risk pool in Singapore (ibid). The idea was to create more transparency and consistency in order to price cyber risks accurately, incentivising more purchasing by companies.

Another role for governments in fostering a more sustainable cyber insurance market is in data collection. The existing lack of standardisation in data collection poses challenges for insurers striving to gather meaningful and accurate data. Governments could establish data standards or minimum requirements to enhance transparency in risk information, thereby optimising efficiency (Woods and Simpson, 2017).

Governments could address these shortfalls by working with the industry to ensure adequate cybersecurity data is available and transparent for accurate pricing and modelling for cyber risks. This data-driven approach allows for more in-depth analysis. For example, it could determine which cybersecurity controls are most effective for preventing ransomware trends by comparing differences in controls between those companies that have filed a ransomware claim and those that remain incident-free.

Conclusion

Cyber-attacks are inevitable. Therefore, cyber resilience and risk mitigation are fundamental for the successful digitisation of the economy. As more sophisticated cyber-attacks emerge, the demand for cyber insurance will continue to increase. As a result, cyber insurance will play an increasingly critical role in delivering a meaningful risk transfer solution, to enable organisations to continue operating and innovating. Cyber insurance should not be the first line of defence against cyber criminals, just as fire insurance does not prevent fires. However, for fire insurance, policyholders usually implement safety measures to mitigate risk and to be eligible for cover, giving policyholders peace of mind and the assurance that they will be covered if there is a fire.

The public sector needs to play an active role in this ecosystem to ensure that the various stakeholders adequately manage their cyber risks, as this relates to national security. It is incumbent on the insurance industry, governments and the private sector to collectively understand and face the challenges of cyber risks in order to proactively address these and to ensure security against cybercrime for this generation and for future generations.

References

Ashraf, C. 2021. Defining cyberwar: towards a definitional framework. *Defense & Security Analysis,* 37(3), pp. 274–294.

Awiszus, K., T. Knispel, I. Penner, G. Svindland, A. Voss and S. Weber 2023. Modeling and pricing cyber insurance. *European Actuarial Journal*, 13, pp. 1–53.

Deloitte Center for Financial Services, 2017. *Demystifying cyber insurance coverage: clearing obstacles in a problematic but promising growth market.* s.l.: Deloitte University Press.

Deloitte, 2019. *Overcoming challenges to cyber insurance growth.* New York: Deloitte.

European Union Agency for Network and Information Security, 2017. *Commonality of risk assessment language in cyber insurance.* Heraklion: ENISA.

Fitch Ratings, 2022. *US Cyber Insurance Payouts Increase Amid Rising Claims, Premium Hikes.* [Online] Available at: https://www.fitchratings.com/research/insurance/us-cyber-insurance-payouts-increase-amid-rising-claims-premium-hikes-06-05-2022 [Accessed 20 November 2023].

Greco, M. 2022. *Cyber attacks set to become 'uninsurable'.* [Interview] (22 December 2022).

Holot, J. and M. Lelarge 2008. *Cyber Insurance as an Incentive for Internet Security*. Seventh Workshop on the Economics of Information Security.

HYPR, 2023. *Notpetya. Five Facts to Know About History's Most Destructive Cyberattack.* [Online] Available at: https://www.hypr.com/security-encyclopedia/notpetya [Accessed 13 November 2023].

IBM, 2023. *Cost of a Data Breach Report 2023.* Armonk: IBM.

IFSEC Insider, 2020. *Two years on from GDPR: Has it driven growth in cyber security insurance?* [Online] Available at: https://www.ifsecglobal.com/cyber-security/two-years-on-from-gdpr-has-it-driven-growth-in-cyber-security-insurance/ [Accessed 19 November 2023].

Institute of Risk Management, 2023. *Cyber risk.* [Online] Available at: https://www.theirm.org/what-we-say/thought-leadership/cyber-risk/ [Accessed 13 November 2023].

Kagan, J. 2023. *What Is a War Exclusion Clause in an Insurance Contract?* [Online] Available at: https://www.investopedia.com/terms/w/war-exclusion-clause.asp [Accessed 19 November 2023].

Kathy, D. 2022. *Does your insurance cover war?* s.l.:s.n.

Kim, Y.-m. 2023. *Why is the 'cybersecurity insurance product' market so weak in Korea?* Bo-an News, 3 April.

Kostka, C. 2022. *The First Ransomware Attack: Lessons Learned from History,* s.l.: Ransomware.org.

Lieberman, M. 2017. Mind The Trust Gap: How Companies Can Retain Customers After A Security Breach. *Forbes*, 8 December.

Liu, A. 2023. *Merck entitled to $1.4B in cyberattack case after court rejects insurers' 'warlike action' claim.* [Online] Available at: https://www.fiercepharma.com/pharma/merck-entitled-14b-payout-cyberattack-case-after-judge-refutes-insurers-warlike-action-claim [Accessed 19 November 2023].

Lloyd's, 2022. *State backed cyber-attack exclusions.* London: Lloyd's Market Bulletin.

Market.us. (2024, January). Global Cyber Insurance Market By Insurance Type. Available at: https://market.us/report/cyber-insurance-market/

Mckinsey, 2022. New survey reveals $2 trillion market opportunity for cybersecurity technology and service providers. *Mckinsey*, 27 October.

Munich Re, 2022. *Munich Re Global Cyber Risk and Insurance Survey 2022,* Munich: Munich Re. Available at: https://www.munichre.com/landingpage/en/global-cyber-risk-and-insurance-survey-2022.html [Accessed 17 November 2023]

One Identity, 2024. *What is attack surface expansion?* [Online] Available at: https://www.oneidentity.com/learn/what-is-attack-surface-expansion.aspx#:~:text=An%20attack%20surface%20is%20the,complexity%20of%20these%20entry%20points [Accessed 20 February 2024].

Richardson, R. and M. North 2017. Ransomware: Evolution, Mitigation, and Prevention. Georgia: *International Management Review,* 13(1)

Shi, C. and A. McNestrie 2023. Munich Re takes hard line on narrower cyber war exclusions. *Insurance Insider,* 19 May.

Tilley, H. and L. Poulsen 2023. *Cyber Attack Not Within War Exclusion.* London: Carter Perry Bailey.

Tilley, H., S. Zaozirny and S. Carter 2022. *It's war but not as we know it?* London: Carter Perry Bailey.

Trend Micro. (2024, March 19). Ransomware. Retrieved from Trend Micro: https://www.trendmicro.com/vinfo/us/security/definition/ransomware

Wolff, J. (2022). *Cyber-Insurance policy: Rethinking international risk for the Internet age*. Cambridge: MIT Press.

Woods, D. and A. Simpson 2017. Policy measures and cyber insurance: a framework. *Journal of Cyber Policy,* 2(2).

Yulchon LLC, 2019. Amendments to the Network Act Coming into Effect in 2019. *Lexology*. [Online] Available at: https://www.lexology.com/library/detail.aspx?g=fa96ac52-003b-4ec9-92b0-22fd0e8c1192 [Accessed 18 November 2023].

About the authors

Serene Chan leads Munich Re's Cyber service offering in the Asia-Pacific region, being one of the founding members of the local set-up since 2018. Prior to joining Munich Re, Serene was working on the primary side at Lloyd's of London, underwriting large US corporate cyber and intellectual property business. Serene is a Barrister-at-law of England and Wales, and is a member of the Lincoln's Inn of London. She is originally from Malaysia and moved to the UK to complete her A-levels, after which she graduated from the King's College London School of Law.

Eric Cho is a Senior Cyber Underwriter at Munich Re, currently based in the Tokyo office. He joined Munich Re in 2020 in the Singapore office, and also spent a year in the South Korea office. Prior to Munich Re, he worked at AIG in Canada, underwriting financial lines for the Western Canada region. Eric graduated from the University of British Columbia with a Bachelor of Commerce (Honors), specialising in accounting and operations and logistics.

110 \ Commonwealth Cyber Journal
ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

The Commonwealth

The National Security Exception in International Trade and Cybersecurity

Kartikeya Garg[1]

Abstract

The national security exception is a crucial part of most trade agreements, and has historically protected against traditional attacks to states' security. However, over time, threats to national security have evolved, to include, among other things, cyberthreats. Article XXI of the General Agreement on Tariffs and Trade, first interpreted by the World Trade Organization's *Russia–Traffic in Transit Panel Report* in 2019, has led to many debates regarding the scope of the provision and its application to cyberthreats.

This article discusses the plausibility of including state-imposed cybersecurity measures within the ambit of this interpretation of Article XXI. Apart from this provision, various Free Trade Agreements (FTAs) have incorporated different formulations of the security exception. This article analyses four main formulations of the exception in various FTAs and discusses the best option to include cybersecurity measures within their ambit, without leaving too broad of a scope to allow misuse. It recommends a balance between an explicit reference to cybersecurity measures and an emphasis on the principle of good faith as an effective check to ensure the balance between trade and cybersecurity is met.

Keywords: national security, Article XXI, cybersecurity, exception

Introduction: Trade, national security and cybersecurity

National security exceptions have formed a crucial part of multilateral trade agreements since the advent of international trade regulation. Given the sensitive nature of the issues at hand, states have rightly valued the security of their territory and their citizens

1 International Trade Policy section, Trade, Oceans and Natural Resources Directorate, Commonwealth Secretariat. Email: k.garg@commonwealth.int; kartikeya.garg@graduateinstitute.ch.

more than the economic benefits that they might derive from international trade.[2] It is for this reason that national security exceptions have been codified not just in World Trade Organization (WTO) Agreements but also in more than 290 regional and preferential trade agreements.[3] However, a perusal of these national security exceptions shows that they were originally drafted taking into consideration only military threats to national security.

In the past two decades, however, cybersecurity has become an increasingly important component of national security,[4] which, according to various lawmakers, requires special policy considerations. These rules in cybersecurity, mostly formulated as a part of domestic policy, cover, among other things, the restriction of cross-border data flows, transactions involving sensitive personal data and the imposition of technical and regulatory standards.[5] However, crucially, despite these innovative rules, there is still considerable ambiguity at the multilateral level on the question of *whether and when cybersecurity concerns should override the demands of trade liberalisation (or vice versa)*.[6]

The global expansion of the internet and increased data flows between businesses and consumers around the world, for e-commerce, for communication and as a source of information, has meant that international trade and cybersecurity are becoming increasingly intertwined. Digital connectivity has therefore transformed international trade and accelerated the global connectivity of businesses, governments and cross-border supply chains.

The interaction between trade and cybersecurity measures can take various forms, with the trading regime having to deal with various kinds of cyberthreats, each suggesting a different response.[7] Broadly, these threats have been classified to include the following: intrusions into military or defence systems; cyberattacks on critical public infrastructure; economic cyber-espionage aimed at stealing intellectual property (IP) and trade secrets; and the manipulation of digital information to create distrust.[8] Also, many cybersecurity measures are likely to restrict cross-border data flows and digital trade, including through data localisation requirements and import restrictions on data and digital products.

While threats to national security have been evolving over time, the national security exception has, for almost 70 years, been treated as a Pandora's box, owing to the highly

2 Van den Bossche, P.L.H. (2020) 'The National Security Exception in International Trade Law Today: Can We Avoid Abuse?', in *Association 'Commercial Law', Pre-Advice 2020: Review of Foreign Investment in Geopolitical and Legal Perspective*, pp. 111–143.

3 Dür, A., Baccini, L. and Elsig, M. (2014) 'The Design of International Trade Agreements: Introducing a New Database', *Review of International Organizations* 9(3): pp. 353–375.

4 Government of the United States of America (2015) *National Security Strategy.*

5 OECD (2009) *Security-Related Terms in International Investment Law and in National Security Strategies*. Paris: OECD.

6 Benton Heath, J. (2020) '*The New National Security Challenge to the Economic Order'*, Yale Law Journal 129: pp. 1020–1098.

7 Ibid.

8 Government of the People's Republic of China (2015) *National Security Law of the People's Republic of China.*

sensitive nature of the provision and the immense scope for misuse.[9] According to various trade officials and policy-makers, states can use the provision as an easy outlet to flout their trade obligations and protect their domestic industry, which could be immensely harmful to the multilateral trading system.[10] At the same time, the provision must not be so rigid as to prevent states from imposing measures to protect their national security from evolving threats, such as cyberattacks.

These are thus two sides of the same coin: while expanding the scope of the national security exception could lead to widespread misuse and hamper international trade, narrowing the scope of the provision to exclude its application for evolving threats to national security could prove highly disadvantageous to states, and would ultimately go against the very purpose of the provision. It thus becomes vital to determine whether such a balance between trade and evolving national security threats exists – and, if not, if it *can* exist, in order to create a situation where states can use the national security exception to implement measures to protect against cyberthreats but at the same time ensure that misuse is prevented.

In this regard, this article seeks to assess the various formulations of the national security exception in international trade law and attempts to determine whether any of these formulations achieve the cybersecurity/trade balance discussed above. First, the article analyses the national security exception under the WTO and attempts to determine whether it can be interpreted broadly to include measures to protect against cyberthreats. It then identifies four different formulations of the exception in various modern Free Trade Agreements (FTAs) and discusses whether any of these formulations are better suited to include evolving threats to national security within their ambit. The article concludes by recommending the formulation that is the most appropriate to combat cyberthreats and suggests a change in its interpretation to prevent misuse.

1. The national security exception under the WTO

The national security exception under the WTO is codified under Article XXI of the General Agreement on Tariffs and Trade (GATT). This provision is also found under Article XIV bis and Article 73, respectively, of the General Agreement on Trade in Services and the Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPS) *mutatis mutandis.*

The provision reads:

> *Nothing in this Agreement shall be construed*
>
> *(a) to require any contracting party to furnish any information the disclosure of which it considers contrary to its essential security interests; or*

9 Van den Bossche (2020), p. 114.
10 Ibid., p. 115.

(b) *to prevent any contracting party from taking any action which it considers necessary for the protection of its essential security interests*

(i) *relating to fissionable materials or the materials from which they are derived;*

(ii) *relating to the traffic in arms, ammunition and implements of war and to such traffic in other goods and materials as is carried on directly or indirectly for the purpose of supplying a military establishment;*

(iii) *taken in time of war or other emergency in international relations; or*

(c) *to prevent any contracting party from taking any action in pursuance of its obligations under the United Nations Charter for the maintenance of international peace and security.*[11] (Emphasis added.)

Among these clauses, subparagraphs (a), (b)(i) and (c) have never been invoked and challenged before the WTO,[12] while subparagraph (b)(ii) has been invoked only once, in the first GATT case[13] dealing with national security exceptions. States therefore rely solely on Article XXI subparagraph (b)(iii), in order to take advantage of its '*controversial and ambiguous wording*.'[14]

1.1 The current interpretation of Article XXI according to the WTO Panel

Article XXI(b)(iii) was for the first time comprehensively analysed by the WTO Panel in the *Russia–Traffic in Transit* case.[15] It was subsequently referred to in a case brought about under Article 73 of the TRIPS, in *Saudi Arabia–IP*.[16] The Panel's landmark interpretation in *Russia–Traffic in Transit* has been widely debated, and has also opened the door for many more cases raising the defence of Article XXI. The Panel analysed the provision in two main parts: (i) determining whether Article XXI(b) is self-judging or not and (ii) understanding the scope of Article XXI(b) as well as Article XXI(b)(iii). This article does not focus on the Panel's interpretation regarding the self-judging nature of the provision, and also does not discuss the facts of these cases. Rather, it deals only with the interpretation of the terms of the chapeau of Article XXI(b), as well as the specific conditions laid down under clause (iii).

11 GATT Art. XXI, 1994.
12 Yoo, J.Y. and Ahn, D. (2016) 'Security Exceptions in the WTO System: Bridge or Bottle-Neck for Trade and Security?' *Journal of International Economic Law* 19(2): pp. 417–444.
13 Article XXI – United States Export Restrictions GD/4, Decision of 8 June 1949.
14 Yoo and Ahn (2016), p. 427.
15 WTO (2019) 'Panel Report, Russia – Measures Concerning Traffic in Transit'. WTO Doc. WT/DS512/R, 5 April.
16 WTO (2020) 'Panel Report, Saudi Arabia – Measures Concerning the Protection of Intellectual Property Rights'. WTO Doc. WT/DS567/R, 16 June.

1. *The chapeau of Article XXI(b): 'any action which it considers necessary for the protection of its essential security interests'*

The first question before the Panel in this regard was whether the phrase '*which it considers*' allows Members to determine on their own their essential security interests as well as the necessity of the measures to protect them; or only the necessity of the measures.[17] Russia argued that the entire provision was self-judging and that both determinations were left entirely to the discretion of the Member, whereas Ukraine contended that it was for the Panel to interpret '*essential security interests*' while applying customary treaty interpretation rules under public international law.[18]

The Panel agreed with the proposed interpretation of Ukraine and attempted to define '*essential security interests*.' Differentiating the term from '*security interests*,' it explicitly qualified the term to mean those relating to *'quintessential functions of the state, namely, the protection from external threats, and the maintenance of law and public order internally.'*[19] It thus expressly provided two functions of the state, the protection of which would fall under the ambit of '*essential*' under Article XXI(b). Despite this rather specific interpretation, the Panel did consider that specific interests that Members sought to protect under this provision would vary depending on situations and changing circumstances.[20]

Therefore, despite initially defining the term rather narrowly, the Panel conferred some discretion to Members to choose what constitutes an essential security interest. However, this discretion was further qualified by the Panel as it obliged Members to exercise this determination keeping in mind the general principle of good faith.

2. *The inherent good faith obligation on Members*

The Panel interpreted that Members had an inherent obligation of good faith when declaring what constituted an essential security interest that needed protection under Article XXI. This obligation is two-fold: (i) Members should not use the provision as a disguise in order to use increasingly protectionist measures; and (ii) there should be a logical link between the essential security interest and the measure imposed by the Member.

The Panel illustrated the first good faith obligation by describing a situation where the invoking Member would seek to evade obligations inherent in the multilateral trading regime, built on principles of non-discrimination, by merely classifying trade interests as '*essential security interests*' falling within the exception. In order to prevent Members from using Article XXI to circumvent obligations under the GATT, the Panel required Members to articulate their essential security interests in such a way that they are '*sufficiently enough to demonstrate their veracity.*'[21]

17 WTO (2019) 'Panel Report, Russia', para. 7.128.
18 Ibid., para. 7.129.
19 Ibid., para. 7.130.
20 Ibid., para. 7.131.
21 Ibid., para. 7.134.

The Panel went on to provide guidelines on what could constitute such 'sufficiency' in articulation by a Member. This would depend on the nature of the '*emergency in international relations*' requiring the imposition of the measure. For more serious emergencies in international relations, the requirement for the Member to sufficiently articulate is less stringent, since essential security interests in these cases would be far more evident. However, for emergencies that are less serious – that is, where defence or military interests, or maintenance of law and public order interests, are not as evident – the obligation to sufficiently articulate to the Panel is enhanced.[22]

The second good faith obligation requires that Members clearly establish the connection between the essential security interest they seek to be protected and the measure actually imposed. In other words, the measure must '*meet a minimum requirement of plausibility in relation to the proffered essential security interests, i.e., that they are not implausible as measures protective of these interests.*'[23] As per the facts of the case, since all measures that Russia had imposed attempted to prevent the transit of goods from the Ukraine–Russia border, and considering that there was a situation of armed conflict between the two countries as recognised by the United Nations General Assembly,[24] the Panel concluded that the measures met the minimum requirement of plausibility.

3. Article XXI(b)(iii): In time of war or other emergency in international relations

The Panel defines '*emergency in international relations*' to include four specific situations: armed conflict; latent armed conflict; a heightened tension or crisis; or of general instability engulfing or surrounding a state.[25] Mere political and economic differences are insufficient to constitute such an emergency.[26] Accordingly, Russia identified various factors that proved that there did in fact exist such an emergency between itself and Ukraine: (i) that the time period during which the emergency arose continued to exist; (ii) that Ukraine was involved; (iii) that it affected the security of the Russia–Ukraine border; (iv) that other countries had imposed sanctions on Russia resultantly; and (v) that the entire situation was publicly known.[27]

The Panel deemed these reasons sufficient to enable a conclusion that there did exist a situation of emergency in international relations in the case. The Panel interpreted the term '*war*' to mean '*armed attack*'[28] and did not analyse it further, presumably because the WTO was not designed as a body for the resolution of conflicts such as wars, insurrections and unrests.[29]

22 Ibid., para. 7.135.
23 Ibid., para. 7.138.
24 Ibid., para. 7.145.
25 Ibid., para. 7.111.
26 Ibid., para. 7.74.
27 Ibid., para. 7.119.
28 Ibid.
29 Ibid., para. 7.112.

