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

The Commonwealth

Cybercrime and the Adoption of Artificial Intelligence Systems for Judicial Decision-Making in Criminal Justice Systems

Dan Jerker B. Svantesson[1]

Abstract

We are now at a stage where cybercrime is more impactful than ever, given the extent to which society operates online. At the same time, there have been significant advances in Artificial Intelligence (AI). Consequently, the temptation to turn to AI to improve the rate of prosecution and adjudication of cybercrime is natural. Clearly, resource limitation is a significant restricting factor; improved efficiencies here may help combat cybercrime.

However, the criminal justice system is one of society's most sensitive functions. Thus, we need to proceed with extreme caution when seeking to rely on AI to improve how we address cybercrime.

Focusing on cybercrime, this article seeks to examine the practicalities of developing guidelines on the adoption of AI systems for judicial decision-making in criminal justice systems.

1. Introduction

Multiple studies have demonstrated that only a small percentage of cybercrime is prosecuted and adjudicated.[2] The reasons for this include resource limitations and cybercriminals' possession of a level of technical expertise comparable with that of those working in cybersecurity. Technologies such as Artificial Intelligence (AI) can help prevent cybercrime by identifying future risks through examining trends in data from earlier cyber-incidents, enabling appropriate authorities to focus on and predict future cyber-attacks.

[1] Faculty of Law, Bond University (Gold Coast, Australia). Email: dasvante@bond.edu.au

[2] See for example Kleijssen, J. and Perri, P. (2017) 'Cybercrime, Evidence and Territoriality: Issues and Options', in M. Kuijer and W. Werner (eds) *Netherlands Yearbook of International* Law 47: 147–173.

Given how common it is today to pursue efficiency via technology, it is only natural to examine the extent to which AI can increase efficiencies in the fight against cybercrime. In doing so, this article approaches AI broadly by examining the adoption of AI systems for judicial decision-making in criminal justice systems.

Importantly, the criminal justice system is a particularly sensitive aspect of society. Consequently, it is critical that any AI-driven efficiencies pursued to address cybercrime are compatible with fundamental values such as the rule of law traditionally emphasised within the Commonwealth.

At the meeting of Commonwealth Law Ministers and Senior Officials in November 2019, it was recognised that not all Commonwealth member countries have access to the same level of technology in their justice systems. The Commonwealth Law Ministers acknowledged the need to remain informed of technological advances and their potential impacts, and stressed the importance of collaborating to increase the scale of knowledge exchange across the Commonwealth in relation to technology.

Law Ministers highlighted the importance of considering the ethical framework surrounding the implementation of new technologies in justice delivery. They supported the development of Commonwealth guidelines to underpin the use of algorithmic decision-making in the legal sphere, based on good practice across the Commonwealth, as well as the formulation of guidance to detail when the public ought to be informed that automated data-driven systems are being used to make decisions of legal consequence. At that occasion, Law Ministers requested the Commonwealth Secretariat to examine the practicalities of developing guidelines on ethical issues linked with the use of technology and to report at the next Commonwealth Law Ministers Meeting (CLMM). As a result, a paper on the adoption of AI systems for judicial decision-making in criminal justice systems was presented at the November 2022 CLMM in Mauritius. At the 2022 CLMM, Law Ministers mandated the Secretariat to adopt a holistic approach to AI in the sector. In particular, the Secretariat should scope emerging practices on the use of AI across the Commonwealth and consider developing principles that align with Commonwealth values and principles.

Within the context of cybercrime, this article seeks to outline current uses of AI systems for judicial decision-making in criminal justice systems, the perceived benefits of such uses, and the risks and challenges involved.

2. Context

The current pandemic has brought about what may be termed a 'COVID-driven trend acceleration' – that is, already existing trends are significantly accelerated because of the COVID-19 pandemic and how society is adjusting to it. One aspect of this is society's increased reliance on the online environment. Put simply, information technology is playing an important role in society. With this come increased opportunities for cybercriminals. This situation also means that cybercrime has a larger societal impact. Thus, it may be argued that it has never been more important to effectively address cybercrime.

The COVID-driven trend acceleration can also be seen in developments in, and the adoption of, AI as a part of this broader and intensified technological uptake. However, AI already has quite a long history.

2.1 A background to AI

AI has been discussed with varying degrees of intensity since the early 1950s.[3] However, significant advances in computing power, data analysis techniques, natural language processing and the availability of large datasets have recently transformed the AI landscape and brought about rapid progress. Despite the attention directed at AI, there are no universally accepted definitions, and AI is not a homogeneous object. However, common themes in discussions of AI focus on the ability to 'mimic human thought'[4] and thereby the capacity to carry out tasks previously falling within the exclusive domain of human capability. The Council of Europe's Ad Hoc Committee on Artificial Intelligence, for example, has concluded that:

> *the term 'AI' is used as a 'blanket term' for various computer applications based on different techniques, which exhibit capabilities commonly and currently associated with human intelligence. These techniques can consist of formal models (or symbolic systems) as well as data-driven models (learning-based systems) typically relying on statistical approaches, including for instance supervised learning, unsupervised learning and reinforcement learning. AI systems act in the physical or digital dimension by recording their environment through data acquisition, analysing certain structured or unstructured data, reasoning on the knowledge or processing information derived from the data, and on that basis decide on the best course of action to reach a certain goal. They can be designed to adapt their behaviour over time based on new data and enhance their performance towards a certain goal.*[5]

[3] See in particular Turing, A. (1950) 'Computing Machinery and Intelligence', in R. Epstein, G. Roberts and G. Beber (eds) (2009) *Parsing the Turing Test*. Dordrecht: Springer; McCarthy, J., Minsky, M., Rochester, N. and Shannon, C. (1955) 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence'. Reprinted in AI Magazine 27(4) (2006).

[4] Simon McDougall, S. (2019) 'Developing the ICO AI Auditing Framework: An Update'. UK Information Commissioner's Office, 4 July. https://ico.org.uk/about-the-ico/news-and-events/ai-blog-developing-the-ico-ai-auditing-framework-an-update/

[5] Council of Europe Ad Hoc Committee on Artificial Intelligence (2020) 'Feasibility Study CAHAI(2020)23'. Strasbourg, 17 December. https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

Change – in the form of AI adoption – is already ongoing and, as is often the case, we are in a situation where regulation is trying to 'catch up' with technological advances. However, especially in the context of criminal justice systems, doubt has been cast on how much in this evolution is driven by societal needs as opposed to by marketing and the industry.[6]

Furthermore, given the seemingly undisputed fact that AI will continue to evolve, the task before us is not merely to ensure that regulation catches up with AI technology. It also involves setting rules and standards now that can guide and regulate the adoption of AI systems for judicial decision-making in criminal justice systems on an ongoing basis. Put simply, this is not a task that can be addressed to completion at this stage; it will require an ongoing commitment, monitoring and review.

2.2 AI and the judicial system

The judicial system is a part of society and, as such, it should evolve with the rest of society, including when it comes to technological developments. However, technological development, such as the adoption of AI systems, is not in itself a goal for the criminal justice system. Rather, it is the potential that AI systems hold to facilitate recognised goals such as efficiency increases, cost minimisation and improved access to justice that makes it a topic worthy of consideration.