1.2 The current interpretation being sufficient to combat evolving threats to national security?

The term '*essential security interests*' was the subject of intense debate during the Preparatory Sessions of the GATT. The purpose of Article XXI, according to the drafters, was to create a balancing act between genuine security interests that warrant protection and to prevent Members from adopting excessively protectionist measures.[30] The only way this balance and the prevention of abuse can be guaranteed is through the '*spirit in which Members would interpret these provisions*.'[31] The Panel attempted to reach this balance by amalgamating the '*deferent standard of review*' provided by the chapeau of Article XXI(b) and the stricter '*objective analysis*' envisaged by subparagraphs (i) to (iii).[32] Thus, although states are allowed to determine their own national security interests as they deem fit, such absolute deference would prevent states from justifying or notifying the imposition of any such trade-restrictive national security measure,[33] and would seriously impair the objectives sought to be achieved by the global trade system.[34]

The Panel thus concluded that most elements of Article XXI(b) required an objective standard of review,[35] including an assessment on whether a measure first concerns an essential security interest and then falls under one of the specified subparagraphs. However, this would generally not be a problem for panels, because traditional security issues would be objectively identifiable. It would be very difficult for governments to disguise protectionist measures as claims of national security.[36]

This position changes, however, when we discuss evolving threats to national security, such as claims regarding cybersecurity. Since cybersecurity policies are more risk-based and generally require long-term adoption,[37] it is necessary to ascertain whether the current interpretation of Article XXI(b) allows for states to take measures they deem necessary to protect themselves against evolving cybersecurity threats, or whether it restricts the scope of the provision to only traditional notions of security. The current interpretation of the provision from the Panel illustrates two possible options.

30 WTO Article XXI: Security Exceptions.
31 Ibid.
32 Blanco, S. and Pehl, A. (2020) *National Security Exceptions in International Trade and Investment Agreements: Justiciability and Standards of Review*. Springer, p. 24.
33 Bhala, R. (1998) *International Trade Law: Theory and Practice*. Second Edition. New York: LexisNexis.
34 WTO (nd) 'Principles of the Trading System'. www.wto.org/english/thewto_e/whatis_e/tif_e/fact2_e.htm
35 Blanco and Pehl (2020).
36 Meltzer, J.P. (2020) *Cybersecurity, Digital Trade and Data Flows: Re-thinking a Role for International Trade Rules*. Working Paper 123. Washington, DC: The Brookings Institution (2020).
37 Ibid.

1. *Cybersecurity measures falling within the ambit of Article XXI(b)?*

Because of the changing character of threats to peace and security and increasing issues of cybersecurity,[38] it becomes crucial to determine whether a dynamic interpretative approach could be used to allow such evolving notions of security threats into the ambit of Article XXI. Interpreting the WTO provisions in light of these contemporary developments, however, does not '*subsume completely different or novel meanings and concepts under their notions*.'[39] This dynamic interpretation is in contrast to an '*overall expansive approach that would subsume all sorts of novel security allegations under Article XXI, irrespective of its boundaries*.'[40]

Although the Panel determined that Article XXI was not wholly self-judging,[41] it still gave states the freedom to articulate their own '*essential security interest'* and noted that, as long as states could sufficiently link this interest to the adopted measure, panels would not hesitate to allow the application of the provision.[42] However, establishing this 'plausible link' may be difficult for states in practice since security vulnerabilities in digital systems are still not known.[43] Nonetheless, if it can be demonstrated, it may be possible for states to avail the exception under any of the three subparagraphs under Article XXI(b).

States have been increasingly digitising their militaries, making them more vulnerable to cyberattacks.[44] Members could justify the imposition of cybersecurity measures to prevent cyberthreats to their nuclear or military facilities under Articles XII(b)(i) and (ii).[45] Since military threats have themselves evolved, the concept of arms used and potentially covered by Article XXI(b)(ii) could also evolve to include measures as necessary for such conflicts.[46] Since combating cyberwarfare would require different goods, it can be argued that Article (b)(ii) could be interpreted in light of these contemporary national security concerns.

However, the exception most suited to expanding the provision to cybersecurity measures is Article (b)(iii). This view emphasises the enhanced discretion that states are conferred under the exception, holding that it provides '*almost no limits*' on governments in justifying

38 Weiß, W. (2008) 'Security Council Powers and the Exigencies of Justice after War'. *Max Planck Yearbook of UN Law*, 1 January.
39 Weiß, W. (2020) 'Interpreting Essential Security Exceptions in WTO Law in View of Economic Security Interests'. *Global Politics and EU Trade Policy, European Yearbook of International Economic Law*, p. 267.
40 Ibid.
41 WTO (2019) 'Panel Report, Russia'.
42 Meltzer (2020).
43 Mishra, N. (2020) 'The Trade: (Cyber)Security Dilemma and Its Impact on Global Cybersecurity Governance', *Journal of World Trade* 54(4): 567–590.
44 Delcker, J. (2027) 'Digitizing Military Will Cost Europe Up to €41 Billion Per Year: Study'. Politico, 23 November. www.politico.eu/article/digitizing-military-will-cost-europe-up-to-e41-billion-per-year-study/
45 Mishra (2020).
46 Weiß (2008).

their security measure.[47] This discretion, coupled with the fact that the Panel interpreted '*emergency in international relations*' expansively to include not just defence and military concerns but also the '*maintenance of law and public order,*'[48] implies that cybersecurity measures could, in fact, be included within the scope of Article XXI(b)(iii). In other words, as long as Members are able to sufficiently articulate their '*essential security interests*' and the 'plausible link' to the cybersecurity measure, panels could be willing to accept the claim that cybersecurity measures are necessary to protect their essential security interests.[49]

For example, a state may impose a ban on foreign digital services during an armed attack in order to minimise risks of cyberattacks.[50] Once a state shows that it does in fact face an emergency in international relations, it can argue that cybersecurity measures were imposed to prevent cyberthreats to critical infrastructure underlying public utility claims,[51] which constitutes an 'essential security interest.' If this measure is analysed using the lens of this broad interpretation, panels might be inclined to allow it as a national security exception under Article XXI(b)(iii).

As the Panel concluded in *Russia–Traffic in Transit*, the onus is on the invoking state to sufficiently articulate its essential security interests. However, this articulation becomes problematic if the state invokes Article XXI(a), since this allows states to refrain from providing information that could hamper its essential security interests. The absence of subparagraphs and qualifiers for this clause implies that it is of a more self-judging nature than Article XXI(b). Therefore, if a state uses Article XXI(a) to justify its non-articulation of its essential security interest, the Panel would not be able to do much in response.[52] This is even more the case if the information that ought to be disclosed is highly confidential, or is limited, which might be the case for an emergency relating to cybersecurity.[53] Thus, states could argue that, by invoking Article XXI(a), they would not be required to provide substantive evidence regarding the 'plausible link' between the cybersecurity measure and its essential security interest, which could relate to the protection of its cyberinfrastructure.[54]

Additionally, the temporal requirement that the measure be taken '*in time of*' an emergency in international relations could potentially extend to '*time-unlimited cybersecurity measures,*'[55] given the uncertain nature of cyberattacks. In this situation, scope for misuse by Members is tackled by the good faith obligation inherent in Article XXI.

47 Benton Heath (2020).
48 WTO (2019) 'Panel Report, Russia'.
49 Meltzer (2020).
50 Mishra (2020).
51 Ibid.
52 Van den Bossche (2019).
53 Ibid.
54 Voon, T. and Mitchell, A. (2019) 'Australia's Huawei Ban Raises Difficult Questions for the WTO'. EastAsia Forum, 22 April. https://eastasiaforum.org/2019/04/22/australias-huawei-ban-raises-difficult-questions-for-the-wto/
55 Benton Heath (2020).

These are thus the various arguments raised in favour of a broad reading of Article XXI to include measures taken to protect cybersecurity based on the interpretation of the Panel.

2. Cybersecurity measures falling outside the scope of Article XXI(b)?

The first problem created by using a broader understanding of the Article XXI interpretation is that it would contradict the intention of the drafters of the provision. In order to prevent the use of the exception to '*permit anything under the sun*,'[56] it was reasoned that the provision should be drafted in a way that '*would take care of real essential security interests and, at the same time, so far as we could, limit the exception so as to prevent the adoption of protection of industries under every conceivable circumstance*.'[57] The wording of Article XXI, as well as the Panel's interpretation that the provision is not entirely self-judging and is therefore subject to an objective standard of review, implies that a broad interpretation is not possible, and it is difficult to include non-traditional notions of security under the provision.

This is more the case given the Panel's definitions of '*war*' and '*emergency in international relations,*' where it discusses traditional notions of security threats that are easily identifiable and objectively discernible. However, when it comes to cyberwarfare or cyberthreats, this is not the case. Thus, even though a Member could classify a cybersecurity measure as one protecting an '*essential security interest,*' it will be difficult for panels to assess when the cyberthreat is grave enough to justify a measure under Article XXI.[58] Further, given that cyberthreats can be characterised into various kinds of security threats (military, political or even commercial), panels would require substantial evidence to assess whether the cybersecurity measure is actually imposed in order to contain an 'emergency in international relations' under Article XXI(b)(iii). For example, a systematic theft of trade secrets of digital companies would not, according to the Panel's interpretation in *Russia–Traffic in Transit*, constitute a situation of '*war*' or '*emergency in international relations.*'[59]

This difficulty also exists when assessing an imminent cyberthreat, whereby panels would have to determine whether the cyberthreat constituted as much of a threat as an armed attack. This is because the industry surrounding cyberwarfare and cyberweapons is highly dynamic and uncertain, making it impossible to predict the nature and intensity of cyberthreats.[60] Therefore, if Members, while acting in good faith, believe there exists a legitimate basis for imposing cybersecurity measures, it will be difficult to prove to panels the necessity of such measures owing to a lack of evidence.[61]

56 Weiß (2008).
57 Ibid.
58 Mishra (2020).
59 Ibid.
60 Mishra (2020).
61 Ibid.

The biggest problem with the inclusion of cybersecurity measures within Article XXI relates to the Panel's interpretation of the temporal requirement provided under Article XXI(b)(iii). By requiring that the measure be taken during the time of the emergency in international relations, the Panel seemed to exclude many 'risk-based' cybersecurity measures from its scope. Since cyberthreats are of such a nature that they may arise from any country with an internet connection, the nature of the risk is such that the only way it can be truly neutralised is if states adopt continuous cybersecurity measures, irrespective of the existence of an emergency in international relations or not.[62]

Therefore, although in theory Article XXI(b)(iii) could be applied to cybersecurity measures that are taken during the specified period of time of 'war or emergency in international relations,' in practice, since cyber-related emergencies are permanent and long in term rather than timebound, and given that they can originate anywhere, it would be rare for such measures to fall under the ambit of Article XXI.[63]

Thus, although Article XXI could possibly be interpreted broadly to include cybersecurity measures, this claim would be difficult to prove, for the reasons explained above. Nonetheless, both the narrow and the broad interpretations lead to the same outcome – that is, a *'lack of an effective governance mechanism to mediate cybersecurity/trade tradeoffs*.'[64]

2. Alternate formulations of the national security exception in FTAs

Despite the various economic and technological developments that have led to evolving threats to national security, there have been no amendments to the text of Article XXI since it was first formulated in 1947.[65] Even in the post-WTO era, with the increase in the number of FTAs between states, little or no attention has been paid to the national security exception. In fact, multiple FTAs do not even contain security exceptions.[66]

This article identifies four different kinds of formulations of the national security exception commonly found in FTAs, and attempts to analyse whether any of these variations are better equipped to cover cybersecurity measures imposed by states. These formulations include the incorporation of Article XXI GATT **(A)**; incorporating the national security exception as part of the General Exceptions **(B)**; formulating a wholly 'self-judging' national security exception **(C)**; and an explicit clause allowing for the imposition of certain cybersecurity measures **(D)**.

62 Meltzer (2020).
63 Benton Heath (2020).
64 Meltzer (2020).
65 Yoo and Ahn (2016).
66 Korea–EU Free Trade Agreement; Korea–India Free Trade Agreement.

2.1 The incorporation of Article XXI GATT

Despite the various debates concerning the interpretation of Article XXI as discussed in the previous section, this formulation has been the most prevalent form of security exceptions in FTAs.[67] This formulation can take two forms.

1. *Directly transposing Article XXI into FTAs*

Many FTAs incorporate the provision as it stands, without any modification;[68] however, a few make minor changes to the language of the provision, especially with respect to that in Article XXI(b)(iii). For instance, many FTAs replace the terms 'in time of war or emergency in international relations' with terms such as '*serious internal disturbances affecting the maintenance of law and order, in time of war or serious international tension constituting threat of war*.'[69] In some cases, the term 'emergency in international relations' is replaced by '*serious international tension*'[70] or '*extraordinary circumstances in international relations*.'[71]

This can be interpreted as either broadening or widening the scope of the provision based on the same interpretation given by the Panel in Russia–Traffic in Transit. In other words, this formulation faces the same issues that Article XXI does. It allows states the discretion to determine what constitutes an essential security interest and sufficiently articulate that the measure has been taken in good faith to protect the same. According to the dictionary, the term '*emergency*' refers to a '*serious situation that occurs suddenly or unexpectedly and requires urgent attention*,'[72] which has a much narrower scope than the term '*serious.*'

Thus, it would seem more likely that states will justify the imposition of cybersecurity measures under the broader ambit of '*serious international tension.*' However, despite this possible additional leeway the provision gives to include measures other than those to protect traditional security concerns, the temporal requirement remains the same as under Article XXI(b)(iii), which severely restricts the scope of the provision to include long-term cybersecurity measures.

2. *Reference to abide by GATT provisions*

Multiple FTAs contain provisions that do not explicitly mention a national security exception; rather, they refer directly to the provisions of the GATT. For instance, they can take the form of the following:

67 Yoo and Ahn (2016).
68 EU–UK Trade and Cooperation Agreement 2020, Article EXC:4; Agreement Establishing the African Continental Free Trade Area 2018, Article 27.
69 Agreement on the European Economic Area 1994, Article 123.
70 Economic Cooperation Organization Trade Agreement 2008, Article 15/B.
71 Free Trade Agreement between Azerbaijan, Armenia, Belarus, Georgia, Moldova, Kazakhstan, The Russian Federation, Ukraine, Uzbekistan, Tajikistan and the Kyrgyz Republic Agreement on the Creation of a Free-Trade Area, Exceptions for the Reasons of Safety 1994.
72 Cambridge Dictionary, Fourth Edition.

> *For the purposes of this Chapter, the rights and obligations of the Parties in respect of Security Exceptions shall be governed by Article XXI of the GATT 1994, which is hereby incorporated into and made part of this Agreement, mutatis mutandis.*[73] (Emphasis added.)

This sort of provision prevents any further interpretation of the national security exception under the FTA, since it allows for the application of Article XXI as it stands. It therefore directly binds states under the interpretation of the provision given by the Panel in *Russia–Traffic in Transit*. Reference to GATT national security exceptions in this manner can be either explicit, as in the formulation mentioned above, specifying the *mutatis mutandis* incorporation of Article XXI, or implied, in the following way:

> *No provision in this Agreement shall be interpreted to prevent either Party from adopting or maintaining exception measures consistent with the rules of the World Trade Organization.*[74](Emphasis added.)

This leaves Members with no room to interpret Article XXI differently and expansively to include cybersecurity measures. They must satisfy the same conditions that the Panel imposed.

2.2 Incorporating the national security exception as a part of the General Exceptions

The first drafts of the International Trade Organization Charter did not contain a separate provision for security exceptions, with the current set of exceptions under Article XXI being listed as separate items in parallel with other clauses of the General Exceptions provision.[75] It was only much later, at the final stage of negotiations in 1947, that the security provisions were divided completely from the General Exceptions.[76] Despite this separation, various FTAs have incorporated the national security exception as one of the clauses in its General Exceptions provision, which often mirrors Article XX GATT. These can take the form of the following:

> *Member States during the mutual trade of goods may apply restrictions (subject to the fact that these measures do not serve as unjustifiable discrimination or covered restriction on trade), if such restrictions are necessary for: 1) protection of human life and health;... 6) the defense and security of the Member state.*[77] (Emphasis added.)

The incorporation of security exceptions into General Exceptions chapters could either be in the form of a general reference to 'security of a state' as mentioned above or

73 EU-Peru Free Trade Agreement 2010, Article 2.20.
74 Cross-Straits Economic Cooperation Framework Agreement 2010, Article 9.
75 Yoo and Ahn (2016).
76 Ibid.
77 Treaty on the Eurasian Economic Union 2015, Article 29.

be specific to certain situations for which cybersecurity measures would be allowed. For example:

> *For the purposes of the following chapters… subject to the requirement that measures are not applied in a manner that would constitute a means of arbitrary or unjustifiable discrimination between parties… or a disguised restriction on trade…, nothing in this Agreement shall be construed to prevent the adoption or enforcement by a Party of measures necessary: a) to protect public security or public morals or maintain public order;… c) to secure compliance with laws or regulations which are not inconsistent with provisions of this Agreement including those relating to:… ii) the protection of the privacy of individuals in relation to the processing and dissemination of personal data and the protection of confidentiality of individual records and accounts;…*[78] (Emphasis added.)

Incorporating the national security exemptions into the General Exceptions could serve to be advantageous to Members because WTO panels have in various instances already given Article XX clauses extremely broad interpretations, taking into consideration contemporary circumstances to include within its ambit different sets of domestic policy objectives.[79] For example, at the time of drafting Article XX(g), the term 'exhaustible natural resources' was thought to be limited to 'stock resources of raw materials or minerals.'[80] The Panel, however, interpreted the term broadly to also include 'fresh air or endangered species.'[81]

This was also the case in *US–Gambling*, where the Panel found that a regulation preventing underage gambling fell within the scope of Article XX(a) as a measure to protect public morals or public order.[82] In the same case, it was decided that Members were empowered under the provision to decide for themselves the content of 'public morals' and 'public order' according to '*their own systems and scales of values*' and that this would vary according to the '*prevailing social, cultural, ethical and religious*' values and concepts of Members.[83]

These examples are encouraging for Members in the sense that they can adopt cybersecurity policies in light of the security exception clause provided for under the General Exceptions, with considerably less backlash from dispute settlement bodies. This is because, although the clauses of Article XX are often interpreted broadly, its chapeau is interpreted very narrowly, in the form of a final litmus test,[84] in order to prevent misuse

78 EU–Canada Comprehensive Economic and Trade Agreement 2014, Article 28.3.
79 Weiß (2008).
80 WTO (2001) 'Appellate Body Report: United States–Import Prohibition of Certain Shrimp and Shrimp Products'. WTO Doc. WT/DS58/23, 21 November.
81 Ibid.
82 WTO (2015) 'Appellate Body Report: United States–Measures Affecting the Cross-Border Supply of Gambling and Betting Services'. WTO Doc. WT/DS285/AB/R, para. 299, 20 April.
83 WTO (2006) 'Panel Report: United States–Measures Affecting the Cross-Border Supply of Gambling and Betting Services'. WTO Doc. WT/DS285/R, para. 6.461, 20 April.
84 Weiß, W. (2019) *WTO Law and Domestic Regulation*. Beck Hart Nomos Publishing.

of the provisions.[85] In other words, the expansive interpretation of the clauses and the limiting effect intended by the chapeau must be seen as 'counteracting movements' and must be analysed together to enable a thorough understanding of the balanced approach sought by the provision to allow Members to pursue domestic policies.[86]

It is precisely this point that differentiates Articles XX and XXI, and this may also be the reason states choose to include national security measures as part of the General Exceptions. Article XXI does not contain a chapeau like in Article XX that can act as a balance to wider interpretations that may be conferred to the clauses under it. Therefore, widely expansive constructions to clauses under Article XXI in the same manner as in Article XX would seem unlikely, as this would lead to immense misuse.[87] Thus, *prima facie*, it seems that this formulation could be the most beneficial for states in implementing cybersecurity measures in order to combat evolving threats to national security.

However, including national security exceptions as part of General Exceptions can be problematic for various reasons. First, this formulation goes against the very reason General Exceptions and security exceptions were split into two different provisions. It was reasoned at the time that the chapeau of Article XX was too constraining, and that, for emergency measures to be taken during military conflicts, a wider basis would be needed.[88] However, at the same time, the formulations in these FTAs give states too much discretion, taking away the qualifiers specified under Article XXI, such as the need to articulate an 'essential security interest,' as well as the conditions specified under Article XXI(b). Having such a broad formulation removes the balance that the framers of the GATT sought, and leaves the provision open to immense abuse. This is also therefore not a very common formulation in FTAs.

2.3 A wholly 'self-judging' national security exception

The most common argument raised by Members invoking Article XXI is that the provision is entirely self-judging, and the WTO Panel therefore has no jurisdiction over it.[89] The Panel in *Russia–Traffic in Transit*, however, much to the discontent of invoking Members, ruled that the exception was not 'wholly self-judging'[90] and that, although the chapeau of Article XXI(b) allows for some measure of discretion, its subparagraphs acted as qualifiers. However, various FTAs, given the interests of many developed countries,

85 WTO (2001) 'Appellate Body Report: United States–Import Prohibition of Certain Shrimp and Shrimp Products'.
86 WTO (2015) 'Appellate Body Report: United States–Measures Affecting the Cross-Border Supply of Gambling and Betting Services'.
87 Weiß (2008).
88 Yoo and Ahn (2016).
89 U.S. First Written Submission 2019, 'US–Certain Measures on Steel and Aluminum Products' (DS548).
90 WTO (2019) 'Panel Report, Russia'.

intentionally broaden this discretion and arbitrariness of security exception clauses.[91] This could be in the form of the following:

> *Nothing in this Agreement shall be construed to:... (b) preclude a Party from applying measures that it considers necessary for the fulfilment of its obligations with respect to the maintenance or restoration of international peace or security, or the protection of its own essential security interests.*[92](Emphasis added.)

This formulation removes the qualifying conditions specified under Article XXI(b), limiting the provision to only its chapeau. As a result, it bypasses the Panel's interpretation and makes the provision wholly self-judging. In other words, as a result of this provision, states not only would be free to classify any interest they deem fit as an 'essential security interest' but also would have the discretion to adopt any measure they deem necessary in order to protect their interest. This unfettered discretion would naturally allow states to impose cybersecurity measures even if the term 'essential security interests' is conferred with the same meaning as that given by WTO panels.