Any discussions of technological reform to judicial decision-making in criminal justice systems must start with the realisation that the justice system is a sensitive core function of society and thus it is something that requires particular care.[7] As a result, technological progress is clearly less important than is the integrity of the system, and progress being slower in relation to the adoption of AI in the judicial system, compared with for less sensitive societal functions, is both natural and desirable. This means that solutions that may be viewed as appropriate in less sensitive areas may not be suitable for how the criminal justice system addresses cybercrime or, indeed, for the criminal justice system as a whole.

[6] For example, the European Commission for the Efficiency of Justice prompts us to 'differentiate between this commercial discourse and the reality of the use and deployment of these technologies' (European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, 31st Plenary Meeting, 3–4 December 2018. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c). The 'hype' around AI is well illustrated in the following statement made in 1960: 'Within the very near future – much less than twenty-five years – we shall have the technical capability of substituting machines for any and all human functions in organisations. ... Duplicating the problem-solving and information-handling capabilities of the brain is not far off; it would be surprising if it were not accomplished within the next decade" (see further the discussion of this 'over-enthusiasm' (Clarke, R., 2019, 'Guidelines for the Responsible Business Use of AI'. Foundational Working Paper Revised Version of 20 February 2019. Xamax Consulting Pty Ltd. www.rogerclarke.com/EC/GAIF.html).

[7] The need to impose strict requirements on judicial decision-making in criminal justice systems is widely accepted. See for example the emphasis on independence, impartiality, integrity, propriety, equality, competence and diligence in the United Nations Bangalore Principles of Judicial Conduct (see UNODC, 2007, *Commentary on the Bangalore Principles of Judicial Conduct*. Vienna: UNODC).

2.3 An enabling framework safeguarding key values

The main message of this article is that, to increase efficiencies in addressing cybercrime, there is a need for a clear framework – preferably at the Commonwealth level – that enables developments for the adoption of AI systems for judicial decision-making in criminal justice systems to take place in a safe, accountable and rights-respecting manner safeguarding the values of the Commonwealth, including the rule of law as an essential protection for the people of the Commonwealth and as an assurance of limited and accountable government. In particular, despite the importance of addressing cybercrime, AI systems should not be adopted in a manner that undermines an independent, impartial, honest and competent judiciary or the independent, effective and competent legal system as an integral component in upholding the rule of law, engendering public confidence and dispensing justice.

The design of such a framework is a task for all Commonwealth member countries equally, regardless of their current state of technological[8] and policy[9] development. Too often, a small number of the most technologically advanced countries set technological and regulatory developments for all countries.[10] By adopting the inclusive and multistakeholder approach to which the Commonwealth is already committed,[11] and by doing so at this early stage of the development of the adoption of AI systems for judicial decision-making in criminal justice systems, a more equitable and better-informed direction can be set.

2.4 Necessity and urgency

What is at stake in the context of the adoption of AI systems for judicial decision-making in criminal justice systems is nothing less than the following six matters of fundamental importance for all Commonwealth member countries:

[8] Unequal access to the Internet remains a major concern. For example, 'Only 40% of Africans have access to the Internet today, compared to 87% in Europe and 95% in North America' (Candelon, F., El Bedraoui, H. and Maher, H. (2021) 'Developing an Artificial Intelligence for Africa Strategy'. OECD Blog, 9 February. https://oecd-development-matters.org/2021/02/09/developing-an-artificial-intelligence-for-africa-strategy/). See more generally the AI Readiness Index 2020 of Oxford Insights and the International Research Development Centre (www.oxfordinsights.com/government-ai-readiness-index-2020).

[9] While some Commonwealth member countries (especially Canada and Singapore) have been world-leading adopters of AI strategies and policies, others are yet to develop such instruments. And, indeed, many still lack strategies and policies for the digital economy more broadly. For example, it has been noted that, 'While most Caribbean economies have developed broad ICT [information and communication technology] policies and strategies throughout the years, very few have developed targeted policies specifically addressing aspects of the digital economy' (Brathwaite, C., 2020. 'Artificial Intelligence & The Caribbean: A Discussion Paper on (Potential) Applications & Ethical Considerations', in C. Aguerre (ed.) *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*. Buenos Aires: CETyS Universidad de San Andrés).

[10] Svantesson, D. (2019) *Internet & Jurisdiction Global Status Report 2019*. Paris: Internet & Jurisdiction Policy Network.

[11] See for example the Commonwealth Cyber Declaration 2018 (https://thecommonwealth.org/commonwealth-cyber-declaration-2018).

1. justice in individual cases
2. the protection of fundamental human rights
3. upholding human dignity
4. adherence to the rule of law
5. what qualifies as a source of law and
6. trust in the legal system.

Given the fundamental significance of these matters, regulation by law is a necessity.[12] The types of ethics-focused guidelines and standards that have preceded discussions of regulation by law remain useful but are no substitutes for regulation by law for such sensitive matters. Thus, the legal regulation of AI systems for judicial decision-making in criminal justice systems is an issue on which Law Ministers may wish to start reflecting now, even where they are not currently in the process of actually adopting such systems in the pursuit of better addressing cybercrime.

3. Uses, including selected illustrative examples

When it comes to the uses to which AI may be put to address cybercrime, it must be understood that much of the discussion is focused on the adoption of AI systems for judicial decision-making in criminal justice systems in a technology-neutral manner as far as the types of crime are concerned. In other words, the efficiency impact of AI is generally pursued across all criminal activities, not specifically for cybercrime.

As this article addresses the adoption of AI systems for judicial decision-making in criminal justice systems, it is prudent to pay attention to the meaning of 'AI systems for judicial decision-making'. In this context, we may note the Australian Human Rights Commission's useful definition of 'AI-informed decision making' as a 'decision or decision-making process that is materially assisted by the use of an AI technology or technique and that has a legal, or similarly significant, effect for an individual'.[13]

As also noted by the Australian Human Rights Commission, 'In some cases, the use of AI will be central to the decision-making process, and the ultimate decision. For others, AI will have only a trivial impact on the decision-making process.'[14] This is an important observation that brings attention to the need for an appropriate analytical granularity and to the fact that each use must be evaluated individually.

[12] For an example of an attempt at comprehensively regulating AI, see Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final.

[13] Australian Human Rights Commission (2021) *Human Rights & Technology Final Report* (p. 38).

[14] Ibid. (p. 39).

The types of uses to which AI has been, and is anticipated to be, put into service within the context of judicial decision-making in criminal justice systems are diverse. In part, this is a consequence of AI not being a homogeneous object. However, it may also be explained by the fact that this is very much a developing field. Indeed, an examination of the extent to which Commonwealth member countries have already adopted AI systems for judicial decision-making in criminal justice systems provides few examples. This is also the case more broadly.[15] For example, the European Commission for the Efficiency of Justice (CEPEJ) notes that, 'For the time being judges in the Council of Europe member states do not seem to be making any practical and daily use of predictive software.'[16] In the literature, much of the discussions are focused on examples from the United States.

Nevertheless, it is possible to discern six different types of uses that warrant mentioning. These uses take place in different settings and at different points in time in the judicial process, but some may apply in parallel.

3.1 AI systems adopted to support the administration of the judicial process

AI systems can be adopted to support the administration of the judicial process. Such systems can, for example, make case management more efficient, and perform evaluations supporting budget and resource predictions. This may be particularly relevant in the context of cybercrime, given the very small percentage of cases prosecuted and adjudicated.

AI systems can also be adopted to support the administration of the judicial process in the form of tools for providing, and making more accessible, legal information, for example via 'chatbots' capable of providing tailored information. One provider describes the functionality of its Artificially Intelligent Legal Information Research Assistant in the following: 'You chat with her just as you a human lawyer. You can ask her questions, and she may ask you questions back to help guide you to helpful legal information. She can help you create documents, and if need be speak with a human lawyer to review those documents and information provided to you.'[17] As victims of cybercrime already are online, such resources may be particulary relevant for such victims.

[15] 'The number of cases in which AI programs have actually been used as a support tool for adjudication does not amount to more than can be counted on the fingers of one hand' (Caianiello, M., 2021, 'Dangerous Liaisons. Potentialities and Risks Deriving from the Interaction between Artificial Intelligence and Preventive Justice'. *European Journal of Crime, Criminal Law and Criminal Justice* 29: 1-23, p. 3).

[16] CEPEJ (2018) 'European Ethical Charter' (p. 14).

[17] www.ailira.com/

Steps like these may strengthen access to justice and cut costs. Adopted in these roles, AI systems introduce limited risks, although, as discussed below (Section 5), risks still exist and must be managed. After all, decisions made about the judicial process, such as the case allocation to specific judges, for example, may be biased, discriminatory and violate fundamental rights relating to procedural fairness.

3.2 AI systems providing decision-making support

AI systems may provide decision-making support. This can take different forms, some of which we are already familiar with, such as information systems (e.g. advanced case law search engines).[18] Others, such as analytical functions and, for example, the use of AI systems to propose possible interpretations of ambiguous terms or statutory provisions, take us into a 'grey zone' between AI systems providing decision-making support and such systems becoming co-adjudication, as discussed directly below. The same is true in relation to AI systems used for 'judge profiling', for example in the form of offering judges detailed quantitative and qualitative assessment of their own activities to help them self-identify patterns indicative of biases. The European Commission for the Efficiency of Justice suggests that, used purely for the informative aim of assisting in decision-making and for the judges' own exclusive use, such profiling could be encouraged.[19]

3.3 AI systems utilised for co-adjudication

AI systems, adopted in the context of judicial decision-making in criminal justice systems, can be utilised for co-adjudication – that is, judicial decisions are made by a human judge together with the AI system. This usage can, for example, include the AI system drafting decisions that are approved and edited by a human judge, or the AI system proposing alternative options based on pre-set criteria and highlighting the most 'suitable' based on those criteria while also ranking the degree of criteria fulfilment of other outcomes.

3.4 AI systems as 'robot judges'

The most advanced form of AI system use in the context of judicial decision-making in criminal justice systems is the 'robot judge'. A robot judge would be an AI system that directly and autonomously adjudicates matters. This is not a current usage but it needs to be flagged here as such a development is part of the discussions of AI systems for judicial decision-making in criminal justice systems. Any application of robot judges in the context of cybercrime must be guided by the sensitivity of the criminal justice system.

[18] CEPEJ (2018) notes, 'The use of machine learning to constitute search engines for case-law enhancement is an opportunity to be taken up for all legal professionals' ('European Ethical Charter', p. 63).

[19] CEPEJ (2018) 'European Ethical Charter' (p. 66).

3.5 AI systems in alternative dispute resolution structures

While more prominent outside the criminal justice system, some forms of alternative dispute resolution (ADR) – like restorative justice – may be used in the context of decision-making in criminal justice systems, potentially also in relation to cybercrime. As AI systems may play a direct or supporting role in various ADR structures, this usage must be noted.

3.6 AI systems facilitating predictive policing, preventative justice and pre-trial risk assessment

The use of AI systems to facilitate so-called 'predictive policing', 'preventative justice'[20] and 'pre-trial risk assessment' may play a role in more than one of the five categories outlined above, and has attracted much attention in the literature.[21] This is only natural given that this is a setting in which there are examples of both adopted systems and systems being tested. At the same time, it may be noted that predictive policing does not need to depend on AI systems, but rather can be carried out entirely by human intelligence. At any rate, the Australian Human Rights Commission notes:

> *Data-driven risk assessment tools are used increasingly to predict the likelihood of future criminal behaviour. These tools are starting to be rolled out in a number of countries to assist decision making in the criminal justice system, including decisions regarding sentencing, bail and post-sentence restrictions on people assessed as being likely to commit further crime.*[22]

The best-known instances of predictive policing relate to physical crime rather than cybercrime. Perhaps the most familiar of these data-driven risk assessment tools, outside the United States, is the Harm Assessment Risk Tool (HART) developed by the

[20] The term 'predictive justice' has been criticised as being 'dangerously misleading' since 'such systems make predictions, but not judicial decisions. Judicial decisions require, as a minimum standard, justifications based on an assessment of the relevant facts and applicable regulations. AI systems make statistical correlations and their forecasts are just the result of those correlations. Hence, it would only be proper to speak of actual predictive justice if the systems were to provide justifications in terms of facts and laws' (Contini, F., nd, 'Artificial Intelligence: A New Trojan Horse for Undue Influence on Judiciaries?' https://www.unodc.org/dohadeclaration/en/news/2019/06/artificial-intelligence_-a-new-trojan-horse-for-undue-influence-on-judiciaries.html).

[21] See for example Ashworth, A. and Zender, L. (2014) *Preventive Justice*. Oxford: Oxford University Press; Lynskey, O. (2019) 'Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing'. *International Journal of Law in Context* 15(2): 162–176. Most pretrial risk assessment tools try to estimate 'recidivism risk' (i.e., how likely a person is to commit a crime or be arrested) and 'flight risk' (i.e., how likely a person is to not show up at trial) (EPIC, 2020, *Liberty at Risk: Pre-trial Risk Assessment Tools in the U.S.* Washington, DC: EPIC).

[22] Australian Human Rights Commission (2021) *Human Rights & Technology Final Report* (p. 43).

University of Cambridge in collaboration with Durham Constabulary.[23] HART is based on a machine-learning algorithm to aid decision-making by custody officers when assessing the risk of future offending.[24] In more detail:

> *the AI-based technology uses 104,000 histories of people previously arrested and processed in Durham custody suites over the course of five years, with a two-year follow-up for each custody decision. Using a method called 'random forests', the model looks at vast numbers of combinations of 'predictor values', the majority of which focus on the suspect's offending history, as well as age, gender and geographical area. …*
>
> *The aim of HART is to categorise whether in the next two years an offender is high risk (highly likely to commit a new serious offence such as murder, aggravated violence, sexual crimes or robbery); moderate risk (likely to commit a non-serious offence); or low risk (unlikely to commit any offence).*[25]

Other noteworthy data-driven risk assessment tools include Connect, widely used by UK police[26] and recently commissioned for the Jamaican police force,[27] the New South Wales Police Force's Suspect Targeting Management Plan (STMP)[28] and Interpol's International Child Sexual Exploitation Database.[29] The latter is an intelligence and investigative tool with an image and video database that allows specialised investigators to share data on cases of child sexual abuse:

> *Using image and video comparison software, investigators are instantly able to make connections between victims, abusers and places. The database avoids duplication of effort and saves precious time by letting investigators know whether a series of images has already been discovered or identified in another country, or whether it has similar features to other images.*

It also allows specialized investigators from more than 64 countries to exchange information and share data with their colleagues across the world.