It would be conceivable for states to argue that cybersecurity measures have been imposed for the protection of the state from 'external threats' as well as for the 'maintenance of law and public order.'[93] However, the absence of the qualifying subparagraphs mentioned in Article XXI(b) leaves states with a large amount of discretion and, consequently, the scope for misuse. A wholly self-judging provision also has the effect that an affected state would have no recourse to a dispute settlement body, since this latter would not have jurisdiction over this clause. This formulation therefore also fails to strike the balance between trade and evolving security concerns.

2.4 Explicit inclusion of cybersecurity measures

The last formulation in FTAs is a fairly recent phenomenon. As many states have begun to recognise the evolving nature of security concerns, they have understood the need to protect their interests from cyberthreats and have incorporated a specific clause in their security exception provisions to that effect. These provisions largely mirror Article XXI, merely adding an additional subparagraph under Article XXI(b). For instance, many FTAs incorporate the following clause:

> *Nothing in this Agreement shall be construed... (b) to prevent any Party from taking any action which it considers necessary for the protection of its essential security interests... (iii) taken so as to protect critical public infrastructures including communications, power, and water infrastructures...*[94](Emphasis added.)

91 Yoo and Ahn (2016).
92 US–Mexico–Canada Agreement 2020, Article 32.2; Dominican Republic–Central America Free Trade Agreement 2004, Article 21.1.
93 WTO (2019) 'Panel Report, Russia', para. 7.130.
94 Regional Comprehensive Economic Partnership 2020, Article 17.13.

This clause has also been drafted in the following way:

> *Nothing in this Agreement shall be construed… (b) to prevent any Party from taking any action which it considers necessary for the protection of its essential security interests… (iii) taken so as to protect critical public infrastructure including communications, power and water infrastructures from deliberate attempts intended to disable or degrade such infrastructures;…*[95] (Emphasis added.)

This formulation expands the scope of the national security exception by including within its ambit security threats other than traditional military threats. The term 'critical infrastructure' has been defined by various states as the '*physical and cyber systems and assets that are so vital that their incapacity or destruction would have a debilitating impact on physical and economic security or public health*'[96] and includes those functions, systems and processes necessary for a country and the daily lives of its people to function,[97] whether publicly or privately owned.[98] As the clause reads, such critical infrastructure includes (but is not limited to) power, communication and water infrastructure; it could also include infrastructure regarding finance, health, food and space.[99] This second formulation increases the scope of the provision by explicitly including 'attempted attacks' as well.

Thus, this formulation recognises not only that protection of these 'critical public infrastructures' can constitute an essential security interest for a state but also that, given the nature of such infrastructure, it can be prone to attacks other than traditional military action, such as cyberthreats. It therefore implies that states may be allowed to impose cybersecurity measures, provided they satisfy the other conditions laid out by the chapeau of the provision and its interpretation by the Panel. States would, therefore, have to articulate with sufficient clarity whether the 'critical public infrastructure' they are seeking to protect by means of the imposition of the cybersecurity measure does in fact constitute an 'essential security interest,' and then whether the measure is actually performing the function it set out to achieve.

An important critique on Article XXI(b)(iii) and why cybersecurity measures fall outside its scope has been the temporal requirement mandated by the provision, requiring that the measure be 'taken in time of war or emergency in international relations.' Since most cybersecurity measures are generally taken as long-term measures without any clear start or end date, it would be difficult for these measures to be valid under Article XXI. This formulation remedies this by removing the temporal requirement altogether. It does

95 Pacific Agreement on Closer Economic Relations Plus 2020, Article 2; Agreement Establishing the ASEAN–Australia–New Zealand Free Trade Area 2009, Chapter 15, Article 2.
96 U.S. Department of Homeland Security (nd) 'Critical Infrastructure Security and Resilience'. www.dhs.gov/topic/critical-infrastructure-security
97 Center for the Protection of National Infrastructure (2021) 'Critical National Infrastructure'. 20 April. www.cpni.gov.uk/critical-national-infrastructure-0
98 WTO (2019) 'Panel Report, Russia', para. 7.130.
99 U.S. Department of Homeland Security (nd) 'Critical Infrastructure Security and Resilience'.

not specify the duration that the measure must be taken for, with the only caveat being the principle of good faith.

Thus, at least *prima facie*, this formulation seems to address concerns of evolving threats of security by allowing states to impose cybersecurity measures. However, these formulations are rather recent, and not many FTAs have included these clauses in their FTAs. Based on this formulation, the US's measures against China's Huawei in early 2020 would, in theory, be accepted as a measure to protect national security but might not be so readily accepted under the Article XXI formulation.

Conclusion

This article has sought to explain how the multilateral trading regime has to equip itself to deal with completely different notions of security.[100] Among these evolving notions, cybersecurity is one of the most important and pressing issues, with the potential to severely hamper international trade. When the security exception was first formulated, the prevailing thought was that security objectives automatically outweighed the trade concerns of states. Since then, however, the scope of the trade and security nexus has widened,[101] and the security exception has failed to reflect this transition. This is evident in that the exception has not been amended since 1947. Two questions were raised at the start of this article: whether Article XXI as it stands is sufficient to combat the evolving threat of cyberwarfare; and, if not, whether any other formulations in FTAs provide a more viable alternative.

As explained earlier, given the dynamic and rather abstract nature of cybersecurity measures, it would be difficult for states, and consequently the Panel, to explain and approve the measure as protecting an essential security interest; and whether cyberwarfare actually constitutes a situation of 'war or emergency in international relations.' According to the Panel, such articulation would not be necessary if the emergency were objectively identifiable.[102] A step in this direction would be to increase the co-ordination between the WTO and the United Nations Security Council. Bhala recommends the establishment of a joint WTO–Security Council Committee, which could render non-binding opinions on whether the use of sanctions by the Security Council comports with the terms of Article XXI(b).[103]

Further, if there is a Security Council Resolution regarding the existence of a particular international situation, panels will be more likely to accept that this does in fact constitute an 'emergency in international relations.' Thus, the first step would be for the Security Council to recognise the credible threat cyberattacks pose, which it has done, by

100 Yoo and Ahn (2016).
101 Ibid.
102 WTO (2019) 'Panel Report, Russia'.
103 Bhala (1998).

realising the threat posed to states' critical infrastructure by possible cyberattacks from terrorists.[104] However, despite all of these steps, the risk-based long-term nature of cybersecurity measures means that it would be extremely difficult to implement them under Article XXI, thereby pointing to the need for reform and new formulations.

In this regard, we have found that modern FTAs have formulated the exception in four main ways: by making minor modifications to Article XXI; by incorporating the provision as it is; by incorporating it as a part of the General Exceptions provision; or by explicitly incorporating cybersecurity measures. Of these four kinds, the last option seems to be the most appropriate formulation in dealing with evolving concerns of national security in relation to trade. By explicitly allowing for the imposition of cybersecurity measures, it addresses all the concerns raised against the application of Article XXI to evolving security threats. This is why newer FTAs such as the Regional Comprehensive Economic Partnership have chosen to incorporate this form.

The caveat is that the only restriction available against the indiscriminate use of this new formulation is the principle of good faith.[105] The Panel has interpreted the good faith principle under Article XXI to include two specific obligations: sufficiently articulating the essential security interest that is meant to be protected; and demonstrating a plausible link between the measure and the interest. However, the good faith principle is applied differently in the application of Article XX. According to this principle, a state imposing a measure under Article XX must demonstrate that it has been imposed a last resort, and is the only way the specific interest can be protected.[106]

This is also reflected in various countries' domestic cybersecurity policies, recognising the need for closer co-operation, improving information exchange, optimising skills and promoting a common global approach to network and information security issues.[107]

A case can be made that the principle of good faith under Article XXI should also include this interpretation in order to ensure the discretion of states is limited. Since arguments have been made regarding the self-judging character of the provision, the only limitation to this is the principle of good faith, which, therefore, requires stronger clarity.[108] A more coherent understanding of the application of the good faith principle to the latest formulation allowing measures to be imposed to protect 'critical public infrastructure' is the best bet to ensure that the international trading regime can sufficiently combat evolving threats to national security.

104 United Nations Security Council Resolution 2341/2017.
105 Vienna Convention on Law of Treaties 1969, Article 31(1).
106 World Trade Organization (2001) 'Appellate Body Report: United States–Import Prohibition of Certain Shrimp and Shrimp Products'.
107 EU Cybersecurity Act (Regulation 2019/881) 2019; U.S. National Cybersecurity Strategy 2023.
108 Trujillo, E. (2020) 'An Introduction to Trade and National Security: New Concepts of National Security in a Time of Economic Uncertainty'. Symposium on Trade and National Security, 7 February.

About the author

Kartikeya Garg is an Assistant Research Officer working in the International Trade Policy section of the Trade, Oceans and Natural Resources Directorate within the Commonwealth Secretariat. He previously worked at the Trade and Environment Division of the World Trade Organization, and has an LLM in International Law from the Geneva Graduate Institute.

ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

The Commonwealth

Cybercrime in the Asia-Pacific Region: A Case Study of Commonwealth APAC Countries

Olajide O. Oyadeyi[1], Oluwadamilola Adeola Oyadeyi[2] and Rofiat Omolola Bello[3]

Abstract

The Asia-Pacific (APAC) region has witnessed a digital transformation in the past decade. There have been many factors behind this shift, including technological breakthroughs, heightened internet accessibility, evolving consumer patterns and efforts by governmental bodies and enterprises to embrace digital solutions. The region has also become a target for cybercrime as a result of its economic potential, expanding internet presence and comparatively inadequate levels of cyber-resilience. This article discusses the susceptibility of APAC to cyberattacks as well as the way recent events have exacerbated its vulnerability, leading to a need to enhance cyber-resilience within the region.

In response to such attacks, there has been a concerted emphasis on bolstering cybersecurity, fostering collaboration among law enforcement agencies and enacting regulatory measures to address cybercrime and related illicit online activity at both international and regional levels. This has entailed the co-operation of governmental bodies, law enforcement entities, financial institutions, technology corporations and international establishments in enhancing cybersecurity measures, exchanging information on potential threats and implementing more stringent regulations to reduce organised criminal activities in the era of digitalisation. Efforts to mitigate cyber-vulnerabilities are continuously evolving and may vary across different countries within the APAC region. In light of the dynamic nature of the threat landscape, ongoing collaboration and proactive actions to foster cyber-resilience and effectively combat cybercriminal activities are vital.

The relatively large surge in cybercriminal activities in APAC compared with other regions presents considerable obstacles for cybersecurity, law enforcement and the broader security environment in the region. There is a need for a

1 Imperial College Business School. Email: jide.oyadeyi@gmail.com
2 PhD candidate, University of Ibadan, Nigeria. Email: dami.oyadeyi@gmail.com.
3 Oxford Brookes University. Email: orb.bello@gmail.com.

comprehensive strategy encompassing the reinforcement of cybersecurity protocols, the augmentation of law enforcement capacities, the facilitation of international collaboration and the elevation of public consciousness regarding the perils associated with cybercrime.

1. Introduction

The Asia-Pacific (APAC) region has witnessed a digital transformation in the past decade. There have been a number of factors behind this shift, including technological breakthroughs, heightened internet accessibility, evolving consumer patterns and the efforts of governmental bodies and enterprises to embrace digital solutions. On the downside, these innovations have also played a key role in increasing cybercriminal activities in the region.

According to Cybersecurity Ventures (2023), the global cost of cybercrime may rise to US$9.5 trillion in 2024. If not appropriately mitigated, it may rise to 10.5 trillion by 2025 (Cybersecurity Ventures, 2022). If cybercrime were a country, it would have the world's third-largest economy behind the USA and China (ibid.). As cybercriminals do not publicise their operations, for obvious reasons, these estimates are based on recent trends and the exposure of corporations, governments, individuals and businesses to different kinds of cyberattacks.

The APAC region has not been spared from these attacks. The region has become an appealing target for cybercrime owing to its economic potential, expanding internet presence and comparatively inadequate levels of cyber-resilience. APAC was the most attacked of all regions in 2022, accounting for roughly 31 per cent of global cyberattacks (Positive Technologies, 2023). Furthermore, the first quarter of 2023 saw cybercriminal activities globally increase by a staggering 1,835 per cent year on year, with APAC bearing the brunt (Check Point Research, 2023). On average, there were 1,835 new cyber-assaults per organization every week in APAC, far above the global average of 1,248. As a result, the potential for cybercriminal activities in APAC is huge, with a potential cost of roughly US$3.3 trillion by 2025 if we take into account the 31 per cent of global cyberattacks attributed to the region and the potential cost of cybercrime of $10.5 trillion by 2025 (ibid.). Indeed, the APAC region has been called the new 'ground zero' for cybercriminal activities (Gullapali, 2023).

The connection between organised crime groups and cybercrime in the APAC region is complex. Organised crime attackers have been quick to recognise the potential for leveraging the digital domain to facilitate the expansion of their operations.

Questions arise about factors contributing to this increase in cyberattacks in the APAC region and the steps that must be taken to counteract the impending danger they

pose. Within this, it is of use to study recent cybersecurity attacks in APAC; we use the Commonwealth countries in the region to ascertain what can be done to mitigate the risks of attacks and improve cyber-resilience and recovery.

If these trends are not taken seriously, there may be further harm as the number of cyber-users increases. For instance, the growing presence of the Metaverse represents a new opportunity for abuse. The use of chatbots, artificial intelligence (AI) and machine learning (ML) in research and analytics (ChatGPT) holds a great deal of potential but these tools also generate vulnerability, since hackers may employ AI tools for sophisticated assaults. Moreover, deepfakes and other malicious bots are already in use, while the Russian invasion of Ukraine has exposed the vulnerability of critical infrastructure to nation-state threats, such as an increase in distributed denial of service (DDoS) assaults on websites and infrastructure (DDI, 2023). A major example of this has been the hacking of a Ukrainian satellite, which has further shown that countries and governments with highly secretive and covert activities are vulnerable to cyber threats and cyberattacks (ibid.).

In essence, this article analyses the susceptibility of the Commonwealth APAC region to cyberattacks and how recent events such as the COVID-19 pandemic and the proliferation of AI and chatbots may exacerbate this vulnerability. The pandemic-related transition to remote working has amplified online engagement and escalated cybercriminal endeavours. The article also aims to suggest ways of enhancing cyber-resilience within the region.

The focus of this article on the Commonwealth countries within APAC is justified by the increasing convergence between internet usage, internet proliferation and cybercriminal activities in the region. This connection is presenting considerable obstacles for cybersecurity, law enforcement and the broader security environment in the APAC region. Addressing it will necessitate a comprehensive strategy encompassing the reinforcement of cybersecurity protocols, the augmentation of law enforcement capacities, the facilitation of international collaboration and the elevation of public consciousness regarding the perils associated with cybercrime and its connection with other criminal activities.

The article first looks into the context behind cybercrime in the APAC region; the situation post-COVID in terms of the proliferation of cybercrime in the region; the pros and cons of the potential use of AI in cybersecurity; cybersecurity initiatives and strategies in the Commonwealth APAC region; and options for policy consideration.

2. The context of cybercrime in the APAC region

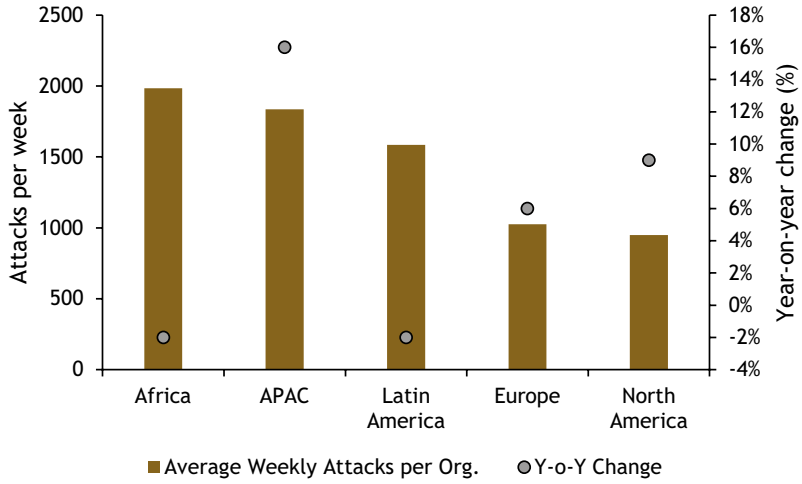
This section provides statistics to support the focus of the article on cyber-vulnerability in the APAC region.

2.1 Average rate of weekly cyberattacks in the APAC region

Even though the African region has the highest average of weekly cyberattacks, Figure 1 shows that the year-on-year percentage change was highest in the APAC region in the first quarter of 2023. This implies that the rate of increase in average weekly cyberattacks in the APAC region was well above that in every other region. For context, the chart also shows that, while it may seem that the African region had the highest number of average weekly attacks, this figure has reduced compared with what it was in the same period of the previous year based on the data collected.

There are a few reasons given for this increasing trend in the APAC region: the trojanising of the 3CXDesktop app for a supply chain attack, the use of ChatGPT for code generation that can help less-skilled threat actors launch cyberattacks without effort, the leveraging of the critical unauthorised remote code exploitation (RCE), and the vulnerability in the Microsoft Message Queuing (MSMQ) service (Gullipalli, 2023).

Figure 1. Average weekly cyberattacks per organisation, by region, Q1 2023

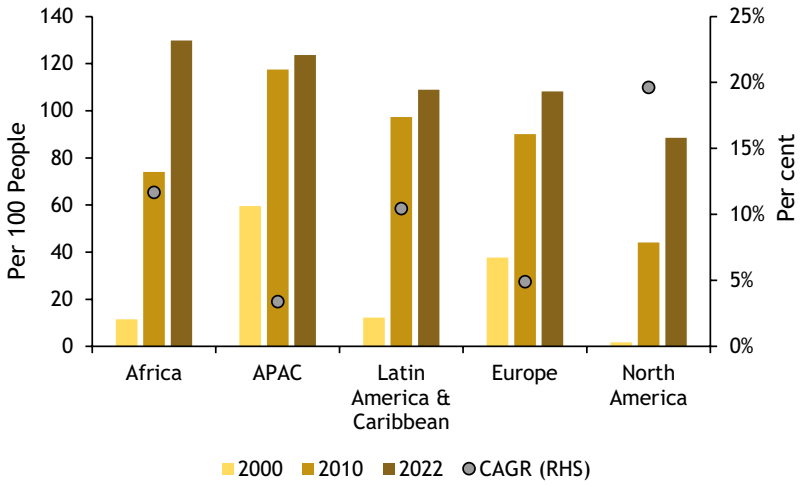


Source: Check Point Research (2023).

2.2 Mobile cellular subscriptions and links to cybercriminal activities

The increase in mobile cellular subscriptions may have contributed to the rise in global cybercriminal activities. Figure 2 reveals that mobile cellular subscriptions expanded significantly in the five regions between 2000 and 2022, with the highest compound annual growth rate (CAGR) found in North America. The results also show that the APAC region had the highest number of mobile cellular subscriptions (per 100 people) in 2010, showing the rising trend in internet activities in the region. However, Africa had the largest

Figure 2. Mobile cellular subscriptions by region (per 100 people)



Source: World Development Indicators 2023.

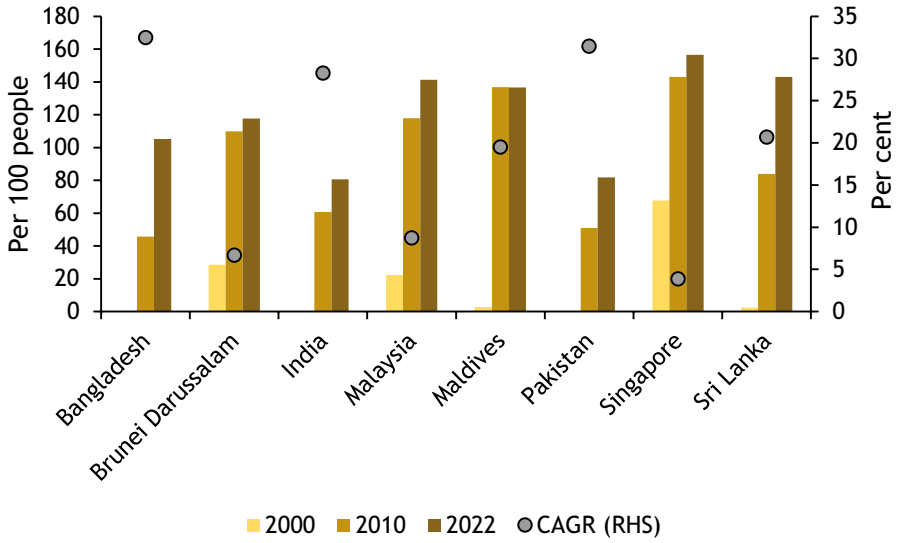
number in 2022, at an average of roughly 130 per 100 people, slightly higher than the APAC region's 124 per 100 people.

The number of mobile cellular subscriptions shown in Figure 2 is likely linked to the average weekly cyberattacks in Figure 1. For example, Africa has the largest number of mobile cellular subscriptions per 100 people, as presented in Figure 2, and at the same time the largest average number of weekly cyberattacks, as shown in Figure 1. This implies that the rise in mobile cellular subscriptions may have led to the rise in weekly cyberattacks, with the APAC region suffering a large chunk of these.