[23] University of Cambridge (2018) 'Helping Police Make Custody Decisions Using Artificial Intelligence'. 26 February. www.cam.ac.uk/research/features/helping-police-make-custody-decisions-using-artificial-intelligence

[24] Oswald, M., Grace, J., Urwin, S. and Barnes, G. (2018) 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality'. *Information & Communications Technology Law* 27(2): 223–250 .

[25] University of Cambridge (2018) 'Helping Police Make Custody Decisions'.

[26] For further details, see the provider's website: www.necsws.com/solutions/police-software/police-record-management-system/

[27] Booth, F. (2021) 'Jamaican Police Force Adopts Technology Platform in Drive to Combat Crime'. NEC Insights, 6 July. www.necsws.com/news/jamaican-police-adopts-technology-platform/

[28] See further Yeong, S. (2020) 'An Evaluation of the Suspect Target Management Plan'. Crime and Justice Bulletin No. 233 Revised. Sydney: NSW Bureau of Crime Statistics and Research; Sentas, V. and Pandolfini, C. (2017) 'Policing Young People in NSW: A Study of the Suspect Targeting Management Plan'. Report of the Youth Justice Coalition NSW. Sydney: Youth Justice Coalition NSW.

[29] See further www.interpol.int/en/Crimes/Crimes-against-children/International-Child-Sexual-Exploitation-database and the discussion in Završnik, A. (2020) 'Criminal Justice, Artificial Intelligence Systems, and Human Rights'. *ERA Forum* 20: 567–583 https://doi.org/10.1007/s12027-020-00602-0

> *By analysing the digital, visual and audio content of photographs and videos, victim identification experts can retrieve clues, identify any overlap in cases and combine their efforts to locate victims of child sexual abuse.*[30]

Importantly, predictive policing may have particular future application in relation to cybercrime, for example based on spatial studies of cybercrime perpetrators and cybercrime victims. This is certainly a topic that requires further academic attention.

4. Perceived benefits

An examination of policy documents and academic literature shows considerable consistency in the perceived benefits associated with the adoption of AI systems for judicial decision-making in criminal justice systems. Some such benefits are clearly related, and partly overlapping. For example, increased efficiency may lead to cost reductions and may, at the same time, facilitate greater access to justice that caters for greater equality. However, whether talking of increased efficiency or related cost reduction, it may be said that the common denominator is the potential to enhance access to justice, clearly a key issue in the context of cybercrime.

4.1 Three forms of enhanced access to justice

With its strong and clearly articulated commitment to, and focus on, facilitating access to justice,[31] it is important that the Commonwealth explore options for the adoption of AI systems for judicial decision-making in criminal justice systems. Considering the uses described in Section 3, AI systems have the potential to enhance access to justice in at least three different ways: (i) improved access, and improved quality of that access, to legal information; (ii) greater procedural efficiency leading to, for example, shorter waiting times; and (iii) increased quality of judicial decision-making. These forms of enhanced access to justice are as relevant for cybercrime as they are for offline crimes.

The first two of these ways in which AI systems may enhance access to justice are rather self-explanatory. However, a few observations must be made about the third. One aspect of the potential for increased quality of judicial decision-making is found in that AI systems may be used to identify, minimise and even eliminate human biases.

[30] www.interpol.int/en/Crimes/Crimes-against-children/International-Child-Sexual-Exploitation-database

[31] See for example Latu-Sanft, J. (2019) 'Commonwealth Law Ministers Resolve to Take Action on Access to Justice'. News, 7 November. https://thecommonwealth.org/media/news/commonwealth-law-ministers-resolve-take-action-access-justice.

4.2 Identifying, minimising and eliminating human biases

Decisions made by humans may be made based on conscious or subconscious biases, and those mental shortcuts that form a natural part of human reasoning.[32] AI systems may be used to address such issues in at least two different ways. First, where AI makes the decisions, such direct human biases are eliminated (although, as discussed below, AI systems may also introduce, or re-introduce, biases).

Second, AI systems may be used to analyse human decisions with the aim of identifying biases. This may be done at the time of the decision-making process, thereby giving the decision-making person the opportunity to become aware of biases and alter the decision before it is finalised. However, AI-based identification of human biases may also be utilised after decisions have been made in a monitoring/review context.

4.3 Other potential benefits

Other potential benefits that the adoption of AI systems for judicial decision-making in criminal justice systems may provide include certainty, equality, consistency, predictability and transparency, resulting in higher quality in judicial decision-making.[33] For example, AI systems may help eliminate excessive variability in court decisions,[34] thus supporting the principle of equality before the law.

5. Identified risks and challenges

As with the perceived benefits discussed above, an examination of policy documents and academic literature shows considerable consistency in the identified risks and challenges associated with the adoption of AI systems for judicial decision-making in criminal justice systems.

5.1 AI systems and legal authority

A fundamental consideration for any adoption of AI systems for judicial decision-making in criminal justice systems stems from the limits on legal authority – that is, can AI systems have legal authority in the judicial decision-making in criminal justice systems?[35] Without such legal authority, their use must always be limited to a support function. The way in which this restriction has been managed is through 'delegated authority',[36] but

[32] Završnik (2020) 'Criminal Justice'.

[33] See for example Caianiello (2021) 'Dangerous Liaisons'.

[34] Re, R. and Solow-Niederman, A. (2019) 'Developing Artificially Intelligent Justice'. *Stanford Technology Law Review* 242–289.

[35] Sourdin, T. (2018) 'Judge v Robot? Artificial Intelligence and Judicial Decision-Making'. *UNSW Law Journal Volume* 41(4): 1114–1133.

[36] For a possible example of this, see Australia's Therapeutic Goods Act 1989 (Cth) s 7C(2) discussed in Sourdin (2018) 'Judge v Robot?'.

different legal systems may impose different limitations on the delegation of decision-making power to machines. Ensuring authority may be particularly complicated in the context of cross-border crime, as often is the case in cybercrime.

5.2 Distinction between outcome and process – human dignity and the risk of alienation

Another issue of overriding importance is the distinction between outcome and process. Even where AI systems are capable of producing results akin to those of a human judge, those systems make no attempt to formalise legal reasoning. Rather, the developers create models aimed at anticipating the likely decisions of a judge in similar situations.[37] Given the important role played by the process itself in judicial decision-making in criminal justice systems – including for cybercrime – it may be argued that AI systems are incapable of meeting certain fundamental rights standards, such as those relating to procedural fairness, and that being judged by a machine undermines human dignity. This points to a need for a deeper discussion of a potential right to be judged by a human. Relatedly, commentators have pointed to the risk of AI systems causing a sense of alienation in relation to the legal system.[38]

5.3 AI systems creating a new, unintentional, source of law?

Another challenge that calls for deeper discussions is the potential for AI systems to, in a sense, create a new source of law. Some commentators have pointed to the risk of 'datafication' – that is, 'by focusing attention on seemingly objective data and adapting legal systems to incorporate this information, "datafication," or emphasis on available data and its uses, might undesirably influence the legal system's operation'.[39] Similarly, as noted by CEPEJ, 'Thought should be given to the transformation of the very logic of the production of case-law. What is the value of the "standard" resulting from the number of decisions given on a specific matter? Does this "standard" add to the law? If so, is this a new source of law?'[40] Put differently, will AI systems turn quantitative caselaw statistics into a source of law? And, if so, how will that source relate to the more qualitative case law usage typical of human judges? Further, the AI systems' quantitative treatment of case law may impact court hierarchy and the court system, as such:

> *would it not be the case that if norms were established according to the majority trend, judicial decisions would be rendered uniform, and no longer be ordered according to the hierarchy of the courts from which they emanate, disregarding the significance of the decisions of supreme courts, which are the guarantors of the uniform interpretation of law in many … States?*[41]

[37] CEPEJ (2018) 'European Ethical Charter'.

[38] Re and Solow-Niederman (2019) 'Developing Artificially Intelligent Justice'.

[39] Ibid. (p. 267).

[40] CEPEJ (2018) 'European Ethical Charter' (p. 23).

[41] Ibid. (p. 24).

Perhaps there is such a thing as 'excessive standardisation of judicial decisions'[42] and perhaps such excessive standardisation is more likely to cause inequality than it is to cater for equality. All this is particulary sensitive in the context of the criminal justice system, including in its application to cybercrime.

5.4 AI systems and the need for discretion

Some commentators[43] have pointed to the prevalence of discretion in adjudication stemming from, for example, the 'under-determinate' nature of law. They highlight the important role of discretion and question whether AI systems can fulfil the aspects of adjudication that involve discretion.[44] In addition, we may question whether AI systems *should* handle the aspects of adjudication that involve discretion or whether such discretion must always be exercised by a human. This is a most significant question. If the answer is that such discretion must always be exercised by a human, the result seems to be that, for most purposes falling within the context of judicial decision-making in criminal justice systems, AI systems may not be used as an autonomous 'robot judge' but may be used only for, for example, co-adjudication and decision-making support.

5.5 The risk of disillusionment

It has also been observed that the adoption of AI systems for judicial decision-making may cause disillusionment in relation to human judges and the legal system more broadly – that is, where an AI system highlights flaws and biases in the decisions of human judges, a resulting 'disillusionment would erode confidence in the legal system's legitimacy. Insofar as increasing use of AI adjudication prompts people to look more skeptically [sic] at human judging, the legitimacy of existing legal activities could be cast into doubt.'[45]

5.6 A lack of transparency

One of the most frequently noted risks with AI systems for judicial decision-making, including in the context of criminal justice systems applied in the cybercrime context, is that stemming from AI creating a lack of transparency. This is not a setting in which a 'black box' effect[46] – put simply, opacity in the step between input and output in a decision-making process – can be accepted. Decisions made in the criminal justice context must be explainable to those directly affected. This is essential, for example

[42] Ibid.

[43] See for example Završnik (2020) 'Criminal Justice' (pp. 580–581).

[44] 'Removing human discretion thus is a double-edged sword: it can reduce human bias, but it can also exacerbate past injustices or produce new ones' (Završnik, 2020, 'Criminal Justice', p. 581).

[45] Richard M. Re & and Alicia Solow-Niederman (2019,) 'Developing Artificially Intelligent Justice', 22 STAN. TECH. L. REV. 242 (2019), at (p. 273).

[46] See for example NITI Aayog (2018) 'National Strategy for Artificial Intelligence'. Discussion Paper. https://indiaai.gov.in/documents/pdf/NationalStrategy-for-AI-Discussion-Paper.pdf; Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

to ensure that informed decisions can be made about whether a person accused of committing cybercrime should appeal the decision, and if so on what grounds, but also more generally for the dignity of the affected party. Such decision must also be explainable to society at large so that the overall fairness and equality of the system may be monitored.

5.7 AI systems and biases

Another frequently highlighted concern with AI systems for judicial decision-making, not least in criminal justice systems, relates to biases. As noted in Section 4, AI systems may be used to identify and even eliminate human biases. But AI systems may also introduce biases in at least three forms. First, any human biases of the creators of the AI system may be transmitted into the AI system as such ('creator biases').[47] Second, the selected training data for an AI system may significantly affect the system's operation, and any intentional or unintentional biases in the data may contaminate how the system operates ('data-driven biases'). In this context, specific mention must be made of the risk that biases that society has moved past are reintroduced where old data (i.e., data that predate the change away from that bias) are used. Third, systems that evolve, such as forms of machine learning, may develop biases over time ('systems-driven biases').

The obvious concern about these types of biases is that they may result in discrimination between individuals and groups of individuals. This is a key concern in the context of the adoption of AI systems for judicial decision-making in criminal justice systems, including in the context of cybercrime. However, such biases are not the only way in which discrimination may arise. As the Australian Human Rights Commission notes, 'Poor technology design can exclude people with disability from work, services and the economy.'[48] In other words, in order to avoid discrimination, technology must be made accessible to all regardless of factors such as level of education, gender, economic position, demographics and disabilities.

5.8 Technology dependence and the 'digital divide'

Increases in technology dependence – including the adoption of AI systems for judicial decision-making in criminal justice systems – may augment the so-called 'digital divide' – that is, the gap between those with reliable access and those who lack reliable access to the technology. The digital divide may stem from a range of sources, such as gender, location, level of education and differences in the available Internet architecture, and may exists on several levels, including between individuals, groups and countries.[49]

[47] 'The neutrality of algorithms is a myth, as their creators consciously or unintentionally transfer their own value system into them' (CEPEJ, 2018, 'European Ethical Charter', p. 57).

[48] Australian Human Rights Commission (2021) *Human Rights & Technology Final Report* (p. 7).

[49] See for example Sibal, P. and Neupane, B. (2021) *Artificial Intelligence Needs Assessment Survey in Africa*. Paris: UNESCO.

5.9 Function creep' and AI systems as an unintended de facto judge

A further concern that has been expressed in relation to the adoption of AI systems for judicial decision-making in criminal justice systems is the risk of so-called 'function creep'.[50] Put simply, function creep occurs where the use or application of a system or technology expands beyond its original purposes.[51] This is a commonly articulated concern in the intersection between law and technology. In the context of judicial decision-making in criminal justice systems, function creep may occur on several levels. For example, systems implemented for decision-making support may end up being used as co-adjudicators, and systems implemented for co-adjudication may end up being used as the main adjudicator.

A partially overlapping risk is that of overreliance making AI systems the *de facto* judge even where the AI system is only meant to be a co-adjudicator or to provide decision-making support. This risk stems from the fact that human judges may experience fear of departing from AI-guided draft decisions or recommendations.[52] Put simply, where a human judge follows a draft decision or recommendation made by an AI system, errors in that decision may be 'blamed' on the AI system. In contrast, where a human judge departs from a draft, proposal or recommendation made by an AI system, errors in the judge's decision may be 'blamed' on the judge in question, and the sense of blame may be intensified by the very fact that the judge did not follow the proposal made by the AI system.

5.10 AI systems as 'self-fulfilling prophecies'

Fears have been expressed about the potential for AI systems to become what may be labelled 'self-fulfilling prophecies'. This may be the case where an error, or misjudgement, made somewhere in the process, becomes incorporated in a decision that then forms the basis for future decision-making. In this context, the difficulty of identifying and eliminating false correlations is noteworthy. For example, a system that identifies a pattern of cyber-fraud conducted by persons with a name starting with a particular letter may assume a correlation between the first letter of names and the propensity to commit cyber-fraud. Put in the context of predictive policing, this may raise issues of confirmation bias. To continue the example above, the AI system's false correlations between the first letter of names and the propensity to commit cyber-fraud may cause investigative resources to be directed towards individuals whose name starts with the relevant letter. Such increased scrutiny is likely to result in a greater number of cyber-fraudsters being found among the group subjected to the increased scrutiny and the AI system is then proven 'right' even though the same result may have been reached regardless of which letter was in focus.