Figures 3 and 4 take the analysis a step further by comparing the Commonwealth APAC countries' performance in terms of their cellular subscriptions over the internet. Interestingly, the findings demonstrate that, even though Singapore has the highest number of mobile cellular subscriptions per 100 people among the Commonwealth Asian countries (Figure 3), the gap is not as wide as we might think: the other Commonwealth countries in the region closed the gap over the 22-year period. As a result, Bangladesh and Pakistan had the highest CAGR in mobile cellular subscriptions per 100 people between 2000 and 2022 and Singapore had the lowest. Nevertheless, the data show that all the countries have embraced connectivity through the use of mobile phones and gadgets to gain access to the internet. Reasons for this include the reduced cost of internet subscriptions on mobile devices, more individuals embracing mobile usage across different age groups and a rise in e-commerce (Brain and Oyadeyi, 2023).

The Commonwealth Pacific region has also witnessed a surge in mobile cellular subscriptions. Figure 4 shows that New Zealand, Australia and Fiji had the highest number of mobile cellular subscriptions per 100 people and Papua New Guinea and Kiribati the lowest.

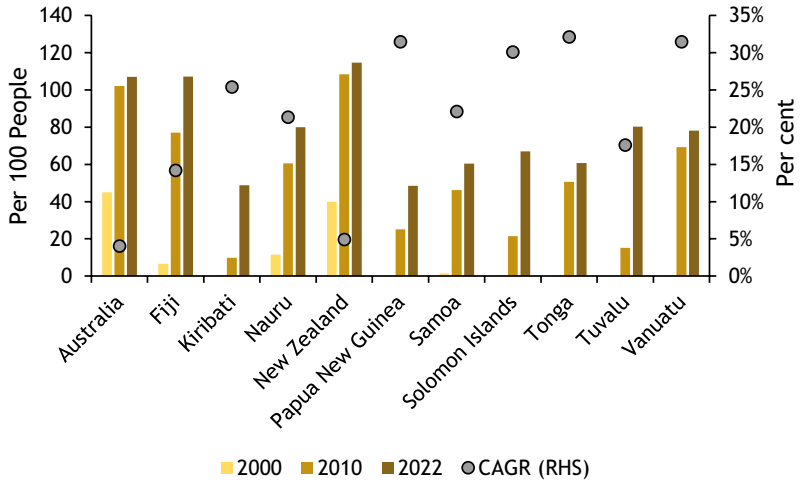
Figure 3. Mobile cellular subscriptions among Commonwealth Asian countries (per 100 people)



Source: World Development Indicators 2023.

In general, these findings imply that the Commonwealth APAC region is making significant strides in its use of mobile cellular subscriptions whether for good or for bad reasons. 'Good' reasons include the fast-paced development in information and communication technology (ICT) and 'bad' reasons include the potential for

Figure 4. Mobile cellular subscriptions among Commonwealth Pacific countries (per 100 people)



Source: World Development Indicators 2023.

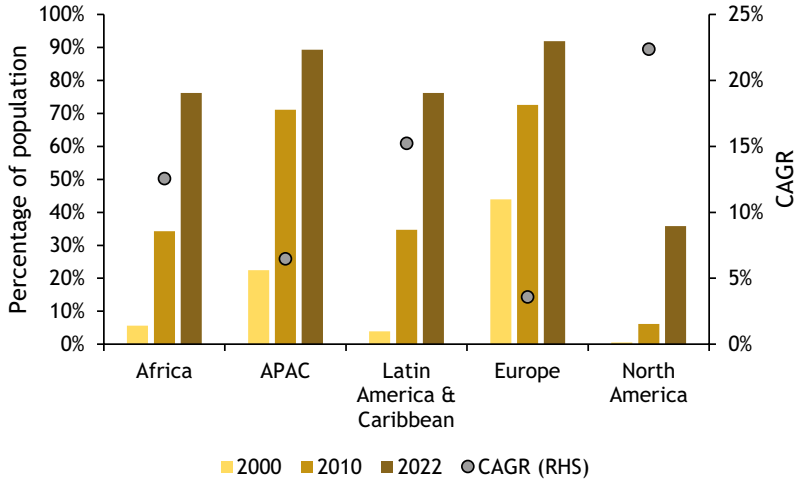
cybercriminals to use such subscriptions to access the internet and target their victims, unlike in the early 2000s, when such activities were performed only using desktops and laptops. Now, with a mobile, cybercriminals can easily access victims to defraud.

Therefore, cybercrime is a particular concern in the Commonwealth APAC region because of the prevalence of internet access on mobile phones. Unlike PCs, mobile phones often lack protective measures like firewalls, antivirus software, encryption and so forth. These heightened possibilities, coupled with the region's inadequate legal framework for cybercrime, make it a prominent target for cybercriminals and their activities.

2.3 Internet penetration and links to cybercriminal activities

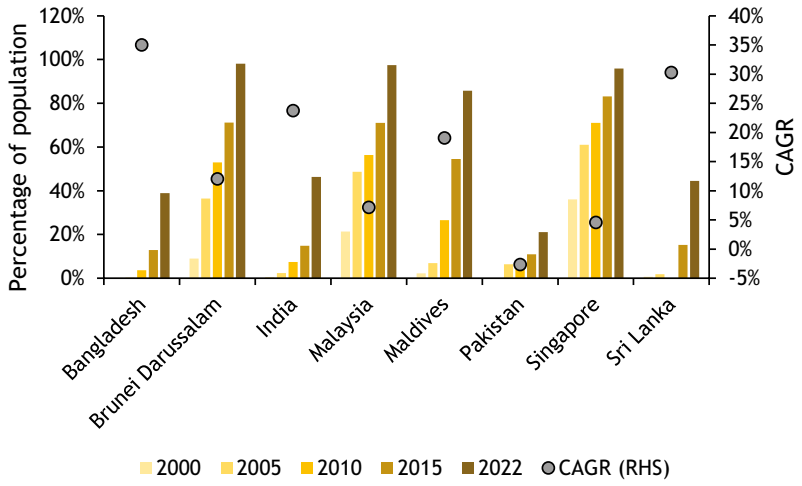
The proliferation of digital technology in the Commonwealth APAC region has resulted in heightened connectivity and a growing dependence on digital infrastructure (Runde et al., 2020). The level of internet penetration within the APAC region ranks high; only Europe rivals (or ranks higher than) the APAC region when it comes to internet use as a percentage of the population. In figures, roughly 89 per cent of the population in the APAC region uses the internet, which is the second highest internet penetration use, beneath only Europe, which has roughly 92 per cent usage. This shows that ICT development within the region is advanced and has been higher than in most other regions except Europe. Weak legislation on cybercriminal activities has also potentially helped cybercriminal activities over the internet proliferate: Figure 1 shows that the region sees the highest number of average weekly cyberattacks globally.

Figure 5. Individuals using the internet by region (% of population, and compound annual growth rate [CAGR])



Source: World Development Indicators 2023.

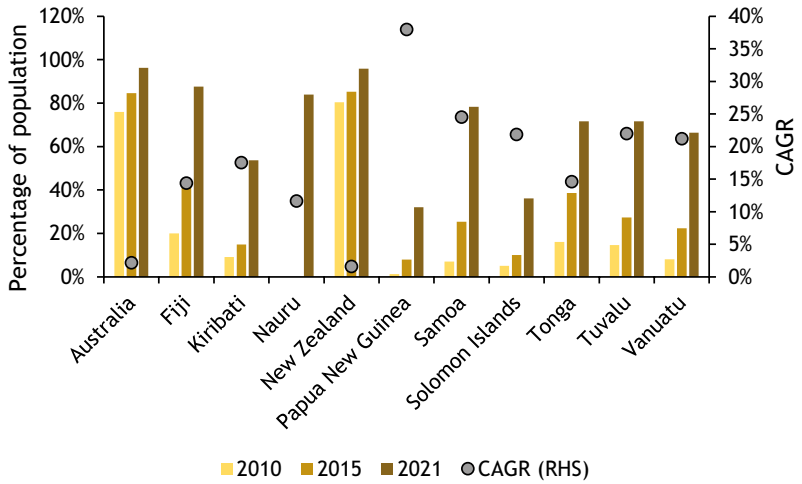
Figure 6. Individuals using the internet in Commonwealth Asian countries (% of population, and compound annual growth rate [CAGR])



Source: World Development Indicators 2023.

In 2022, Brunei Darussalam, Malaysia and Singapore had the highest share of people using the internet in Commonwealth Asia, at 98, 97 and 96 per cent of the population, respectively (Figure 6). The figures were the lowest in India (46 per cent), Bangladesh (39 per cent) and Pakistan (21 per cent). Therefore, from Figure 6, we can see that Brunei Darussalam, Malaysia, Singapore, and Maldives, are the key driving forces behind the

Figure 7. Individuals using the internet in Commonwealth Pacific countries (% of population, and compound annual growth rate [CAGR])



Source: World Development Indicators 2023.

Commonwealth Asian region's internet penetration of 89 per cent in Asia as illustrated in Figure 5.

Figure 7 shows internet penetration as a share of the population in Commonwealth Pacific countries. Australia (96.2 per cent), New Zealand (95.9 per cent) and Fiji (88 per cent) have the highest rate and Kiribati (54 per cent), Solomon Islands (36 per cent) and Papua New Guinea (32 per cent) had the lowest in 2021.

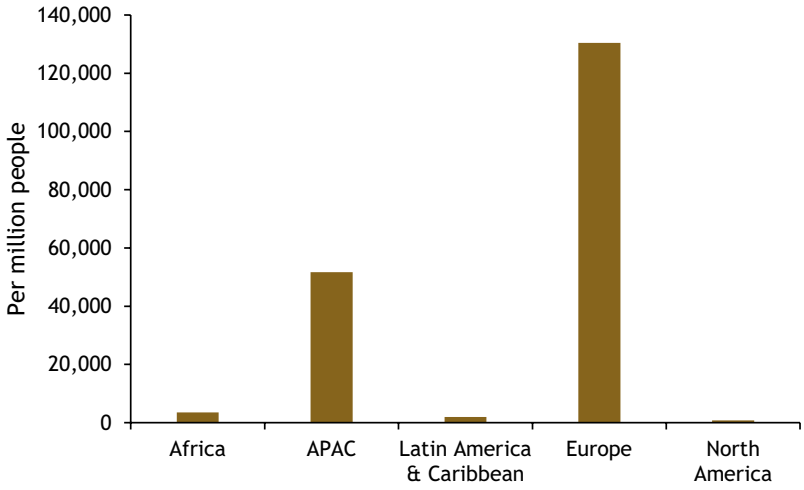
2.4 Internet server security and links to cybercriminal activities

Figure 8 shows how successful each region has been in enforcing stringent cybercrime policies. Europe has the most secure internet users in the world. That is, for every 1 million people, 130,402 persons have secure access to the internet or roughly 13 per cent of the population. This rate is low but still the highest of all regions. Europe's position may be linked to the high presence of cybercriminal laws and regulations, as well as stringent penalties for anyone caught engaging in cybercriminal activities.

The APAC region comes in second at roughly 52,000 secure internet users per 1 million people or roughly 5.2 per cent of the population in the region. However, many of the countries driving these numbers are not Commonwealth APAC countries (Table 1). Finally, Africa, Latin America and the Caribbean, and North America have extremely low figures per 1 million people.

Table 1 sheds more light on the information provided in Figure 8 on the APAC region. Among Commonwealth Asian countries, Singapore had the highest rate of secure

Figure 8. Secure internet servers by region (per million people)



Source: World Development Indicators 2023.

Table 1. Secure internet servers in Commonwealth APAC countries (per 1 million people)

Asian countries	2010	2015	2020	CAGR	Pacific countries	2010	2015	2020	CAGR
Bangladesh	0	2	138	92%	Australia	1,403	4,574	39,853	40%
Brunei Darussalam	40	565	15,598	81%	Fiji	14	86	259	34%
India	2	12	474	76%	Kiribati	0	9	40	20%
Malaysia	44	228	7,306	67%	Nauru	0		325	35%
Maldives	25	122	1,124	46%	New Zealand	1,389	3,921	20,509	31%
Pakistan	1	3	72	63%	Papua New Guinea	1	13	52	49%
Singapore	532	3,585	128,378	73%	Samoa	26	123	475	34%
Sri Lanka	3	21	384	60%	Solomon Islands	2	21	62	42%
					Tonga	9	104	561	51%
					Tuvalu			271	72%
					Vanuatu	41	123	359	24%

Source: World Development Indicators 2023.

internet servers in 2020, at roughly 13 per cent of the population. This was followed by Brunei Darussalam, at roughly 1.6 per cent. The rest of the Commonwealth Asian countries had a rate of less than 1 per cent, meaning that public and private organisations and investors need to put more emphasis on building secure internet servers within the region. Based on these figures, the potential for businesses that want to delve into creating a secure cyberspace is huge; likewise, the potential for cybercriminals to perpetrate their activities is also huge, as many people in Commonwealth Asian countries are not using a secure internet connection and this makes them vulnerable to cyber-scams, attacks and threats.

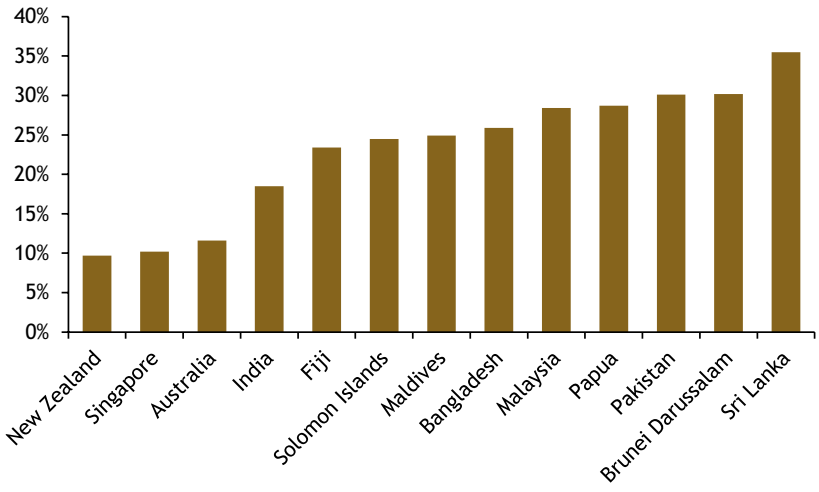
In the Commonwealth Pacific countries, only about 4 per cent of people in Australia use the internet securely, and only about 2 per cent in New Zealand. The rest of the Commonwealth countries in this region do not reach 1 per cent.

Overall, these figures highlight a huge gap in infrastructure for many Commonwealth countries and enormous susceptibility to cyberattacks.

2.5 The shadow economy and the APAC region

The underground economy can be described as all economic activities that are hidden from official authorities for monetary, regulatory and institutional reasons (Medina and Schneider, 2019). Reasons may include tax avoidance, corruption and weak institutions or regulatory frameworks. Figure 9 shows that the shadow economy is largest in Sri Lanka (35.5 per cent), Brunei Darussalam (30.2 per cent) and Pakistan (30.1 per cent). This implies that the size of the shadow economy compared with overall economic activities in these countries is huge, at over a third in Sri Lanka and at a little less than a third in Brunei

Figure 9. Shadow economy (% of GDP) in the Commonwealth APAC region



Source: Medina and Schneider (2019).

Darussalam and Pakistan. On the other hand, New Zealand (9.7 per cent), Singapore (10.2 per cent) and Australia (11.6 per cent) have the least informal economies in the Commonwealth APAC region. This points to the level of development in these countries and the high level of regulatory frameworks guiding against underground activities.

In essence, the high level of cybercriminal activities in the Commonwealth APAC region (as presented in Figure 1) may relate to the level of underground activities going on (Figure 9). This is particularly so in countries that have higher shadow economy levels and a weaker secure internet server situation (as presented in Table 1), such as Papua New Guinea, Brunei Darussalam and Pakistan, to mention a few. Consequently, a larger shadow economy is a concern for the Commonwealth APAC region, since it may aid cybercriminal activities.

3. Post-COVID situation and cybercriminal activities in the APAC region

The COVID-19 pandemic had a significant impact on the cybercrime landscape in the APAC region (Germanos and Georgiou, 2021). It led to a rise in the reliance on AI-based technologies for work, education and entertainment (Khan et al., 2022). The heightened dependence on digital technology during the pandemic engendered new prospects for cybercriminals while at the same time intensifying pre-existing susceptibilities. Consequently, there was a notable increase in cyberattacks inside the APAC region, affecting various entities both public and private organisations as well as individuals (Christine and Thinyane, 2020; Singh, 2022).

Key trends in cybercrime in the Commonwealth APAC region post-COVID include the following.

Increase in phishing and malware attacks

Phishing and malware attacks are among the most common types of cyberattacks, and they have become significantly more frequent since the pandemic (Lallie et al., 2021). Cybercriminals are using phishing emails and websites to trick individuals into revealing their personal information or clicking on malicious links (Alkhalil et al., 2021). Malware is also being used to infect devices with ransomware, which can encrypt data and demand a ransom payment. There has been a substantial increase in the occurrence of phishing attacks and data breaches among Commonwealth APAC countries. For instance, the Tasmanian education department in Australia was the victim of a cyberattack that led to the exposure of roughly 30,000 documents on the dark web, many of which contained sensitive personal information of schoolchildren. Both Australian and New Zealand citizens were victims of the Latitude Financial attack in Australia, which affected over 14 million customers. In Bangladesh, over 14 million details were breached through the Office of the Registrar General. The objectives of these attacks were not only the acquisition of financial benefits but also the deliberate disruption of vital services and the exploitation of confidential information (Smail, 2023).

Rise in ransomware attacks

The emergence of ransomware has resulted in significant monetary damage and the disruption of essential services within diverse industries (Laitinen and Armstrong-Smith, 2022). Ransomware attacks in the APAC region encompass the unauthorised penetration of computer systems or networks, wherein sensitive data is encrypted and a ransom is subsequently demanded in exchange for the provision of decryption keys (Dimitrov, 2020; Amankwah-Amoah et al., 2021; Skouby et al., 2022). The APAC region has experienced a significant rise in both the number and the severity of these attacks, especially since the pandemic. In addition, the pandemic expedited the adoption of remote work practices, which increased dependence on digital platforms (Amankwah-Amoah et al., 2021), making many firms vulnerable to ransomware attacks.

Within Commonwealth APAC countries, Australian commercial law firm HWL Ebsworth has been a victim of ransomware attacks (Smail, 2023). Singapore's Cyber Security Agency reported 8,500 cases of phishing attempts in 2022, a rise from 3,100 cases in 2021 (Chakravarti, 2023). Also, an Indian cybersecurity firm exposed plans by cybercriminals originating from Indonesia and Pakistan to disrupt the G20 summit in India using distributed denial of service (DDoS) attacks and mass defacement. According to the Centre for Strategic and International Studies (CSIS), Bangladesh shut down access to its central bank and commission website in 2023 when it received information that an Indian group was trying to hack it, similar to the hack that happened in 2016 that cost the country almost $1 billion (CSIS, 2023). A Pakistani-based hacker group infiltrated the Indian army and education sector in a wave of attacks against Indian government institutions (ibid.).

Finally, geopolitical factors have also served as potential catalysts for cyberattacks. The use of cyber-tools for purposes of disruption or espionage has been observed during Russia's war on Ukraine (Solar, 2023). Actors now utilise advanced strategies including double extortion, which involves the threat of leaking sensitive data if the ransom is not paid (Ryan, 2021). This strategy puts pressure on victims to comply.

Targeting of critical infrastructure

The deliberate targeting of critical infrastructure by cybercriminals presents an increasing danger, particularly to critical sectors of the economy, such as healthcare, electricity, transportation and telecommunications (Mizrak, 2023). The strategies cybercriminals have employed in targeting critical infrastructure have exhibited a notable enhancement in sophistication. Threat actors utilise sophisticated techniques like ransomware, supply chain attacks and zero-day exploits to infiltrate systems (Diogenes and Ozkaya, 2019; Xu et al., 2021). Incidents targeting healthcare systems pose a significant threat to patient care as they disrupt critical medical services and compromise the security of confidential patient information (Argaw et al., 2020). These disturbances not only impede prompt medical attention but also present potential long-term hazards to patient safety and confidentiality. Cyberattacks also pose significant threats to energy systems, which are crucial for providing power to urban areas and facilitating industrial operations (Zhao et al., 2021).

Commonwealth APAC region countries affected by these types of cybercriminal activities include Malaysia and Singapore, among others (Commonwealth Secretariat, 2022).

Exploitation of remote work vulnerabilities

The pandemic mandated a swift transition to telecommuting in various sectors. This sudden shift resulted in numerous organisations being exposed to potential risks, as they hastily attempted to modify their infrastructure to facilitate remote accessibility (Dwivedi et al., 2020). The transition has led to an increase in cyber-vulnerabilities, rendering organisations more prone to cyberattacks. Cybercriminals in the APAC region have exploited these vulnerabilities by utilising remote access points, specifically targeting home networks lacking security measures and employing social engineering strategies to gain unauthorised access to business systems (Aslan et al., 2023).

4. The link between AI and cybersecurity and the role of AI in building cyber-resilience

In recent years, the incorporation of AI into various digital services has experienced continuous and extensive growth. Governments around the world are currently contemplating the implementation of AI systems to assist in a multitude of endeavours, such as the identification and prediction of criminal activities (Engstrom et al., 2020). National security and intelligence organisations acknowledge the potential impacts of AI in achieving cyber-resilience and public cyber-safety (Schmidt et al., 2021). Nevertheless, if the advancement of AI technology (such as face recognition, drones and lethal autonomous weaponry) is not appropriately regulated or supervised, it presents risks to the protection of individual rights and liberties (Ala-Pietilä and Smuha, 2021).

AI and ML have promising prospects for the identification and mitigation of cyberattacks targeting essential sectors of critical infrastructure in the APAC region. Despite this, issues remain regarding its legislation, particularly for small and medium enterprises (SMEs) whose funds for cybersecurity are constrained. Cybercriminals use AI to create and carry out targeted attacks against government entities, businesses and people. Although there is currently a lack of substantial proof regarding cybercriminals possessing extensive technical knowledge in AI manipulation, they are aware of its potential for illicit and disruptive activities (Caldwell et al., 2021). The current trends in cybercrime underscore a growing dependence on the Internet of Things (IoT) for the dissemination of malware, as well as the utilisation of AI to enhance ransomware attacks (Cascavilla et al., 2021). The anticipated growth of this phenomenon is projected to coincide with the rapid proliferation of interconnected gadgets, potentially heightening the susceptibility of both businesses and individuals to cybercriminal activities.

Moreover, the emergence of deepfakes has raised substantial concerns within the realms of national politics and law enforcement, as these have the potential to facilitate

fraudulent acts using impersonation (van der Sloot and Wagensveld, 2022). According to a report by The Asset (2023), the APAC region experienced a 1,530 per cent surge in deepfake cases from 2022 to 2023 amid a growing trend in sophisticated scams and money laundering cases globally, with Commonwealth countries such as Bangladesh (5.44 per cent) and Pakistan (4.59 per cent) the biggest culprits. The report further showed that Singapore, another Commonwealth country in the region, stands out compared with many other countries in APAC, maintaining a low level of 0.89 per cent. Australia has also been able to maintain a low rate, of 2 per cent, despite increased incidents within the region. The use of deepfakes has presented law enforcement agencies with hurdles to climb as a result of the intricate legal considerations involved in cross-border investigations.

How can AI be deployed to foster cyber-resilience? AI has emerged as a pivotal tool in addressing cybersecurity threats by employing ML to monitor and track illegal and malicious activities within similar digital environments (Zeadally et al., 2020). AI-based security systems play a crucial role in distinguishing between 'good' and 'bad' behaviour but more advanced iterations can analyse vast datasets, identifying interconnected activities that may indicate suspicious behaviour by anonymous entities. The proliferation of network computers, the internet and mobile applications has led to a rise in diverse and prevalent cyberattacks, particularly through connected devices with insufficient security measures, exacerbated by the expansion of the IoT. This surge in cybercrime has highlighted the limitations of traditional 'signature-based' cybersecurity methods (Zhang et al., 2022). These conventional approaches require substantial human effort to identify risks, develop risk features and integrate threat characteristics into software, often falling short of addressing modern, complex cyberattacks, which is an area where AI can effectively be used to boost cyber-resilience.

Furthermore, the concept of CAPTCHA exemplifies the intersection between AI and cybersecurity, requiring users to identify distorted letters or images as a test to distinguish between humans and computers (Al-Maliki et al., 2023). When conventional security systems prove ineffective against evolving threats, AI-driven approaches enhance the overall security architecture, offering robust protection against a diverse range of intricate cyberattacks. Companies integrating AI into their operations witness improved business processes and financial outcomes, particularly through AI-powered cybersecurity solutions that swiftly develop data-driven security models across various domains (Allioui and Mourdi, 2023). This is because AI-based monitoring systems continuously track user behaviour, promptly identifying anomalies, providing a significant advantage in today's dynamic cybersecurity landscape (Zeadally et al., 2020). In essence, AI and ML technologies serve as effective anti-malware defences against sophisticated cybercriminal tactics such as camouflaging malware and ransomware to evade detection (Ferdous et al., 2023). These technologies enable systems to cross-reference new malware with existing databases, assess code and pre-emptively prevent potential attacks, even when malicious code is concealed within large volumes of benign or irrelevant data.

5.1 How international and national legal measures have helped combat cybercrime

One major impact of regional and national legal frameworks in the APAC region is the creation of vast awareness of the dangers of cyberattacks and the ways to mitigate such attacks. The various frameworks put in place by many of these countries have created cyber-awareness among government institutions, business owners and individual users. Australia, for example, has avenues for training in cyber-awareness on protecting emails, personal phones and online transactions (ASD, 2024). India has an incident report framework giving a duration within which cyberattacks must be reported to the appropriate agency. Cases of business email compromise (BEC) are on the rise; the case on BEC heard at the ACT Civil and Administrative Tribunal (ACAT) is one such cyberattack incident (ibid.). The Australian Court puts the responsibility on companies to protect their systems to prevent the future occurrence of BEC when dealing with clients and other businesses (Falk, 2022).

5.2 Gaps in national laws and the regional or international legal framework

The United Nations Office on Drugs and Crime noted that the Commonwealth of Independent States' CIS Agreement on Cooperation in Combating Offences Related to Computer Information of 2001 calls on countries to adopt national laws to implement the Agreement's provisions to harmonise their national cybercrime laws (UNODC, 2019). Singapore, Australia and New Zealand are some of the top Commonwealth APAC countries in terms of formulating cybersecurity strategies, establishing computer emergency response teams (CERT), creating governmental agencies and enacting laws to protect their critical information infrastructure, economy, businesses and people from incessant cyberattacks while paying attention to international frameworks such as the CIS Agreement and to UNODC, Council of Europe and EU cybersecurity policies.

The Association of Southeast Asian Nations (ASEAN) is another example of regional co-operation on cybersecurity in APAC. The focus of ASEAN is threefold: ensuring member states make provision for (i) cybersecurity incident response; (ii) CERT policy and co-ordination; and (iii) cybersecurity capacity-building (ASEAN, 2017).

There are gaps in the regional framework of ASEAN and in the national strategies of these countries (Gan, 2024). Also, the APAC region has uneven cybersecurity development (ASEAN Cyber Security Cooperation Strategy, 2021). Singapore and Australia have more legal and judicial frameworks in place than countries such as Kiribati, Solomon Islands and Tuvalu. India's cybersecurity legislation has not been updated recently, and Malaysia has unspecialised cybersecurity legislation (Positive Technologies, 2023).

Table 2. Major cyberattacks in the APAC region in 2023

Institutions	Date	Country	Affected
Tasmanian Education Department	April 2023	Australia	30,000 documents
Samsung ChatGPT incident	April 2023	APAC	Company and all Samsung users
Toyota cyberattack	May 2023	APAC	Over 2 million customers
Latitude Finance	March 2023	Australia	14 million customers
Bangladesh Registrar General cyberattack	July 2023	Bangladesh	14 million citizens
Tissupath Clinic attack	August 2023	Australia	10 years' worth of data breached

Source: Smail (2023).

5.3 Cybersecurity incidents, court judgements and sanctions

The APAC region has experienced several cyberattacks and has been labelled highly vulnerable to cybercriminal activities. The Medibank Group cyberattack of October 2022, the Optus Pty Limited cyberattack and the Costa Group cyberattack were some of the many devastating cyberattacks in the APAC region in 2022. The Medibank cyberattack led to the exposure of 9.7 million customers' data. Table 2 presents major cyberattacks occurring in the Commonwealth APAC region in 2023.

These cyberattacks have led to several sanctions. For instance, the Australian government sanctioned Aleksandr Ermakov, who was linked to the Medibank cyberattack. It has been deemed criminal for anyone to use or deal with Aleksandr Ermakov's assets, cryptocurrency wallet and ransomware payments, with a 10-year imprisonment or heavy fines imposable (Wong et al., 2024).

Australia also saw some court judgments delivered in cyberattack cases at the Federal Court and the ACAT. In 2020, the Australian Security and Investments Commission (ASIC) instituted an action against RI Advice Group at the Federal Court. The action was instituted owing to RI Advice Group's lack of documented cybersecurity measures to protect its client's information, leading to losses for its clients for six years. The company was penalised to the tune of AU$750,000, to be paid to ASIC as compensation (Falk, 2022). These sanctions have helped solidify the Australian authority's efforts in the fight against cybercriminal activities; other Commonwealth countries in the APAC region that are yet to make strong cybercriminal legislation can adopt such measures.

5.4 Relevance of regional and international co-operation in fighting cybercrime

The transnational nature of cyberthreats means regional and international co-operation play a crucial role in combating cybercrime. In the APAC region, several organisations play a crucial role in tackling cybercrime and enhancing cybersecurity (Benincasa, 2020). These include the Asia-Pacific Computer Emergency Response Team**,** Interpol Global Complex for Innovation**,** ASEAN's Cyber Capacity Programme, Asia-Pacific Telecommunity, Asia-Pacific Economic Cooperation's Telecommunications and Information Working Group, Asia-Pacific Network Information Centre, Asia-Pacific Regional Internet Governance Forum and Asia-Pacific Security Forum. These agencies foster collaboration, share expertise and promote best practices in cybersecurity (Sarowa et al., 2020). Their relevance in the APAC region in fighting cybercrime is briefly discussed in the subsections below.

Information-sharing and co-ordination

Regional organisations serve as platforms for member states to share threat intelligence, cyber-incident data and best practices in cybersecurity (Rui, 2023). By facilitating information exchange and co-ordination, they contribute to a more effective global response to cyberthreats (Sarowa et al., 2020).

Capacity-building and training

Regional organisations enhance the skills and capabilities of cybersecurity professionals in member states (Quimba and Barral, 2022). By investing in human resources development and knowledge-sharing, they contribute to building a global network of cyber-experts equipped to address evolving cyberthreats (Kumar, 2020).

Harmonisation of cyber-policies

Regional organisations work towards harmonising cybersecurity policies, standards and frameworks across member states. By promoting policy coherence and alignment on cybersecurity issues, they contribute to a more unified approach to combating cybercrime at the international level (Tien and Cheng, 2016).

Promotion of international cyber-norms

Regional organisations advocate for the adoption of international cyber-norms, principles and best practices to promote responsible behaviour in cyberspace (Herko, 2023). By endorsing global cybersecurity standards and norms, they contribute to a more secure and stable international cyber-environment (Ang, 2021).

Collaboration with international partners

Regional organisations engage in partnerships and collaborations with international entities, such as INTERPOL, the United Nations and other global organisations (Araki,

2022). In doing so, they contribute to a co-ordinated and comprehensive global response to cyberthreats (Rui, 2023).

Advocacy for cyber-resilience

Regional organisations advocate for cyber resilience and the importance of cybersecurity at the international level (Herko, 2023). By raising awareness, promoting good governance and advocating for cybersecurity measures globally, they contribute to strengthening the overall resilience of the international community against cybercrime (Slayton and Clarke, 2020).

Contribution to global cybersecurity initiatives

Regional organisations actively participate in global cybersecurity initiatives, conferences and forums to share insights, experiences and expertise on cybercrime prevention and response (Ang, 2021). In this way, they play a vital role in shaping international cybersecurity agendas and strategies (Bahuguna et al., 2020).

5.5 Challenges in enforcing legal provisions on cybercrime at the regional level

Enforcing legal provisions on cybercrime in the APAC region faces several challenges to the effective prosecution and deterrence of cybercriminal activities.

Jurisdictional issues

Cybercrimes are often transnational, making it challenging to determine jurisdiction and prosecute offenders operating across multiple jurisdictions (Quimba and Barral, 2020). Lack of clear legal frameworks for cross-border co-operation and extradition complicates the enforcement of cybercrime laws in the region (Araki, 2022).

Lack of harmonised legislation

Variations in cybercrime laws and regulations among countries in the APAC region create inconsistencies and gaps in legal frameworks (Kumar, 2021). The absence of harmonised legislation hampers international co-operation and co-ordination in combating cybercrimes effectively (Araki, 2022).

Capacity and resourcing difficulties

Many countries in the region face resource constraints, including on funding, technical expertise and specialised cybercrime units (Tien and Cheng, 2016). This hinders law enforcement agencies' ability to investigate cybercrimes, gather digital evidence and prosecute offenders (Roberts, 2022).

Technological challenges

Rapid technological advancements present challenges for law enforcement agencies in keeping pace with evolving cyberthreats (Ang, 2021). Cybercriminals often use sophisticated techniques and encryption methods to conceal their activities, making it difficult for authorities to detect and investigate effectively (Herko, 2023).

Data privacy concerns

Balancing the need for law enforcement access to digital evidence with data privacy rights poses a significant challenge in enforcing cybercrime laws (Sarowa et al., 2022). Striking a balance between investigating cybercrimes and protecting individuals' privacy rights is a complex issue that requires careful consideration and legal safeguards (Benincasa, 2020).

Cross-border co-operation

Effective enforcement of cybercrime laws requires close co-operation and information-sharing among law enforcement agencies across borders (Sarowa et al., 2022). Challenges such as differing legal systems, language barriers and cultural differences can impede seamless collaboration in combating transnational cybercrimes (Bahuguna et al., 2020).

Inadequate cybersecurity capacity-building

Building the technical capabilities and expertise of law enforcement agencies to investigate cybercrimes is essential for the effective enforcement of legal provisions (Kumar, 2021). A lack of specialised training programmes and cybercrime units in some countries hinders their ability to respond to and investigate cyber-incidents.

Difficulties in setting up public–private partnerships

Collaboration between government agencies, private sector entities and civil society organisations is crucial in combating cybercrime (Slayton and Clarke, 2020). However, establishing effective public–private partnerships and information-sharing mechanisms can be challenging owing to concerns about data protection, trust issues and differing priorities (Ang, 2021).

The need for legal framework adaptation

Cyberthreats evolve rapidly, with continuous updates and adaptations to existing legal frameworks required to address emerging challenges (Araki, 2022). The process of revising laws and regulations to keep pace with technological advancements can be slow and complex, delaying the enforcement of legal provisions on cybercrime (Kumar, 2021).

6. Efforts toward the enforcement of cybersecurity laws to build cyber-resilience

To combat cybercrime, the Commonwealth APAC countries have adopted different national strategies.

6.1 Asia region

Singapore

Singapore launched its first Cybersecurity Strategy in 2016 and updated this in 2021 to develop a vibrant cybersecurity ecosystem and grow a robust cyber talent pipeline. Singapore has laid out three pillars – 'build resilient infrastructures,' 'enable safe cyberspace' and 'enhance international cyber co-operation' – to accomplish its goals. The Cyber Security Act of 2018 is one of the laws passed to put this into action. The Cyber Security Agency (CSA) has been established to manage the establishment of a cyberthreat response team and the development of cyberspace awareness initiatives for organisations, companies and people. Other government departments receive security consultation services from CSA. Given the global nature of cybercrime, CSA actively promotes international and regional capacity-building initiatives with other nations (CSA Singapore, 2023). Cybercrime monitoring, and policy generation within existing regulatory frameworks for implementation by various financial institutions, payment platforms and the general public, is a primary function of Singapore's Monetary Authority.

Brunei Darussalam

Brunei Darussalam has formed Cyber Security Brunei (CSB). Despite being a government body, this assists both public and private entities in their fight against cybercrime. CSB raises public awareness of cybercrime, develops policies to deal with cyber-risks and plans to strengthen the national enforcement agency's enforcement capabilities (ibid.). Brunei's Computer Emergency Response and the National Digital Forensic Laboratory are the ways by which they carry out their mandate. CSB has rules and regulations pertaining to cybercrime, including the Cyber Security Order 2023 and the National Cybersecurity Framework, which provide organisations with standards, guidelines and protocols to fight cybercrime.

India

India has purposefully established a multistakeholder ecosystem to handle cybercrime. The National Cyber Security Strategy, launched in 2013, with a revamp in 2023 (in its final stages of approval) aims to help raise awareness about cybercrime, educate the public on data protection, develop strategies to prevent cyberattacks and bring those responsible to justice (Inamdar, 2023). The strategy will help agencies including the National Cybercrime Reporting Portal, Platform for Joint Cybercrime Investigation Team, National Cybercrime Forensic Laboratory and National Cybercrime Threat Analytics Unit.

The country's National Cyber Crime Research and Innovation Centre keeps an eye out for emerging cybercrime trends and develops strategies to combat them.

Malaysia

To ensure the safety of its critical and national information infrastructure, Malaysia established the National Cyber Security Agency (NACSA) in 2017 and created the National Cybersecurity Strategy in 2020. Its cybersecurity strategy plan spanning the years 2020–2024 is built around five main pillars: (i) efficient administration and control; (ii) increasing the robustness of law enforcement; (iii) facilitating first-rate innovation, technology, research and development, and business; (iv) improving awareness, education and capacity-building; and (iv) building international co-operation. The tenets of these pillars include legislation, awareness-raising initiatives, cyberthreat agencies and enhanced cyberattack prevention measures.

Maldives

Cybersecurity Maldives was established in 2013 with the mandate of performing security audits, assessing network threats and developing solutions for cyberthreat hunting. In addition to assisting businesses and organisations in preventing cyberattacks and safeguarding their data, it conducts penetration tests. Meanwhile, Maldives has initiated the Digital Maldives for Adaptation, Decentralisation, and Diversification Project, with cybersecurity as one of its aims. Maldives and India signed a multipronged international co-operation partnership in 2022 to fund a wide range of development initiatives, one of which is cybersecurity on the island country. Maldives and Bahrain have also signed a memorandum of understanding.

Sri Lanka

As a result of the establishment of a National Cybersecurity Agency in 2016, Sri Lanka now has a Centre for Computer Incident Response Team. The government approved the Cybersecurity Bill 2023 in July of that year. The agency will oversee the creation of cybercrime awareness, the formulation of regulatory frameworks for organisations, training and the creation of a computer incidence response team. Sri Lanka has an information and cybersecurity plan for the years 2019–2023. The country also has laws that make it a felony to commit certain types of cybercrime. Sri Lanka's legal framework for dealing with cybercrime and malevolent online actions includes the Computer Crimes Act 2007, the Payment Devices Fraud Act 2006, the Intellectual Property Act 2006, the Electronic Transaction Act 2006 and the Information and Communication Technology Act 2003.

Pakistan

Many cybercrimes and internet-related offences are punishable by law in Pakistan. A national cybersecurity policy was unveiled in July 2021. This allows for the establishment

of a national cybersecurity agency and the development of public and private sector cybersecurity rules, frameworks, processes and standards. Nevertheless, these rules remain unenforced.

Bangladesh

According to the Bangladesh Gazette 2018, the country's legal framework for dealing with cybercrimes and the prosecution of those accused of such crimes is embedded within the Digital Security Act 2018. The Bangladesh Cybersecurity Act 2023 passed into law in 2023 and supersedes the Act of 2018. Reportedly, the substance of the two Acts is the same, although they go by different names. Amnesty International (2022) has speculated that the most recent version includes language that may lead to abuses of human rights. In addition, Bangladesh has formulated its Cybersecurity Strategy 2021–2025 but it is yet to implement this.

6.2 Pacific region

Australia

Australia has changed its cybersecurity strategy, laws, agencies, structure, standards and guidelines, joining other Pacific nations in this effort. The government previously unveiled two cybersecurity plans, in 2016 and 2020. After soliciting inputs from the private sector, the government, academia and business, it then unveiled its 2023–2030 Cybersecurity Strategy, aiming to become a global leader in cybersecurity. The Australian Signals Directorate set up the Australian Cyber Security Centre in 2023 to enhance cyberattack awareness, conduct cybersecurity assessments, spread information about cybercrime reporting and recovery, and put into action many other aspects of the Australian Cybersecurity Strategy.

Vanuatu

The National Security Strategy of Vanuatu has cybersecurity as its fifth pillar. The aim is to protect cyberspace and the nation's critical information and infrastructure. Cybercriminal activities are now a crime, punishable by law, according to the Cybersecurity Act of 2021. The government has also developed the National Cybersecurity Strategy 2030, the National Harmful Digital Communication Strategy 2023 and Vanuatu National Data Protection & Privacy Policy. Vanuatu has intergovernmental programmes involving many of its agencies on training, cyberattacks and threat reporting, among others. The main agency in charge of Cybersecurity in the country is CERT Vanuatu.

Fiji

Fiji has in place the Cybercrime Act 2021. The Ministry of Information and Communication Technology is responsible for the country's cybersecurity. To strengthen the country's overall digital presence and to control and mitigate unwanted cybercriminal

activities, it oversees programmes such as Digital FIJI and many more. The country's CERT and an updated National Cybersecurity Strategy are nearing completion. To enhance its security policy, the country takes part in regional partnerships with Australia as well as other international collaborations.

Kiribati

The Ministry of Information, Communications and Transport of Kiribati prepared a national cybersecurity strategy in 2020 and enacted the Cybercrime Act in 2021. No dedicated government body has been established to carry out this plan; the ministry oversees implementation of the cybersecurity strategy.

Nauru

As a member of the Pacific Cybersecurity Operational Network, Nauru passed a law in 2015 to combat cybercrime and make it easier to prosecute those responsible. The Ministry of Information and Telecommunications is responsible for implementing cybersecurity policy.

Tonga

The Strategy and Computer Emergency Response Team in Tonga oversees the implementation of the National Cybersecurity Framework 2022. This specialist agency enforces Tonga's cybersecurity rules. Critical infrastructure operators, as well as public and private sector organisations, obtain guidance and assistance from the response team.