[50] Završnik (2020) 'Criminal Justice'.

[51] Koops, B.J. (2021) 'The Concept of Function Creep'. *Law, Innovation and Technology* 13(1): 29–56.

[52] See further CEPEJ (2018) 'European Ethical Charter'.

5.11 AI systems and the many roles of the judge

Commentators have also pointed to the multiple roles fulfilled by human judges.[53] For example, the role of a judge goes beyond dispute resolution to include activities such as education activities and social commentary. It seems clear that AI systems are not positioned to replace the human judge for such activities.[54] However, this may be a relatively minor issue since human judges can carry on with those activities even if complemented by AI systems. More importantly, concerns may be expressed about AI systems' argued inability to assess the social impact of their decisions and inability to ensure protection of societal values.

5.12 AI systems and the many roles of the law

It must be noted that the law fulfils multiple roles, and AI systems' ability to provide outputs that go beyond the case at hand has been called into question. Without going to deep into the quagmire of legal philosophy, it may be suggested that the law fulfils three different roles: it is a tool to (i) decide legal disputes, (ii) provide a framework to control, guide and plan life out of court and (ii) express and communicate the values of those who created the law.[55] For the second and third of these functions, the process and reasoning that led to the outcome (including what a judge states in *obiter*) are almost as important as the outcome itself. Doubt exists as to how capable AI systems are to fulfil these functions of the law.

5.13 A negative impact on fundamental rights and fundamental values

Much of the discussion of risks and challenges above one way or another relates back to concerns about the adoption of AI systems for judicial decision-making in criminal justice systems having a negative impact on fundamental rights, and fundamental values such as the rule of law. As already alluded to, several such rights are of relevance in the context of judicial decision-making in criminal justice systems – including in relation to cybercrime – and may be at risk as a result of the adoption of AI systems. Most obviously, principles of equality before the law, of the presumption of innocence, of the right to a fair and public hearing by a competent, independent and impartial tribunal established by law and of the right to be tried without undue delay are at stake. Article 14 of the International Covenant on Civil and Political Rights (ICCPR) is illustrative. It includes rights such as:

- All persons shall be equal before the courts and tribunals (14(1)).

[53] Sourdin (2018) 'Judge v Robot?'.

[54] Ibid.

[55] The first two of these roles may be derived from Hart, H., Raz, J. and Bulloch, P. (2012) *The Concept of Law*. 3rd ed. Oxford: Oxford University Press: 'The principal functions of the law as a means of social control are not to be seen in private litigation or prosecutions, which represent vital but still ancillary provisions for the failures of the system. It is to be seen in the diverse ways in which the law is used to control, to guide, and to plan life out of court' The third role is articulated in Svantesson, D. (2015) 'A Jurisprudential Justification for Extraterritoriality in (Private) International Law'. *Santa Clara Journal of International Law* 13(2): 517–552.

- Everyone shall be entitled to a fair and public hearing by a competent, independent and impartial tribunal established by law (14(1)).
- Any judgement rendered in a criminal case or in a suit at law shall be made public except where the interest of juvenile persons otherwise requires or the proceedings concern matrimonial disputes or the guardianship of children (14(1)).
- There exists the right to be presumed innocent until proved guilty according to law (14(2)).
- There exists the minimum guarantee to be informed promptly and in detail in a language which he understands of the nature and cause of the charge against him (14(3)(a)).
- There exists the minimum guarantee to have adequate time and facilities for the preparation of his defence and to communicate with counsel of his own choosing (14(3)(b)).
- There exists the minimum guarantee to be tried without undue delay (14(3)(c)).
- There exists the minimum guarantee to examine, or have examined, the witnesses against him and to obtain the attendance and examination of witnesses on his behalf under the same conditions as witnesses against him (14(3)(e)).
- In the case of juvenile persons, the procedure shall be such as will take account of their age and the desirability of promoting their rehabilitation (14(4)).
- There exists the right to his conviction and sentence being reviewed by a higher tribunal according to law (14(5)).

There can be little doubt that the adoption of AI systems for judicial decision-making in criminal justice systems may be in conflict with obligations such as those noted above. Consequently, it is necessary to appreciate that these requirements are not obstacles for AI adoption; rather, their fulfilment is a necessary component of AI adoption in the pursuit of better addressing cybercrime.

In the context of the risk that AI systems may pose to fundamental rights, specific attention should also be given to the concerns that have been raised about how AI may affect data privacy. The data-intensive nature of AI is a direct threat to data privacy any time the data include 'personal data'.[56] Furthermore, as noted by the Council of Europe Ad Hoc Committee on Artificial Intelligence, 'A right to privacy implies a right to a private space free from AI-enabled surveillance as necessary for personal development and

[56] For a useful illustration of the relationship between data privacy law and AI, see Sartor, G. and Lagioia, F. (2020) 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence'. European Parliamentary Research Service, June. See also Gonzalez Fuster, G. (2020) 'Artificial Intelligence and Law Enforcement - Impact on Fundamental Rights'. Study for the European Parliament (pp. 38–40).

democracy.'[57] There are, of course, many forms of 'AI-enabled surveillance', both for the online cybercrime environment and for the 'real' offline world. However, limiting the focus to the criminal justice systems context, some forms of predictive policing may fall within the category of AI-enabled surveillance.

5.14 Practical concerns

The above discussion of risks and challenges has focused mainly on risks posed by the adoption of AI systems for judicial decision-making in criminal justice systems, with an emphasis on the cybercrime context. Turning to challenges of a more practical nature, it may be noted that a study published in July 2021 points to data quantity, quality or availability as the main challenge in developing or deploying AI.[58] Other challenges noted in the same study include 'lack of available talent in the workforce', 'bias' and 'understanding appropriate governance and control'.[59]

Further, concerns have been expressed about intellectual property issues and security risks. The former relates to the impact that intellectual property and trade secret protection may have in preventing transparency. As to the latter, commentators have pointed both to risks of manipulation and to risks of disruption (e.g., via ransomware attacks).

6. Conclusion

Section 4, describing perceived benefits, demonstrates that there is a value in pursuing the adoption of AI systems for judicial decision-making in criminal justice systems as a tool to improve how we handle cybercrime. At the same time, given the sensitive nature of the AI uses discussed (Section 3) and taking account of the many well-founded risks and challenges outlined (Section 5), Law Ministers may wish to proceed with caution. Establishing a framework that enables developments to take place in a safe, accountable and rights-respecting manner may be regarded as an essential first step.

In the end, all the above may arguably be distilled into one thesis – that is, when adopting AI systems for judicial decision-making in criminal justice systems in the pursuit of important goals such as access to justice, it is necessary to ensure that the quality of the justice system is not compromised in the process. Thus, requirements imposed on AI systems, and their use, for judicial decision-making in criminal justice systems should not be viewed as obstacles to progress. Rather, such requirements – including those recommended in the section below – are necessary building blocks for the long-term

[57] Council of Europe Ad Hoc Committee on Artificial Intelligence (2020) 'Feasibility Study'.

[58] AI Asia Pacific Institute (2021) 'Trustworthy Artificial Intelligence in the Asia-Pacific Region'. July. https://aiasiapacific.org/wp-content/uploads/2021/07/2021-Trustworthy-Artificial-Intelligence-in-the-Asia-Pacific-Region.pdf (p. 20). As for access to data, it has been noted that, via the recourse to remote online legal proceedings, a wealth of new data is being generated that may serve to inform and train AI systems (Caianiello, 2021, 'Dangerous Liaisons', p.7).

[59] AI Asia Pacific Institute (2021) 'Trustworthy Artificial Intelligence' (p. 20).

success of AI systems for judicial decision-making in criminal justice systems in the fight against cybercrime. Against this background, it is now appropriate to present some recommendations on how we ought to proceed in this area.

7. Recommendations

A study of relevant policy documents, academic literature, ethical frameworks and proposed legal instruments shows a considerable degree of consistency in what is held to be required in the adoption of AI systems for judicial decision-making in criminal justice systems. Drawing upon such works, as well as the discussion above, several recommendations can be made for consideration by the Commonwealth member countries, and beyond. These recommendations are specifically discussed in the context of the adoption of AI systems for judicial decision-making in criminal justice systems in the fight against cybercrime. However, they ought to have equal applicability in any discussions of the adoption of AI systems for judicial decision-making in criminal justice systems. Indeed, the principles canvassed here should serve as a useful guide for any adoption of AI systems for judicial decision-making whether in the civil or the criminal justice systems.

As already alluded to, this article strongly advocates the creation of a clear framework that enables developments for the adoption of AI systems for judicial decision-making in criminal justice systems to take place in a safe, accountable and rights-respecting manner safeguarding the values of the Commonwealth,[60] including the rule of law as an essential protection for the people of the Commonwealth and as an assurance of limited and accountable government. AI systems aimed at combatting cybercrime should not be adopted in a manner that undermines an independent, impartial, honest and competent judiciary or the independent, effective and competent legal system as an integral component to upholding the rule of law, engendering public confidence and dispensing justice.

The Commonwealth Secretariat is well placed to play a crucially important role by facilitating the development of such a framework, whether it is ultimately articulated as a coherent Model Law at the Commonwealth level[61] or merely as model provisions recommended for consideration at a domestic level.

7.1 The fundamental rights principle

A framework for the safe, accountable and rights-respecting adoption of AI systems for judicial decision-making in criminal justice systems must be anchored in, and take care to integrate, international and domestic human rights law, and all other fundamental rights

[60] See for example the Commonwealth Cyber Declaration 2018, emphasising the 'Commonwealth values of human rights, tolerance, respect and understanding, freedom of expression, rule of law, good governance, sustainable development and gender equality'.

[61] See the Commonwealth Cyber Declaration 2018.

and values of free and democratic societies (the 'fundamental rights principle'). Several such rights are of relevance in the fight against cybercrime, including principles of equality before the law, of the presumption of innocence, of the right to a fair and public hearing by a competent, independent and impartial tribunal established by law and of the right to be tried without undue delay.

7.2 The rule of law principle

In addition, such a framework must be supportive of the multifaceted concept of the rule of law – both in the sense of directly supporting the rule of law and in the sense of working to enhance trust in the rule of law (the 'rule of law principle'). In fact, a rule of law focus may usefully be applied as a filter in the sense that any adoption of AI systems for judicial decision-making in criminal justice systems that supports the rule of law ought to be explored, and any adoption of AI systems in this setting that undermines the rule of law must be rejected even where they may otherwise prove effective tools against cybercrime – the cure should not be worse than the issue it seeks to address.

7.3 The lifecycle principle

Steps to guarantee adherence to, and support for, fundamental rights and the rule of law must be taken throughout the AI systems' entire 'lifecycle' (the 'lifecycle principle'), and that lifecycle may be split into a number of stages.[62] Thus, the regulation of AI in general, and the adoption of AI systems for judicial decision-making in criminal justice systems in particular, must be subject to ongoing monitoring, review and evaluation. This ongoing work should involve both public and private actors, and benefit from extensive consultation, audits, democratic scrutiny[63] and multistakeholder input. This is especially important in the context of cybercrime, where private actors play a significant role in how we respond.

7.4 The justification principle and the precautionary principle

Various tools may be pursued during the different stages of the AI lifecycle. For example, an important aspect in the 'design stage' is to promote responsible innovation through tools such as 'human rights by design', 'privacy by design' and 'ethical by design'. In addition, in relation to the 'deployment stage', Law Ministers may wish to consider and

[62] For example, a document published by The Alan Turing Institute speaks of the 'design stage', the 'development stage' and the 'deployment stage' (Leslie, D., Burr, C., Aitken, M. et al. (2021) 'Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: A Primer'. The Council of Europe, pp. 10-12). Similarly, the Organisation for Economic Co-operation and Development speaks of four different phases: 'AI system lifecycle phases involve: *i)* "design, data and models"; which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building; *ii)* "verification and validation"; *iii)* "deployment"; and *iv)* "operation and monitoring"' (OECD, 2019, 'Recommendation of the Council on Artificial Intelligence'. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449).

[63] See for example the Montreal Declaration for a Responsible Development of Artificial Intelligence (www.montrealdeclaration-responsibleai.com/the-declaration), Principle 5.

adopt the 'justification principle' and the 'precautionary principle'. The justification principle means that, for any proposed adoption of AI systems for judicial decision-making in criminal justice systems, that proposal must be justified by reference to specific benefits and the achievability of the postulated benefits must be demonstrated. The justification principle will encourage a purposeful, rather than hype-driven, adoption of AI in this sensitive context. The precautionary principle signifies that, in any situation where it may reasonably be suspected that an AI system may cause harm, those proposing the adoption of the AI systems for judicial decision-making in criminal justice systems bear the burden to show that the system may be safely adopted. As one commentator notes, 'Ill-advised uses of AI need to be identified in advance and nipped in the bud, to avoid harm to important values.'[64]

In addition to these overriding considerations, a framework for the safe, accountable and rights-respecting adoption of AI systems for judicial decision-making in criminal justice systems may usefully incorporate several more specific key principles, many of which are recurring in the literature about the regulation of AI.

7.5 The appealability principle

Law Ministers may wish to consider and adopt the 'appealability principle' – that is, for any decision made by AI for the purpose of judicial decision-making in criminal justice systems, it must always be possible to appeal the decision to a human. In fact, Law Ministers may wish to consider embracing a ban on AI as final arbiter in the context of judicial decision-making in criminal justice systems in the context of both cybercrime and offline crime.

7.6 The explainability principle

Law Ministers may wish to consider and adopt the 'explainability principle' essential for the above-mentioned appealability principle, for upholding justice and dignity for those affected by a decision, and for facilitating society's monitoring of justice and equality in judicial decision-making. Under this principle, any decision made, or supported, by an AI system must be explainable to be valid. The principle covers both 'ex ante explainability' (i.e., the decision-making process being explainable prior to its use) and 'ex post explainability' (i.e., the decision-making process being explainable after its use).[65] This is particularly significant in the context of the adoption of AI systems for judicial

[64] Clarke (2019) 'Guidelines for the Responsible Business Use of AI'.