New Zealand

New Zealand passed the Intelligence and Security Act 2017. In addition, the country launched a cybersecurity strategy in 2019 managed by CERT New Zealand.

Papua New Guinea

In Papua New Guinea, cybersecurity matters are handled by the Department of Information and Communication Technology and other departments that are members of the National Cyber Co-ordinating Centre. There is national legislation and a cybersecurity policy in place to make cybercrime a punishable offence. An example of this is the Cybercrime Code Act 2016. Launched in 2020, the National Cybersecurity Policy 2021 details the government's cybersecurity objectives and plans to attain them.

Samoa, Solomon Islands and Tuvalu

While Samoa has a national cybersecurity strategy for 2016–2021, with SamCERT as its main cybersecurity body, it has no specialised cybercrime legislation. Solomon Islands also does not have cybercrime legislation but the government ICT Unit keeps tabs on the country's cybersecurity activities. Tuvalu does not yet have a dedicated body to

supervise cybersecurity matters; issues are handled by the Department of Information and Communication Technology under the Ministry of Justice.

7. Conclusion and recommendations

7.1 Conclusion

This article has looked in depth into cybercrime in APAC and found that the region is not as resilient to cybercriminal activities as initially thought. This is particularly the case for Commonwealth APAC countries, used as the focus for the study, although some Commonwealth countries, such as Singapore, New Zealand and Australia, are very heavy on cybersecurity initiatives that uncover and address the activities of cybercriminals and promote cyber-resilience. Our study has highlighted the context of cybercrime in the region, and within this the role that the COVID-19 pandemic has played in shifting work online, thereby exposing many organisations in the region to cyberthreats. It has also highlighted the positive and negative roles that AI may play in the fight against cybercriminal activities.

The Commonwealth APAC region needs to strengthen its resilience to and reduce the risk of cyberattacks and threats, and protect its digital infrastructure in the ever-changing landscape. Therefore, we now highlight options to build the resilience of Commonwealth APAC countries to prevent the proliferation of cybercriminal activities in the region.

7.2. Policy considerations to prevent cybercrime

Increased awareness, advocacy and education: The governments of APAC countries should fund campaigns to educate the public about the risks of cybercrime. An example of an existing awareness campaign is Singapore's Better Cyber Safe than Sorry campaign, which began with private e-commerce retailers like Shopee and supermarket chain NTUC Fairprice, and has since expanded to include instructional videos, national television advertisements and posters at most bus stops (Gullapalli, 2023). Individuals and businesses alike can benefit from increased vigilance and preparedness in the face of cybersecurity threats if they are made more aware of the risks they face and given direction on how to react (ibid.).

Improved public and private co-operation: Co-operation between cybersecurity authorities, organisations and governments may aid in the prevention of attacks and proactively address emerging threats. By working together, organisations can improve their defences in response to cyberthreats more quickly.

Creation of taskforces: Learning from the successes of countries such as Singapore, it is pertinent to establish national taskforces devoted to developing, co-ordinating and implementing comprehensive strategies and policies to successfully tackle cybercrime.

Improved government regulations: To safeguard their citizens, APAC countries should consider enacting strict and standardised cybersecurity laws. These policies can create

baselines for security, promote frequent evaluations and impose consequences for failing to comply, drawing inspiration from the practices already in place in Australia and Singapore. By establishing rules that strengthen cybersecurity resilience, APAC nations can push businesses to prioritise safety and implement best practises.

Strengthening cybersecurity governance and leadership: Organisations in the APAC region should hire skilled experts with experience in cybersecurity to senior roles and boards of directors to strengthen cybersecurity leadership and governance frameworks. Fostering a culture of responsibility and giving security measures their appropriate priority can be achieved when cybersecurity is prioritised at the highest levels of decision-making within organisations. Public and private organisations require a chief information security officer, who is given authority and a clear mandate to implement an 'intelligence-led, prevention-first cybersecurity approach' to compete on the new cyber-battlefield (Gullapalli, 2023).

Collaboration with international partners: Cybercrime is a global phenomenon, hence APAC countries need to collaborate with global partners to put an end to it. By doing this, APAC countries will enhance their defence and reduce the threats presented by cybercriminals who may operate from other countries.

Consistent expenditure on security: The public and private entities in the APAC region need to invest heavily in cyber-resilience. To remain ahead of threats and reduce their exposure to attacks, entities must invest in strong security protocols, update and patch systems regularly and undertake thorough security audits.

Responding to the misuse of AI for cybercriminal activities: To respond to the misuse of AI for cybercriminal activities, the APAC region can adopt policy initiatives like the Digital Services Act and the Regulation proposal for Artificial Intelligence Systems adopted by the EU (Schneider and Werle, 2023). Such policy will serve the purpose of creating rigorous guidelines and responsibility for risk assessments using AI services to guarantee the safe use of AI services. Other initiatives to draw lessons from include the Budapest Convention and the recent inclusion of the Second Additional Protocol to the Budapest Convention, initiated by the Council of Europe, which aims at criminalising offenders who perpetrate cybercriminal activities (Chang, 2020; Mantelero, 2022). In addition, the adoption of these initiatives will foster co-operation within the region, thereby supporting efforts to identify, investigate and prosecute cybercriminals (Velasco, 2022).

Finally, it is crucial to note that recognising that a multidimensional strategy comprising awareness, co-operation, regulation and continual improvement from all stakeholders is necessary to make APAC the least targeted region for cyberattacks. By adopting these policies and promoting a cybersecurity-aware culture, APAC can strengthen its defence against cybercriminals; safeguard its digital infrastructure, businesses and citizens from evolving threats; and reduce the associated risks. Since the threat environment is always changing, it is vital to stress the need for constant communication and preventative measures in building cyber-resilience and stopping cybercrime.

References

Ala-Pietilä, P. and N.A. Smuha (2021) 'A Framework for Global Cooperation on Artificial Intelligence and Its Governance'. *Reflections on Artificial Intelligence for Humanity* 237–265.

Alkhalil, Z., C. Hewage, L. Nawaf and I. Khan (2021) 'Phishing Attacks: A Recent Comprehensive Study and a New Anatomy'. *Frontiers in Computer Science* 3: 563060.

Allioui, H. and Y. Mourdi (2023) 'Unleashing the Potential of AI: Investigating Cutting-Edge Technologies That Are Transforming Businesses'. *International Journal of Computer Engineering and Data Science* 3(2): 1–12.

Al-Maliki, S., A. Qayyum, H. Ali et al. (2023) 'Adversarial Machine Learning for Social Good: Reframing the Adversary as an Ally'. arXiv preprint arXiv: 2310.03614.

Amankwah-Amoah, J., Z. Khan, G. Wood and G. Knight (2021) 'COVID-19 and Digitalization: The Great Acceleration'. *Journal of Business Research* 136: 602–611.

Amnesty International (2022) 'Bangladesh: Government Must Remove Draconian Provisions from the Draft Cyber Security Act'. 31 August. www.amnesty.org/en/latest/news/2023/08/bangladesh-government-must-remove-draconian-provisions-from-the-draft-cyber-security-act/

Ang, B. (2021) 'Singapore: A Leading Actor in ASEAN Cybersecurity'. In Romaniuk, S. and M. Manjikian (eds) *Routledge Companion to Global Cyber-Security Strategy*. Basingstoke: Routledge, pp. 381–391.

Araki, N. (2022) 'Report on the 34th Asia-Pacific Telecommunity Standardization Program Meeting'. *NTT Technical Review*: 81–84.

Argaw, S.T., J.R. Troncoso-Pastoriza, D. Lacey et al. (2020) 'Cybersecurity of Hospitals: Discussing the Challenges and Working towards Mitigating the Risks'. *BMC Medical Informatics and Decision Making* 20: 1–10.

Aslan, , S.S. Aktuğ and M. Ozkan-Okay (2023) 'A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions'. *Electronics* 12(6): 1333.

Association of Southeast Asian Nations (ASEAN) Cybersecurity Cooperation Strategy (2021) 'Broadening and Deepening Cybersecurity Cooperation for a Secure and Resilient ASEAN Cyberspace, 2017–2020'. https://asean.org/wp-content/uploads/2021/08/ASEAN-Cybersecurity-Cooperation-Strategy.pdf

Australian Signals Directorate (ASD) (2024) 'Steps for Organisations to Protect their IT Environment'. www.cyber.gov.au

Bahuguna, A., R.K. Bisht and J. Pande (2020) 'Country-Level Cybersecurity Posture Assessment: Study and Analysis of Practices'. *Information Security Journal: A Global Perspective* 29(5): 250–266.

Benincasa, E. (2020) 'The Role of Regional Organizations in Building Cyber Resilience: ASEAN and the EU'. *Pacific Forum Issues & Insights* 20.

Brain, S. and O. Oyadeyi (2023) 'Funding Crime Online: Cybercrime and Its Links to Organised Crime in the Caribbean'. *Commonwealth Cybercrime Journal* 1(1): 84–110.

Caldwell, M., J.T. Andrews, T. Tanay and L.D. Griffin (2020) 'AI-Enabled Future Crime'. *Crime Science* 9(1): 1–13.

Cascavilla, G., D.A. Tamburri and W.J. van den Heuvel. (2021) 'Cybercrime Threat Intelligence: A Systematic Multi-Vocal Literature Review'. *Computers & Security* 105: 102258.

Chakravarti, J. (2023) 'Phishing Attacks Rise Sharply in Southeast Asia'. Bank Info Security, 27 July. www.bankinfosecurity.asia/phishing-attacks-rise-sharply-in-southeast-asia-a-22669.

Chang, L.Y. (2020) 'Legislative Frameworks against Cybercrime: The Budapest Convention and Asia'. In Holt, T. and A. Bossler (eds) *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. London: Palgrave, pp. 327–343.

Check Point Research (2023) 'Global Cyberattacks Continue to Rise with Africa and APAC Suffering Most'. 27 April. https://blog.checkpoint.com/research/global-cyberattacks-continue-to-rise/

Christine, D. and M. Thinyane (2020) 'Cyber Resilience in Asia-Pacific: A Review of National Cybersecurity Strategies'. Macau: United Nations University.

Commonwealth Secretariat (2022) 'Commonwealth Experts Meet in Singapore to Explore Solutions to Increasing Cyber Risks in Asia'. 29 September. https://thecommonwealth.org/news/commonwealth-experts-meet-singapore-explore-solutions-increasing-cyber-risks-asia

Centre for Strategic and International Studies (CSIS) (nd) 'Significant Cyber Incidents'. Accessed online at: www.csis.org/programs/strategic-technologies-program/significant-cyber-incidents (accessed 29 October 2023).

Cybersecurity Ventures (2022) '2022 Official Cybercrime Report'. https://s3.ca-central-1.amazonaws.com/esentire-dot-com-assets/assets/resourcefiles/2022-Official-Cybercrime-Report.pdf

Cybersecurity Ventures (2023) 'Cybercrime to Cost the World $9.5 Trillion USD Annually in 2024'. www.esentire.com/web-native-pages/cybercrime-to-cost-the-world-9-5-trillion-usd-annually-in-2024?utm_medium=email&utm_source=pardot&utm_campaign=autoresponder

Development Dimensions International (DDI) (2023) 'Global Leadership Forecast 2023'. www.ddiworld.com/global-leadership-forecast-2023?utm_source=google&utm_medium=display&utm_campaign=|Brand|DA|GLF23_Launch|EU|EN|&gclid=EAIaIQobChMIn6y0kfLegQMVHUf2CB2Xlw_xEAEYASAAEgLpP_D_BwE

Dimitrov, W. (2020) 'The Impact of the Advanced Technologies Over the Cyber Attacks Surface'. In Artificial Intelligence and Bioinspired Computational Methods: Proceedings of the 9th Computer Science Online Conference: 509–518.

Diogenes, Y. and E. Ozkaya (2019) *Cybersecurity–Attack and Defense Strategies: Counter Modern Threats and Employ State-of-the-Art Tools and Techniques to Protect Your Organization against Cybercriminals*. Birmingham: Packt Publishing Ltd.

Dwivedi, Y.K., D.L. Hughes and C. Coombs (2020) 'Impact of COVID-19 Pandemic on Information Management Research and Practice: Transforming Education, Work and Life'. *International Journal of Information Management* 55: 102211.

Engstrom, D.F., D.E. Ho, C.M. Sharkey and M.F. Cuéllar (2020) 'Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies'. NYU School of Law Public Law Research Paper 20–54.

Falk, R. (2022) 'First Australian Court Judgments on Cyber Security'. AICD, 8 June. www.aicd.com.au/economic-news/world/global-risk-report/first-australian-court-judgments-on-cyber-security.html

Ferdous, J., R. Islam, A. Mahboubi and M.Z. Islam (2023) 'A State-of-the-Art Review of Malware Attack Trends and Defense Mechanism'. *IEEE Access* 11: 121118–121141.

Gan, G.S. (2024) 'Filling the Gaps: The Story of APAC's Cyber Security Capacity Building'. Kaspersky www.kaspersky.com/about/policy-blog/filling-the-gaps-the-story-of-apacs-cyber-capacity-building

Germanos, G. and N. Georgiou (2022) 'How Did Cybercriminals "Survive" during the Pandemic?' *Urban Crime, An International Journal* 3(2): 110–123.

Gullapalli, V. (2023) 'Why Is the Asia Pacific Region a Target for Cyber Crime & What Can Be Done'. Check Point Research, 4 August. www.cybertalk.org/2023/08/04/why-is-the-asia-pacific-region-a-target-for-cyber-crime-what-can-be-done/

Herko, T. (2023) 'The INTERPOL Global Complex for Innovation in Singapore: A Personal Retrospective'. *Belügyi Szemle* 71(3. ksz): 45–55.

Inamdar, N. (2023) 'National Cyber Security Strategy 2023 to Be Released Soon'. *Hindustan Times*, 13 June www.hindustantimes.com/cities/pune-news/national-cyber-security-strategy-2023-to-be-released-soon-101686596627065.html

Khan, J.I., J. Khan, F. Ali et al. (2022) 'Artificial Intelligence and Internet of Things (AI-IoT) Technologies in Response to the COVID-19 Pandemic: A Systematic Review'. *IEEE Access* 10: 62613–62660.

Kumar, S. (2021) 'The Missing Piece in Human-Centric Approaches to Cybernorms Implementation: The Role of Civil Society'. *Journal of Cyber Policy* 6(3): 375–393.

Laitinen, M. and S. Armstrong-Smith (2022) 'Tackling Cybercrime and Ransomware Head-On: Disrupting Criminal Networks and Protecting Organisations'. *Cyber Security: A Peer-Reviewed Journal* 5(3): 190–205.

Lallie, H.S., L.A. Shepherd and J.R. Nurse (2021) 'Cyber Security in the Age of COVID-19: A Timeline and Analysis of Cyber-Crime and Cyber-Attacks during the Pandemic'. *Computers & Security* 105: 102248.

Malja, M. and M. October (2022) 'The Lanzarote Convention: National Action Plan for the Years 2022–2025'. http://urn.fi/URN:ISBN:978-952-00-5443-4

Mantelero, A. (2022) 'Regulating AI'. In *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*. The Hague: TMC Asser Press, pp. 139–183.

Medina, L. and F. Schneider (2019) 'Shedding Light on the Shadow Economy: A Global Database and the Interaction with the Official One'. CESIFO Working Paper 7981.

Mizrak, F. (2023) 'Integrating Cybersecurity Risk Management into Strategic Management: A Comprehensive Literature Review'. *Research Journal of Business and Management* 10(3): 98–108.

Positive Technologies (2023) 'Cybersecurity Threatscape of Asia: 2022–2023'. 12 September. www.ptsecurity.com/ww-en/analytics/asia-cybersecurity-threatscape-2022-2023/

Quimba, F.M.A. and M.A.A. Barral (2020) 'Exploring the Feasibility of Content Analysis in Understanding International Cooperation in APEC'. PIDS Discussion Paper 2020–58.

Roberts, W. (2022) 'Role of IGF and APrIGF in reference to Libraries in Nepal'. *Access: An International Journal of Nepal Library Association* 1(1): 139–142.

Rui, W. (2023) 'ASEAN Cybersecurity Policy and China-ASEAN Cooperation' *China International Studies* 98: 55.

Runde, D.F., C.M. Savoy and O. Murphy (2020) 'Post-Pandemic Infrastructure and Digital Connectivity in the Indo-Pacific'. Brief, 2 November. Washington, DC: CSIS.

Ryan, M. (2021) *Ransomware Revolution: The Rise of a Prodigious Cyber Threat* . Berlin/Heidelberg: Springer.

Sarowa, S.K., B. Bhanot and V. Kumar (2022) 'Analysis of Cyber Attacks and Cyber Incident Patterns over APCERT Member Countries'. In 4th International Conference on Artificial Intelligence and Speech Technology (AIST): 1–6.

Schmidt, E., B. Work, S. Catz et al. (2021) 'National Security Commission on Artificial Intelligence (AI) Final Report'. https://digital.library.unt.edu/ark:/67531/metadc1851188/

Schneider, V. and R. Werle (2023) 'International Regime or Corporate Actor? The European Community in Telecommunications Policy'. In Dyson, K. and P. Humphreys (eds) *The Political Economy of Communications*. Basingstoke: Routledge, pp. 77–106.

Singh, L. (2022) 'Cyber Crime, Cyber Resilience and Security Strategy in Post Pandemic World'. *Supremo Amicus* 28(305): 1–11.

Skouby, K.E., P. Dhotre, I. Williams and K. Hiran (2022) *5G, Cybersecurity and Privacy in Developing Countries*. Boca Raton, FL: CRC Press.

Slayton, R. and B. Clarke (2020) 'Trusting Infrastructure: The Emergence of Computer Security Incident Response, 1989–2005'. *Technology and Culture* 61(1): 173–206.

Smail, J. (2023) 'The Top 10 Data Breaches'. Cybers Security Hub, 15 September www.cshub.com/attacks/articles/the-top-10-apac-data-breaches.

Solar, C. (2023) *Cybersecurity Governance in Latin America: States, Threats, and Alliances*. Albany, NY: State University of New York Press.

The Asset (2023) 'Asia-Pacific Deepfake Incidents Surge'. 30 November. www.theasset.com/article/50495/asia-pacific-deepfake-incidents-surge

Tien, H.M. and T.J. Cheng (2016) *The Security Environment in the Asia-Pacific*. Basingstoke: Routledge.

United Nations Office on Drugs and Crime (UNODC) (nd) 'International and Regional Instruments'. E4J University Module Series. www.unodc.org/e4j/en/cybercrime/module-3/key-issues/international-and-regional-instruments.html

Van der Sloot, B. and Y. Wagensveld (2022) 'Deepfakes: Regulatory Challenges for the Synthetic Society'. *Computer Law & Security Review* 46: 105716.

Velasco, C. (2022) 'Cybercrime and Artificial Intelligence. An Overview of the Work of International Organizations on Criminal Justice and the International Applicable Instruments'. *ERA Forum* 23(1) 109–126.

Wong, P., R. Marles and C. O'Neil (2024) 'Cyber Sanctions in Response to Medibank Private Cyber-attack'. Release, 23 January. www.foreignminister.gov.au/minister/penny-wong/media-release/cyber-sanctions-response-medibank-private-cyber-attack

Xu, L., Q. Guo, Y. Sheng et al. (2021) 'On the Resilience of Modern Power Systems: A Comprehensive Review from the Cyber-Physical Perspective'. *Renewable and Sustainable Energy Reviews* 152: 111642.

Zeadally, S., E. Adi, Z. Baig and I.A. Khan (2020) 'Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity'. *IEEE Access* 8: 23817–23837.

Zhang, Z., H. Al Hamadi, E. Damiani et al. (2022) 'Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research'. *IEEE Access* 10: 93104–93139.

Zhao, P., Z. Cao, D.D. Zeng et al. (2021) 'Cyber-Resilient Multi-Energy Management for Complex Systems'. *IEEE Transactions on Industrial Informatics* 18(3): 2144–2159.

About the authors

Dr Olajide O. Oyadeyi is a passionate economist who specialises in the economic dynamics of cybercrime to inform strategic solutions and policies in combating digital threats worldwide. He is dedicated to unravelling the economic incentives driving cybercriminal activities and developing innovative frameworks to mitigate risks and enhance cybersecurity resilience on a global scale.

Oluwadamilola A. Oyadeyi is a passionate researcher on a range of societal issues. She has served as an independent researcher and writer, contributing her expertise to publications covering a wide array of topics, from public health to energy, climate change and cybercriminal activities.

Rofiat O. Bello is a seasoned legal expert who specialises in the multifaceted legal landscape surrounding cybercrime on a global scale, adept at navigating intricate regulatory frameworks and addressing emerging challenges in cybersecurity law. Passionate about exploring the intersections of technology and jurisprudence, she is committed to advancing legal solutions that safeguard individuals and organisations against digital threats in an ever-evolving digital ecosystem.

The Commonwealth

Commonwealth Cyber Journal \ 161
ISSN 2959-3018 (print), ISSN 2959-3026 (online)
thecommonwealth.org/cyber-journal

Towards a Victim-Centred Approach? Reflections on Existing Cybercrime Instruments and the Draft United Nations Convention on Cybercrime

Brenda Mwale[1]

Abstract

In recent years, increasing attention to cybercrime and its impacts on society has led to an unprecedented focus on the prosecution of offenders. Indeed, anti-cybercrime conventions such as the Council of Europe Convention on Cybercrime and the African Union Convention on Cyber Security and Personal Data provide a good framework for addressing cybercrime by setting out provisions on substantive criminal law, procedural law and international co-operation, which are essential for prosecuting offenders. Yet, while these provide adequate frameworks, they barely focus on victims of cybercrime (those who suffer harm as a result of cybercrime). Unlike for traditional crimes, perpetrators of cybercrime do not require physical proximity to the victim, and they can anonymously reach a significant number of victims with limited detectability. As a result, the way in which cybercrimes occur and how their consequences manifest challenge the way we traditionally think about victims of crime.

This points to a need to specifically address the plight of victims of cybercrimes. Currently, negotiations on a binding international cybercrime treaty under the UN framework provide new hope for victims of cybercrime: the current draft not only recognises victims but also contains specific provisions relating to them. If adopted, the convention will be the first international cybercrime convention to adopt a victim-centred approach.

[1] Post-doctoral Fellow, Faculty of Law, University of Pretoria. Email: mwale17@gmail.com.

1. Introduction

As society continues to be digitalised, so too does the nature of crimes – and their victims. While the actual number of cybercrime victims is unknown, by the year 2020 alone it was estimated that the global cost of cybercrime was US$ 1 trillion per year, a 50 per cent increase on figures reported in 2018 (Scroxton, 2020). Estimates indicate that, in 2018, the global cost was as much as $600 billion per year (Lewis, 2018, p. 6). These figures are particularly concerning because it is generally known that the global scale of cybercrime is often underestimated, frequently because of underreporting.

Despite these worrying figures, the plight of victims of cybercrime (those who suffer harm as a result of cybercrime) has been ignored. Today, the primary focus of cybercrime instruments is to punish perpetrators by creating cyber-specific offences and imposing penalties on cybercrimes. While the needs of cybercrime victims may be partially met through the prosecution of cyber-offenders, their specific needs are often overlooked. As a result, their plight is often relegated to a secondary consideration. Vincent (2017, p. 27) highlights some of the reasons why criminal law overlooks victims of cybercrime:

> First, and perhaps most surprisingly to many..., victims and their harms are best of only marginal interest to... criminal law. Second, core features of criminal law doctrine are conceptually incompatible with recognizing and adjudicating cybercrime. Consequently, for largely doctrinal and conceptual reasons, criminal law makes a very poor ally for victims of cybercrime.

Besides, some forms of cybercrimes may be deemed victimless crimes. These could involve phishing offences, in which an unsuspecting victim unintentionally joins a criminal network (Halder, 2022, p. 6). In such cases, if victims' claims of victimhood are not supported by relevant facts, the criminal justice system may choose to dismiss them, place the responsibility on them or treat them like perpetrators.

These factors highlight the need for renewed reflection on how cybercrime instruments respond to the unique circumstances of victims of cybercrime. For a long time, there has been little scrutiny on this topic. However, since early 2022, UN member states have been negotiating a cybercrime treaty that has the potential to reshape the criminal justice approach to victims of cybercrime. This is because the current draft text of the convention (published in February 2024) contains specific provisions on victims, relating to the protection of victims who are witnesses, assistance to and protection of victims, mutual legal assistance, remedies such as recovery and return of proceeds of cybercrime, preventive measures and technical assistance. If adopted, the treaty has the potential to cement certain guarantees for cybercrime victims.

In this context, this article argues that legal measures aimed at addressing cybercrime, as provided in the draft text of the convention, should focus on a victim-centred approach that addresses the unique needs of cybercrime and its victims. Accordingly, as a preliminary matter, the article explains what cybercrime entails, who the victims are and

the impact of cybercrime on its victims, and highlights the specific needs of those victims. Building on this, the article analyses their legal position in the context of the current anti-cybercrime frameworks and the current draft text of the UN convention on cybercrime.

2. What is cybercrime?

A critical starting point in looking at the question of victims of cybercrime involves examining what cybercrime entails. This is because a clear definition of cybercrime has an impact on the definition of victims of cybercrime and measures to respond to their plight. That said, while cybercrime instruments aim to prevent, investigate and punish cybercrimes, there is no universally accepted definition of the term. Cybercrimes are conceptualised differently across different jurisdictions, with significant variations on what constitutes a criminal offence. Besides, most international and regional instruments shy away from defining cybercrime.

Nonetheless, at its broadest, it could be argued that the notion signifies illegal activities committed through information and communication technology (ICT). Based on this understanding, Thomas and Loader (2000, p. 3) conceptualise and define cybercrimes as 'computer-mediated activities which are either illegal or considered illicit by certain parties and which can be conducted through global electronic networks.' In addition, Brenner (2007, p. 386) notes that cybercrimes comprise 'the use of computer technology to commit crime; to engage in activity that threatens a society's ability to maintain internal order.'

A popular distinction is drawn between two categories of cybercrimes based on the role that technology plays in committing a crime. The first category, computer-assisted crimes, involves crimes that occur because cyberspace allows them to be committed in new ways. Crimes that predate the internet, like fraud, hate speech and money laundering, can occur even without the internet, but the internet gives them a new life. In such cases, computer systems and networks merely facilitate the commission of existing crimes (Gillespie, 2016). On the other hand, computer-focused crimes – the second category – came into existence with the advent of computer technology and cannot be perpetrated without the use of computer systems and networks. A classic example of a computer-focused crime would be hacking.

The main criticism of this distinction is that 'technological advancements have... blurred the distinction between assisted and focused' (Gillespie, 2016, p. 9). Thus, it is argued that, while this classification is helpful, it may be limiting from a criminal law perspective as it focuses on the technology at the cost of the relationship between the perpetrator and the victim (Yar, 2006).

For these reasons, alternative approaches that are slightly more nuanced have been proposed. Some of these are broader and account for the role of technology in the perpetration of crime or specific types of offences (Phillips et al., 2022). For instance, Wall (2010, in ibid., p. 385) proposes a three-tier classification that distinguishes between:

1. Cyber-dependent crimes or true cybercrimes,
2. Cyber-enabled crimes or hybrid crimes; and
3. Cyber-assisted crimes or the use of computers in traditional crime.

It can be argued that this approach does not add new categories per se but extends the two-category system identified above. Other approaches focus on different categories of offences. For instance, the Council of Europe (COE) Convention on Cybercrime (2001) sets out four broad categories of cybercrime:

1. Offences against the confidentiality, integrity and availability of computer data and systems;
2. Computer-related offences such as fraud and forgery;
3. Content-related offences – such as child pornography; and
4. Offences related to infringements of copyright and related rights.

These categories focus on a range of online harms that can occur as a result of online conduct (e.g., hacking) or material (e.g., child pornographic material). Rather than explicit definitions of cybercrimes, the use of these broader categories and the above classification systems has gained popularity (Phillips et al., 2022). The problem, in the context of the topic at hand, is that lack of uniformity can pose challenges to determining who the victims of cybercrimes are across different jurisdictions.

3. Who are the victims of cybercrimes?

It follows from the above that defining victims of cybercrimes is not straightforward. Like with the concept of 'cybercrime,' there is no universally accepted definition of the term 'victim of cybercrime,' and it may be difficult to envisage an 'ideal' victim because of the breadth of prohibited acts that constitute 'cybercrime' across different legal instruments. However, as a point of departure, one can seek guidance from international instruments focusing on victims of crime. An oft-quoted reference point is the United Nations Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power (1985), which defines victims as:

> ... persons who, individually or collectively, have suffered harm, including physical or mental injury, emotional suffering, economic loss or substantial impairment of their fundamental rights, through acts or omissions that are in violation of criminal laws operative within Member States, including those proscribing criminal abuse of power (para. 1).

According to the Declaration, a person may be regarded as a victim irrespective of whether the perpetrator of the crime is identified, apprehended, prosecuted or convicted and whether there is a familial relationship between the perpetrator and the

victim (para. 2). Where appropriate, the definition extends to 'the immediate family or dependants of the direct victim and persons who have suffered harm in intervening to assist victims in distress or to prevent victimization' (ibid.).

Drawing on the above, two key points are critical. First, victims are defined by reference to the harm suffered. The focus on the harm suffered is emphasised by Suryanto et al. (2020, p. 155), who conceptualize victims as 'those who have been harmed both materially and non-materially as a result of cybercrime.' Second, two categories of victims can be identified: direct victims (those who suffered harm resulting from the cybercrime) and indirect victims (immediate family or dependants of the direct victim). In general, primacy is given to those who suffer harm as victims. However, where appropriate, there may be circumstances when immediate family or dependents are considered victims, as in the above definition.

It is noteworthy that, while the Declaration focuses on persons who individually or collectively suffer, 'victims of cybercrime [can] range from individuals and communities to entire businesses and governments' (Wilkinson, 2023). Thus, notwithstanding the difficulties in defining cybercrime and the lack of a common definition thereof, it can be concluded that a victim of a cybercrime is anyone who suffers harm resulting from a cybercrime. Such crime may be cyber-dependent or cyber-enabled, or arise from the list of cyber-offences in a cybercrime instrument.

4. What are the impacts of cybercrimes on their victims?

The focus on the harm suffered leads us to a discussion on the impact of cybercrimes on their victims. Generally, the negative impacts of cybercrimes are material and non-material and can fall into the categories highlighted below.

4.1 Financial impacts

Cybercrime can have devastating financial impacts on businesses and individuals. Even a single cyber-incident can result in significant financial losses. For instance, in 2017, the WannaCry ransomware cyberattack affected around 230,000 computers globally, causing a financial loss of US$4 billion across the globe (Kaspersky, nd). Besides, as noted earlier, the global financial cost of cybercrime scales up to $1 trillion a year. However, these estimates reflect the aggregate cost to countries, not to individuals or companies, and may not be reflective of the cost to individual victims (Lewis, 2018).

4.2 Reputational impact

In addition to causing financial losses, cybercrime can lead to reputational harm at different levels. For instance, at an organisational level, cybercrime can affect a company's image, erode customer trust, damage public perceptions and reduce

business opportunities (Agrafiotis et al., 2018). At the individual level, a cyber-incident that leaks personal information can have a negative reputational impact when the information leaked damages the individual's reputation.

4.3 Psychological and emotional impacts

At the individual level, cybercrime can result in psychological and emotional impacts on its victims. Often, victims have feelings ranging from anger, outrage, anxiety and fear to a total loss of trust in information technology. Victims can also feel ashamed, vulnerable and powerless, leading to depression. In many cases, victims of online abuse, such as cyberstalking, doxing, online harassment, non-consensual dissemination of intimate images and so on, often blame themselves for what happened.

4.4 Disruptive impacts

Cybercrimes, such as those aimed at causing data or system interference, can be disruptive, causing operational disruptions that result in downtime, loss of productivity, loss of access to critical services and delays in service delivery. For example, ransomware attacks can encrypt files on a computer system, rendering computer files inaccessible and unusable. As a result, victims may spend a lot of time trying to recover their data, which sometimes cannot be recovered.

4.5 Physical impacts

Cybercrimes can also result in physical damage to physical assets (such as computer hardware, infrastructure, etc.). There are 'targeted and specific intrusions capable of creating functional and even physical damage' (Rid and McBurney, 2012, p. 8). Damage to physical infrastructure can also have cascading effects on individuals, organisations and society as a whole (Agrafiotis et al., 2016).

5. What are the needs of victims of cybercrime?

By all accounts, victims' needs can be diverse, as the impacts of cybercrime on victims can range from material (e.g., financial loss) to non-material (e.g., reputational damage, psychological and emotional impacts). Some victims may have suffered financial loss and require compensation. When victims still feel the negative consequences of cybercrime, they may develop a more punitive stance and pursue prosecution of the offender (Pemberton and Vanfraechem, 2015). However, for others there can be more pressing matters than retribution. These may include an apology, assistance and protection, recognition and condemnation of the harm, guarantees of safety, restoration to the situation preceding the cyber-incident, prevention of the cybercrime's recurrence or participation in criminal proceedings.

Some may require immediate responses, such as removing illegal content posted online, while others may require responses that consider their particular vulnerabilities and needs. Thus, special consideration should be made of the fact that 'victimhood varies depending on a number of identified dimensions, including vulnerability aspects, psychological perspectives, [and] age-related differences' (Sikra et al., 2023, p. 28). This means that a wide range of measures are required to respond to the diverging needs of cybercrime victims.

6. The legal response: international responses to cybercrime and its victims

Despite the growing number of cybercrime victims, the impact of cybercrime on its victims and the specific needs of victims, the predominant focus of cybercrime instruments is on the prosecution of offenders by creating new cyber-offences or adapting existing offences to address the challenge of cybercrime. As the ensuing discussion demonstrates, the existing cybercrime instruments 'recognize' the harm to victims only when the crime is defined as a criminal offence (Vincent, 2017, p. 31). To further elaborate on the situation and the prospects of a victim-centred approach, this section analyses the relevant provisions of two multilateral cybercrime conventions (the COE Convention on Cybercrime and the African Union Convention on Cyber Security and Personal Data Protection) and the draft text of the UN convention on cybercrime (published on 6 February 2024).

6.1 Council of Europe Convention on Cybercrime

In 2001, the COE Convention on Cybercrime (the Budapest Convention) was adopted as the first international convention to deal with cybercrimes. With the aim of establishing 'a common criminal policy aimed at the protection of society against cybercrime,' highlighted in its Preamble, the convention deals with issues of substantive criminal law, procedural law and international co-operation. It takes a retributive approach that focuses on prosecuting the following offences: illegal access, illegal interception, data interference, system interference, misuse of devices, computer-related forgery, computer-related fraud, offences related to child pornography and offences related to infringements of copyright and related rights.

In terms of procedural law, Article 15 of the Convention provides that states parties should ensure that the establishment, implementation and application of the procedural powers are subject to the conditions and safeguards laid down in domestic laws. The Convention further provides in Article 15(3) that 'to the extent that it is consistent with the public interest, in particular the sound administration of justice, each state party shall consider the impact of [such] powers and procedures upon the rights, responsibilities and legitimate interests of third parties.' The Explanatory Report to the Convention notes that the initial consideration is given to the sound administration of justice and

other public interests, such as the interest of victims (COE, 2001a, para. 148). In this regard, victims' interests are implicitly recognised. Thus, while the Convention provides an adequate legal basis for prosecuting cybercrime, victims of cybercrime are considered to be of only secondary interest in terms of the conditions and safeguards that states should place while implementing procedural powers in the Convention.

Nonetheless, the Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence (2022) recognises victims of cybercrime**.** It recognises in its Preamble 'the growing number of victims of cybercrime and the importance of obtaining justice for those victims.' The Second Protocol is one of the first examples of a cyber-convention recognising victims. However, although it constitutes a step forward, it is also worth noting that there are no detailed provisions on victims in the Protocol.

6.2 African Union Convention on Cyber Security and Personal Data Protection

In June 2014, the African Union Assembly adopted the African Union Convention on Cyber Security and Personal Data Protection (the Malabo Convention) as a regional instrument on cybersecurity, cybercrime and data protection. For many years after its adoption, the Convention did not receive the required number of ratifications to enter into force. It did so on 8 June 2023 following the deposit of the 15th Instrument of Ratification by Mauritania.

Similar to the Budapest Convention, the Malabo Convention sets out provisions on substantive criminal law, procedural law and international co-operation and takes a retributive approach. It lays down three broad categories of prohibited conduct that states should criminalise: (i) attacks on computer systems; (ii) computerised data breaches; and (iii) content-related offences. However, unlike the Budapest Convention and its Second Additional Protocol, the Malabo Convention does not refer to victims explicitly or impliedly. Save for setting out substantive offences, which could by implication cater to victims' needs through retribution, the Malabo Convention does not refer to victims in any way. Of course, victims can resort to instituting criminal proceedings against offenders. However, the assumption that punishing perpetrators alone will cater to victims' needs is flawed.

6.3 Draft UN convention on cybercrime

For a long time, the Budapest Convention has been regarded as the most important international agreement on cybercrime and electronic evidence and the only international convention on the subject. Initially, it was open for signature only to the COE states and four observer states that participated in the negotiations. But today, any state prepared to implement its provisions and engage in co-operation can accede to it. Despite this, the lack of a UN convention on cybercrime (open to all member states) is still glaring.

In 1998, Russia introduced a draft resolution at the General Assembly that mentioned the need to prevent the misuse of information technology for criminal purposes. But progress towards adopting a UN treaty has been slow. It was only in 2019 that the General Assembly, pursuant to Resolution 74/247, decided to establish an open-ended ad hoc intergovernmental committee of experts, representative of all regions, to elaborate a comprehensive international convention on countering the use of information and communications technologies for criminal purposes (para. 2).

In early 2022, UN member states convened their first session to negotiate a binding international cybercrime treaty draft pursuant to the work plan for the delivery of the mandate of the Ad Hoc Committee. Negotiations continued into early 2024; the concluding session was held in New York from 29 January to 9 February. The concluding session ended with no consensus on the scope of the convention, including on what crimes it should cover and on key provisions on criminalisation, international co-operation and human rights safeguards. As a result, a reconvened concluding session will be held at a future date. While proponents may view the lack of consensus as a dream deferred, it is an opportunity to reflect deeply on the provisions of the current draft text of the convention.

6.3.1 The scope of the draft text

During the first session of the Ad Hoc Committee, several member states submitted statements on what the scope of the convention should entail. South Africa captured the current state of affairs concerning the regulation of cybercrime by submitting that:

> The international system in its current form is not equipped to deal with the growing scourge of cybercrime, thus necessitating that the world unites in formulating a true international instrument that will protect the victims of crimes committed in cyberspace and guarantee maximum protection and legal remedies (South Africa, 2021, p. 1).

Several states agreed that the convention should focus primarily on substantive criminal law, criminal procedure and international co-operation. While these three areas are the main focus of existing cybercrime instruments, some states also drew attention to the need to focus on victims. For Switzerland, the convention should provide for a co-ordinated approach in the fight against cybercrime and create a common understanding of what cybercrime offences entail and a framework for international co-operation 'to protect ICT users and to obtain justice for the victims of cybercrime' (Switzerland, 2021, p. 2). In addition to provisions on criminalisation and co-operation, the EU and its member states stressed that the convention should also 'comply with international human rights standards and strive to fight cybercrime most effectively and thus protect victims' (EU, 2022, p. 2). Chile's view on the scope, objectives and structure (elements) of the convention focused on, among others, the preventive role of the convention. In Chile's view, the convention should 'promote victim-centered prevention strategies [that] deal with interpersonal cybercrimes' (Chile, 2021).

Prior to the concluding session, a draft text of the convention provided that the convention would apply to the prevention, investigation and prosecution of the offences set out therein, including the freezing, seizure, confiscation and return of the proceeds of such offences (UN General Assembly, 2023). It also provided that the convention would apply to the collecting, obtaining, preserving and sharing of evidence in electronic form. However, as noted above, no consensus was reached on the scope of the convention's application during the concluding session.

The current draft text, published on 6 February 2024 (UN General Assembly, 2024), while not addressing the scope of the convention, lists a number of prohibited conducts. These include illegal access, illegal interception, interference with electronic data, interference with an ICT system, misuse of devices, ICT system-related forgery, ICT system-related theft or fraud, offences related to online child sexual abuse or child sexual exploitation material, solicitation or grooming for the purpose of committing a sexual offence against a child, non-consensual dissemination of intimate images and laundering of proceeds of crime (Articles 6–16). While mirroring some of the offences laid down in the Budapest and Malabo Conventions, the proposed offences go beyond the list of core cybercrimes and may broaden the list of who victims of cybercrime are.

6.3.2 Recognition of victims in the Preamble

In the early stages of the treaty negotiations, the chair of the Ad Hoc Committee (2022a, p. 11) came up with guiding questions for invited delegations to consider and address in their interventions, including questions on victims. The following guiding questions were put across:

> Should the convention contain a provision on the assistance to and protection of victims? If yes, which factors of protection are important to include in such a provision, and what level of detail, in terms of definitions and description of related procedures, should be expected? What role should victims and reporting persons have? Would the committee like to follow the formulation of UNTOC [United Nations Convention Against Transnational Organized Crime] (article 25)?

These questions provided a useful framework for discussions on victims of cybercrime.

To begin with, the Preamble of the draft text emphasises the need to protect society against cybercrime. It also recognises 'the increasing number of victims of cybercrime, the importance of obtaining justice for those victims and the necessity to address the needs of persons in vulnerable situations in measures taken to prevent and combat the offences covered by [the] Convention.' This recognition represents a significant shift in how states view the objects and purposes of cybercrime instruments and has been welcomed by some stakeholders.

The CyberPeace Institute, for instance, noted in its submission to the Ad Hoc Committee that the recognition of victims was a 'key statement as the main purpose of a new international treaty on cybercrime should be to protect and bring remedy to its victims

through evidence-led accountability, allowing those affected by cybercrime to seek redress and for measures to prevent their re-victimisation'. However, the Institute argued that the Preamble should consider the various types of harm that cybercrime can cause by highlighting the distinct impacts of cybercrime that may be felt by those disproportionately targeted or affected in cyberspace. This article argues that, although the Preamble does not provide for the various types of harm, it acknowledges the needs of persons in vulnerable situations in measures taken to prevent and combat the offences covered by the convention.

6.4 Specific provisions relating to victims

In addition to recognising victims in the Preamble, the draft text contains provisions relating to victims. It specifically provides for the protection of victims who are witnesses, assistance to and protection of victims, mutual legal assistance, remedies such as recovery and return of proceeds of cybercrime, preventive measures and technical assistance in relation to victims.

6.4.1 Protection of victims who are witnesses

First, the draft text provides for the protection of victims who are witnesses. Article 33 requires states parties to establish measures to protect witnesses who give testimony, provide information concerning offences established in the convention or co-operate with investigative or judicial authorities. Protective measures may include establishing physical protection procedures and evidentiary rules to ensure the witness' safety when testifying. States parties should also consider entering into agreements or arrangements with other states for the relocation of witnesses. Article 33(4) states that these measures also apply to victims insofar as they are witnesses and are based on Article 32 of the United Nations Convention Against Corruption (UNCAC).

Practically speaking, the idea of victims who are witnesses of cybercrimes sounds complex. When someone thinks of a 'witness,' what comes to mind is someone who saw a crime take place. Thus, in the context of cybercrimes, it may be difficult to envision a victim who is a witness given that cybercrimes occur via computers and computer systems and can result in online harms. As Wilkinson and Swali (2022) put it, a 'witness' in cyberspace is more ambiguous than a 'victim.' Nonetheless, while Article 33 does not define who 'witnesses' are, it limits its scope to witnesses who give testimony or provide information concerning offences established in the convention. Although not definitive in describing cybercrime witnesses, international conventions such as the UNCAC may be 'a productive starting point, demonstrating the merit of adapting language that enjoys international consensus in existing anti-crime frameworks' (ibid.).

On the other hand, states may need to consider the specific realities of cybercrime witnesses and victims. Most importantly, protective measures should go beyond physical protection procedures and evidentiary rules to ensure the safety of witnesses when

testifying. This is because there are many ways in which victims and witnesses can be intimidated in a digital environment that require cyber-specific measures. For instance, victims who are witnesses of cybercrimes may require an assurance that their data will be protected while giving digital evidence.

6.4.2 Assistance to and protection of victims

Second, Article 34 of the draft text makes provision for the assistance and protection of victims. To begin with, states parties are required to take appropriate measures to provide assistance and protection to victims of cybercrime, especially in cases of threat of retaliation or intimidation. States parties are also required to establish appropriate procedures to provide access to compensation and restitution for those victims, subject to their domestic laws. Indeed, compensation and restitution are key pillars of victim assistance as they acknowledge victims' losses by providing financial relief. In practice, however, particular difficulties may arise from the fact that a single cyber-incident can result in multiple victims spread across different jurisdictions, and multiple claims for compensation instituted in several states could raise jurisdictional issues. Therefore, it would be helpful if the provision on compensation and restitution is tied to a requirement for states to co-operate in cases where cybercrimes traverse borders.

On criminal justice matters, states parties are required, subject to domestic laws, to consider the views and concerns of victims at appropriate stages of criminal proceedings against offenders in a manner not prejudicial to the rights of the defence. This provision gives wide discretion to domestic courts to determine the appropriate stage at which victims can participate in criminal proceedings. It must be noted, however, that expressing 'views and concerns' is not the same as giving evidence. Although the views and concerns of victims may assist courts in approaching evidence, they do not form part of trial evidence (ICC, 2014). Thus, unless victims are witnesses, as provided for under Article 33, their participation in criminal proceedings is limited to expressing their views and concerns.

Special provision is also made for the assistance and protection of children. Article 34 covers victims of offences related to online child sexual abuse or child sexual exploitation material; solicitation or grooming for the purpose of committing a sexual offence against a child; and non-consensual dissemination of intimate images. States parties are required to take appropriate measures to assist such victims, including their 'physical and psychological recovery, in cooperation with relevant international organizations, non-governmental organizations, and other elements of civil society' (Article 34(4)). Many would argue that it is important that these actions are undertaken in collaboration with relevant stakeholders such as international organisations, non-governmental organisations and civil society because these are leading providers of services to victims of various crimes. Even so, including other stakeholders, such as technology companies, is crucial, given their technical expertise.

It is also important to note that the protection of child victims of sexual abuse and sexual exploitation can be improved by aligning the relevant provisions of the draft text with the minimum standards for the protection of children outlined in the UN Convention on the Rights of the Child (UNCRC) 1989. Specifically, Article 39 provides that:

> States Parties shall take all appropriate measures to promote physical and psychological recovery and social reintegration of a child victim of: any form of neglect, exploitation, or abuse; torture or any other form of cruel, inhuman or degrading treatment or punishment; or armed conflicts. Such recovery and reintegration shall take place in an environment which fosters the health, self-respect and dignity of the child.

In cases where children are victims of sexual abuse and sexual exploitation, Article 34 of the UNCRC provides that states parties undertake to take measures to prevent (i) the inducement of a child to engage in unlawful sexual activity and (ii) the exploitative use of a child in prostitution, other unlawful sexual practices or pornographic performances. These protective measures should be guaranteed both offline and online (UN Committee on the Rights of the Child, 2021). To further assist and protect child victims, the United Nations Children's Fund argues that the proposed cybercrime convention can be used to further strengthen the protection of children by adopting special measures such as child-friendly practices in the criminal justice system and different forms of platforms for compensating child victims (UNICEF, 2022).

In applying the above measures, the draft text provides that states parties should take appropriate measures that consider the age, gender, particular circumstances and needs of victims, including the particular circumstances and needs of children. Further, states parties should take steps, in accordance with domestic laws, to ensure compliance with requests to take down such content relating to online child sexual abuse or child sexual exploitation material; solicitation or grooming for the purpose of committing a sexual offence against a child; and non-consensual dissemination of intimate images, or render them inaccessible. On that basis, the proposed treaty should also include specific provisions for online platforms and service providers to remove such content on their platforms without delay.

6.4.3 General principles and procedures relating to mutual legal assistance

Third, Article 40 of the draft text contains a lengthy provision on general principles and procedures relating to mutual legal assistance in (i) investigations, prosecutions and judicial proceedings relating to offences laid down in the convention and (ii) with the aim of collecting evidence in electronic form. While victims do not take centre stage in Article 40, they are mentioned in subparagraph 18 alongside witnesses or experts in relation to hearings in court proceedings. Article 40 (18) reads as follows:

> Wherever possible and consistent with fundamental principles of domestic law, when an individual is in the territory of a State Party and has to be heard as a

> witness, victim or expert by the judicial authorities of another State Party, the first State Party may, at the request of the other, permit the hearing to take place by videoconference if it is not possible or desirable for the individual in question to appear in person in the territory of the requesting State Party. States Parties may agree that the hearing shall be conducted by a judicial authority of the requesting State Party and attended by a judicial authority of the requested State Party. If the requested State Party does not have access to the technical means necessary for holding a videoconference, such means may be provided by the requesting State Party, upon mutual agreement.

Noting that cybercrimes can traverse multiple jurisdictions and hearings can occur outside the jurisdiction where the cybercrime occurred, special measures for hearings to take place by videoconference are highly welcomed.

6.4.4 Recovery and return of proceeds of cybercrime

Fourth, Article 52 provides for the recovery and return of proceeds of cybercrime. It provides that the disposal of confiscated shall be in accordance with its domestic law and administrative procedures. Here, primacy is given to the victims of cybercrime and prior legitimate owners in the return of such proceeds. Article 52 (2) provides that:

> When acting on a request made by another State Party [...] States Parties shall, to the extent permitted by domestic law and if so requested, give priority consideration to returning the confiscated proceeds of crime or property to the requesting State Party so that it can give *compensation to the victims* of the crime or return such proceeds of crime or property to their prior legitimate owners. [Emphasis added.]

This means that a victim can be compensated for loss or damage arising from a cybercrime. Indeed, this is a positive step in recognising victims by allowing them to benefit from the prosecution of cybercrimes, particularly in the context of compensation or return of confiscated proceeds of crime or property. The challenge, however, is that compensation in this regard is dependent on domestic law and the confiscation and return of the proceeds of crime. In cases where a state does not have domestic legal redress measures or mechanisms, the guarantees that they may offer to victims are weakened. Commenting on Article 52 of the draft text, Microsoft calls on states to 'enable victims to initiate civil action in courts of other states to protect their property rights violated by cybercriminals' (Microsoft Corporation, 2023).

6.4.5 Preventive measures, technical assistance and capacity-building

Earlier draft texts of the convention did not provide for preventive measures, technical assistance and capacity-building in relation to victims. This might have led the CyberPeace Institute to call for the 'mainstreaming [of] victims' perspectives throughout the chapters on preventive measures and technical assistance.' Perhaps in realising the need to do so, the negotiations towards the current draft text led to the inclusion of

Article 53(3), which provides for preventive measures, including a requirement for states to develop or strengthen support programmes for victims. It also includes provisions on technical assistance and capacity-building in relation to victims.

Specifically, Article 54 requires states parties to afford one another the widest measure of technical assistance and capacity-building, including training, exchange of expertise and, where possible, transfer of technology for the purposes of preventing, detecting, investigating and prosecuting offences under the convention. States parties should also establish, implement or improve specific training programmes for those involved in preventing, detecting, investigating and prosecuting offences under the convention. To the extent permitted by domestic law, these measures may deal with methods 'used in the protection of victims and witnesses who cooperate with judicial authorities' (Article 54(3)(h)).

Capacity-building is particularly important to help criminal justice authorities deal with cybercrimes. Given that evidence relating to cybercrime may be stored in computer systems and networks, those involved in preventing, detecting, investigating and prosecuting cybercrimes should be trained to handle electronic evidence. In addition, this article argues that special training programmes should be made available to relevant stakeholders who provide services to victims of crime. In so doing, technical assistance and capacity-building measures become valuable tools in responding to the specific needs of cybercrime victims.

7. Towards a victim-centred approach?

As demonstrated above, the draft text to the UN convention on cybercrime constitutes the first time since the adoption of the Budapest Convention that states have attempted to respond to the plight of victims of cybercrime by coming up with draft provisions on victims. From the outset of the negotiations, many state and multistakeholder groups agreed that a future cybercrime treaty needed to include provisions for victims of cybercrime. One of the arguments in favour of recognising victims is the need for victims to obtain justice.

During the second session of the Ad Hoc Committee (2022b), held in Vienna between 30 May and 10 June 2022, Mexico proposed the following clause: 'In all actions aimed at the implementation of the present Convention, the best interests of the victims -individuals and institutions and organizations- of the crimes recognized in the present Convention shall be a primary consideration' (Ad Hoc Committee (2022b, p. 6). Although not included in the draft text, Mexico's proposal captured the crux of a victim-centred approach that puts the best interests of victims at the forefront of efforts to implement the convention. Similarly, as the CyberPeace Institute (2023) puts it, 'mainstreaming the victims' perspectives throughout the chapters on preventive measures and technical assistance can support the development of targeted, needs-driven, and context-specific responses to mitigating and preventing cybercrime.'

As noted above, the draft text focuses on measures related to the protection of victims who are witnesses, the assistance and protection of victims, mutual legal assistance provisions, effective remedies, preventive measures and technical assistance measures. While these provisions can guide states on how to address victims of cybercrime, certain shortcomings in the relevant provisions are likely to present challenges. First, the relevant provisions do not require these specific measures to be taken in conformity with international human rights standards. States could therefore define how to implement the provisions, subject to their domestic laws, in ways that disregard their human rights obligations. If the cybercrime instrument is to be effective, human rights standards must be included to guarantee fundamental rights and freedoms.

Second, many of the measures listed in the draft text are to be taken subject to domestic laws. The challenge at the domestic level is threefold: (i) victims of cybercrime are often marginalised by national cybercrime laws; (ii) some states have not adopted adequate cybercrime laws, let alone specific provisions that address victims of cybercrime; and (iii) in the absence of legal safeguards, states could use domestic laws as a tool for arbitrary abuse of power. These challenges may be a stumbling block to the implementation of a victim-centred approach in addressing cybercrime.

Third, little is done to address the unique needs of victims in the cyber context. It must be pointed out that using technology as a means, medium or target of crime creates a wide range of cyber-specific needs. It is therefore imperative that special consideration is given to the nature of online harm and the specific impact of cybercrime on its victims, while also bearing in mind the fact that victimhood depends on a number of aspects such as vulnerability, psychological aspect and age-related differences (Sikra, 2023). These issues should be addressed in future treaty negotiations.

8. Conclusion

While the recognition of victims' needs can be implied through existing cybercrime conventions' retributive approaches, such approaches do not put the needs of victims centre stage. Prosecution is only one aspect of addressing cybercrime and its victims. An appropriate response to the plight of cybercrime victims should be holistic, involving much more than bringing perpetrators to justice. Although, as of the time of writing this article, no consensus had been reached to adopt the draft text to the convention, the proposed treaty has the potential to transform how states view and address victims of cybercrimes. If adopted, it will mark a significant step towards realising specific guarantees for cybercrime victims. However, if it is to make a valuable contribution, the problematic areas on the specific criminal justice needs of cybercrime victims, human rights guarantees and adequate safeguards to protect victims should be resolved by addressing the gaps in the current draft text.

References

Ad Hoc Committee to Elaborate a Comprehensive International Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes (2022a) 'Guiding Questions'. Second Session, 30 May–10 June. www.unodc.org/documents/Cybercrime/AdHocCommittee/Second_session/Documents/Letter_from_AHC_Chair_-_2nd_session_methodology_and_guiding_questions4115.pdf

Ad Hoc Committee to Elaborate a Comprehensive International Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes (2022b) 'Compilation of Draft Provisions Submitted by Member States on Criminalization, General Provisions and Procedural Measures and Law Enforcement'. Note by the Secretariat, Vienna, UN Doc A/AC.291/CRP, 30 May–10 June. www.unodc.org/documents/Cybercrime/AdHocCommittee/Second_session/Documents/CRP11.pdf

Agrafiotis, I., M. Bada, P. Cornish et al. (2016) 'Cyber Harm: Concepts, Taxonomy and Measurement'. Saïd Business School Research Papers 2016–23. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2828646

Agrafiotis, I., J. Nurse, M. Goldsmith et al. (2018) 'A Taxonomy of Cyber-Harms: Defining the Impacts of Cyber-Attacks and Understanding How They Propagate'. *Journal of Cybersecurity* 4(1): yy006.

African Union (2014) 'Convention on Cyber Security and Personal Data Protection'. Adopted on 27 June 2014, entered into force on 8 June 2023.

Brenner, S.W. (2007) 'At Light Speed: Attribution and Response to Cybercrime/Terrorism/Warfare'. *Journal of Criminal Law & Criminology* 97(2): 379–475.

Chile (2021) 'Submissions from Member States Related to the First Session of the Ad Hoc Committee: Chile's Views on the Scope, Objectives, and Structure (Elements) of the New Convention, Regarding the Implementation of UN General Assembly Resolutions 74/247 and 75/282'. www.unodc.org/documents/Cybercrime/AdHocCommittee/First_session/Comments/CHILE_national_views_AHC_05.11.2021.pdf

COE (Council of Europe) (2001a) 'Convention on Cybercrime'. European Treaty Series 185, adopted in Budapest on 23 November 2001 and entered into force on 1 July 2004.

COE (2001b) 'Explanatory Report to the Convention on Cybercrime'. Budapest, 23 November. https://rm.coe.int/16800cce5b

COE (2022) 'Second Additional Protocol to the Convention on Cybercrime on Enhanced Cooperation and Disclosure of Electronic Evidence'.

CyberPeace Institute (2023) 'CyberPeace Institute's Submission to the Fifth Session of the Ad Hoc Committee to Elaborate a Comprehensive International Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes'. https://cyberpeaceinstitute.org/news/submission-to-ad-hoc-committee-on-cybercrime/

EU (2022) 'Contribution from the European Union and Its Member States: Preparation for the First Session of the United Nations Ad Hoc Committee to Elaborate a Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes, Taking Place from 17–28 January 2022 in New York'. www.unodc.org/documents/Cybercrime/AdHocCommittee/First_session/Cwomments/EU_Position_for_AHC_first_session.pdf

Gillespie, A.A. (2016) *Cybercrime: Key Issues and Debates*. New York: Routledge.

Halder, D. (2022) *Cyber Victimology: Decoding Cyber-Crime Victimisation*. New York: Routledge.

ICC (International Criminal Court) (2014) 'Representing Victims before the International Criminal Court: A Manual for Legal Representatives'. Office of Public Counsel for Victims.

Kaspersky (nd) 'What Is WannaCry Ransomware?' www.kaspersky.com/resource-center/threats/ransomware-wannacry (accessed 1 March 2024).

Lewis, J. (2018) *Economic Impact of Cybercrime—No Slowing Down Report*. Washington, DC: CSIS.

Microsoft Corporation (2023) 'Microsoft's Submission to the Sixth Session of the Ad Hoc Committee to Elaborate a Comprehensive International Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes'. www.unodc.org/documents/Cybercrime/AdHocCommittee/6th_Session/Submissions/Multi-stakeholders/Microsoft_Submission_-_AHC_Sixth_Substantive_Session.pdf

Pemberton, A. and I. Vanfraechem (2015) 'Victims' Victimization Experiences and Their Need for Justice'. In Vanfraechem, I., D.B. Fernández and I. Aertsen (eds) *Victims and Restorative Justice*, pp.15–47. New York: Routledge.

Phillips, K., J.C Davidson, R. Farr et al. (2022) 'Conceptualizing Cybercrime: Definitions, Typologies and Taxonomies'. *Forensic Science* (2(2): 379–398.

Rid, T. and P. McBurney (2012) 'Cyber-Weapons'. *The RUSI Journal* 157(1): 6–13.

Scroxton, A. (2020) 'A Trillion Dollars Lost to Cyber Crime Every Year'. *Computer Weekly*, 7 December. www.computerweekly.com/news/252493157/A-trillion-dollars-lost-to-cyber-crime-every-year

Sikra, J., K.V. Renaud and D.R. Thomas (2023) 'UK Cybercrime, Victims and Reporting: A Systematic Review'. *Commonwealth Cybercrime Journal* 1(1): 28–59.

South Africa (2021) 'Submissions from Member States Related to the First Session of the Ad Hoc Committee: South Africa's Views on Scope, Objectives and Structure (Elements) of the Envisaged International Convention on Countering the use of Information and Communications Technologies for Criminal Purposes'. www.unodc.org/documents/Cybercrime/AdHocCommittee/Comments/SOUTH_AFRICA_SUBMISSION_ON_SCOPE_OBJECTIVES_AND_STRUCTURE_17_DECEMBER_202171.pdf

Suryanto, T., H. Hamzah, S. Wahab et al. (eds.) (2020) ICETLAWBE 2020: Proceedings of the International *Conference on Environmental and Technology of Law, Business and Education on Post Covid 19*. European Alliance for Innovation Publishing.

Switzerland (2021) 'Elaboration of a Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes: Switzerland's View on the Objectives, Scope and Structure'. www.unodc.org/documents/Cybercrime/AdHocCommittee/First_session/Comments/Cybercrime_input_Switzerland_102021.pdf

Thomas, D. and B.D. Loader (2000) *Cybercrime: Law Enforcement, Security and Surveillance in the Information Age*. Abingdon: Routledge.

UN Committee on the Rights of the Child (2021) 'General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment'. UN Doc CRC/C/GC/25, 2 March.

UN General Assembly (1989) 'Convention on the Rights of the Child'. *Treaty Series*, 1577, 3.

UN General Assembly (2019) 'Countering the Use of Information and Communications Technologies for Criminal Purposes'. UN Doc A/RES/74/247, 27 December.

UN General Assembly (2023) 'Revised Draft Text to the Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes'. UN Doc A/AC.291/22/Rev.1, 6 November.

UN General Assembly (2024) 'Further Revised Draft Text of the United Nations Convention against Cybercrime'. UN Doc A/AC.291/22/Rev.2, 6 February.

UNICEF (United Nations Children's Fund) (2022) 'Renewed Opportunities: The Convention on Countering the Use of ICTs for Criminal Purposes to Further Strengthen the Protection of Children'. www.unodc.org/documents/Cybercrime/AdHocCommittee/Third_intersessional_consultation/Presentations/Panel_2_Afrooz_Kaviani_Johnson_UNICEF.pdf

Vincent, A.N. (2017) 'Victims of Cybercrime: Definitions and Challenges'. In Martellozzo, E. and E.A. Jane (eds) *Cybercrime and Its Victims*, pp. 27–42. London & New York: Routledge.

Wilkinson, I. and A. Swali (2022) 'Cybercrime Convention Could Help and Harm Victims: The Proposed UN Cybercrime Convention Has Risks and Opportunities for Defining and Protecting Vulnerable Groups'. Chatham House, 19 July. www.chathamhouse.org/2022/07/cybercrime-convention-could-help-and-harm-victims

Wilkinson, I. (2023) 'What is the UN Cybercrime Treaty?' Chatham House, 2 August. www.chathamhouse.org/2023/08/what-un-cybercrime-treaty-and-why-does-it-matter

Yar, M. 2006. *Cyber Crime and Society*. London: SAGE Publications.

About the author

Brenda Mwale is a postdoctoral fellow under the South African Research Chair in International Constitutional Law, University of Pretoria. She holds an LLD from the University of Pretoria, and her LLD thesis is titled 'The Prevention and Repression of Cyber Terrorism in Africa: An Analysis of the Applicable Legal Regimes'. She also holds an LLM in transnational criminal justice from the University of the Western Cape in conjunction with Humboldt-Universität zu Berlin, a postgraduate diploma in law from the Kenya School of Law, and an LLB from Kenyatta University. She is an Advocate of the High Court of Kenya with experience in teaching and legal research. Her research interests lie in public international law, criminal justice and cyber law.