[65] 'Only some algorithmic methods lend themselves to ex ante transparency, notably those relying on decision trees. ... in the case of other algorithmic technologies, such as neural networks, the machine is learning as it processes the data and it is not possible to set out the reasoning in advance' (Black, J. and Murray, A., 2019, 'Regulating AI and Machine Learning: Setting the Regulatory Agenda'. *European Journal of Law and Technology* 10(3)).

decision-making in criminal justice systems. Indeed, it may be said to flow from relevant fundamental rights. Adherence to the explainability principle may usefully incentivise further work on what has been termed 'explainable AI'.[66]

7.7 The transparency principle

Law Ministers may wish to consider and adopt the 'transparency principle'. This principle is related to, and partly overlaps with, the explainability principle. However, it does not only relate to the need for transparency in the sense of explainability. It also calls for transparency in the sense of persons being made aware of the fact that AI has played a role in a decision made about them, what methods were used and, at least, what parameters the AI system considered. Further, the transparency principle emphasises the need for law to clearly identify what decisions may be made by, or partially made by, AI systems. The transparency principle may be in tension with intellectual property and trade secret protections afforded to the developers of AI systems. In such situations, the rule of law demands that only systems that fulfil the transparency principle are adopted. This may be considered in the rules governing the procurement process of AI systems for judicial decision-making in criminal justice systems and may even point to a need to explore governments playing a role in the design and creation of AI systems for judicial decision-making in criminal justice systems.

7.8 The non-discrimination principle

Given the prominence of the risk of AI systems introducing, augmenting or re-introducing discrimination between individuals or groups of individuals, Law Ministers may wish to specifically consider and adopt the 'non-discrimination principle', requiring an ongoing commitment to eliminate discrimination, and risks of discrimination, in the adoption of AI systems for judicial decision-making in criminal justice systems. There are two dimensions to this principle. It aims to utilise AI systems to eliminate existing discrimination and it aims to prevent AI systems, one way or another, introducing discrimination. On a practical level, this may take several forms. For example, attention can be directed at what variables AI systems use as the basis for their decisions. Where the variables include examples of sensitive data, such as on gender, political opinions or ethnicity, that may be used in a discriminatory manner, special steps may be required to ensure the system does not unfairly discriminate between individuals or groups of individuals. Furthermore, as noted by CEPEJ, 'the use of machine learning and multidisciplinary scientific analysis to combat such discrimination should be encouraged.'[67]

[66] See further Deeks, A. (2019) 'The Judicial Demand for Explainable Artificial Intelligence'. *Columbia Law Review* 119(7): 1829-1850.

[67] CEPEJ (2018) 'European Ethical Charter' (p. 9).

7.9 The quality assurance principle

Law Ministers may wish to consider and adopt the 'quality assurance principle'. A reliable application of AI systems for judicial decision-making in criminal justice systems must be able to maintain quality assurance and should reliably operate in accordance with its intended purpose, over its lifecycle; while close enough may be good enough in some settings, that is not the case where AI is applied in a criminal justice setting such as in the fight against cybercrime. This places quality and robustness requirements on both the AI system as such and the data it uses. Further, it places quality requirements on the operation of the system by those using it. Certification schemes and external audits, and the involvement of external, independent, expert assessment, may be valuable in this context.

7.10 The resilience principle

While the Commonwealth has already commenced important work on cybersecurity,[68] the world's cyber-dependence has by far outpaced efforts aimed at ensuring cyber-resilience. This has created serious societal vulnerabilities, which are frequently exploited by criminals and hostile state actors. Thus, Law Ministers may wish to consider and adopt the 'resilience principle'. Under this, there should be no situation of full AI dependence. All systems must include back-up features ensuring continuous functionality of the judicial system even where a particular AI system is attacked or otherwise fails. The resilience principle also imposes cybersecurity obligations on users of AI systems, meaning that all reasonable steps must be taken to ensure system integrity, and to avoid manipulation and unlawful access. In this context, users must be mindful that manipulation can take many forms. For example, and also where the algorithms are operating properly, data may have been manipulated either to either cause undue outcomes in a specific instance or to impact the long-term operation of the system.

7.11 The human oversight principle

Many of the risks and challenges identified in Section 5 may be mitigated where the structures adopted include appropriate human oversight,[69] review, audits and intervention. Thus, Law Ministers may wish to consider and adopt a 'human oversight principle' mandating such oversight.

[68] See the Commonwealth Cyber Declaration 2018.

[69] See for example New Zealand Government (2020) 'Algorithm Charter for Aotearoa New Zealand'. Statistics NZ. https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

7.12 The accountability principle

Law Ministers may wish to consider and adopt the 'accountability principle' that partly overlaps with some of the previously noted principles. The important role accountability can play in technology regulation is widely recognised,[70] and, as outlined by the Australian Human Rights Commission:

> *Accountability involves ensuring that the law is followed in a decision-making process. It includes both a* corrective *function, facilitating a remedy for when someone has been wronged, as well as a* preventive *function, identifying which aspects of a policy or system are working and what needs adjustment.*[71]

7.13 The human-centricity principle

Law Ministers may wish to consider and adopt the 'human-centricity principle' often highlighted in works discussing the regulation and ethics of AI systems.[72] As noted by one such work, 'Put simply, a human-centric approach to AI is placing humans and the human experience at the centre of design considerations and intended outcomes of AI technologies.'[73] Importantly, even where used to pursue legitimate goals such as enhancing the efficiency of tools against cybercrime, the adoption of AI systems for judicial decision-making in criminal justice systems must not undermine human dignity.

7.14 The need for 'red lines'

The above has already alluded to the value of a framework, such as that discussed, containing clear 'red lines' delineating types of AI uses that are incompatible with the values of the Commonwealth member countries. For example, Law Ministers may, as noted, wish to articulate a ban on AI as final arbiter in the context of judicial decision-making in criminal justice systems. Similarly, AI systems' use for general biometric surveillance and for 'social scoring', for example, could be specifically banned.

7.15 Structural arrangements, collaboration, co-ordination and information-sharing

In addition to the framework and principles canvassed above, Law Ministers may wish to explore structural arrangements that support the safe adoption of AI systems for judicial decision-making in criminal justice systems so as to support the fight against cybercrime. For example, the Australian Human Rights Commission has recommended

[70] See for example the work of the Institute for Accountability in the Digital Age: https://i4ada.org/.

[71] Australian Human Rights Commission (2021) *Human Rights & Technology Final Report* (p. 51). See also UNOHCHR (2013) *Who Will be Accountable? Human Rights and the Post-2015 Development Agenda*. HR Pub/13/1, 2013.

[72] See for example Singapore's Model AI Governance Framework (Second Edition) (2020). www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

[73] AI Asia Pacific Institute (2021) 'Trustworthy Artificial Intelligence' (p. 15).

establishment of an 'AI Safety Commissioner'.[74] The ambit of such a role would include, but also go beyond, the topic of adoption of AI systems for judicial decision-making in criminal justice systems. Law Ministers may wish to consider establishing such a position in their respective countries; where this is done, they may also wish to consider establishing a structure for active collaboration and co-operation between their respective AI Safety Commissioners, perhaps in the form of a Council of Commonwealth AI Safety Commissioners.

More broadly, there are clear benefits to be gained from collaboration, co-ordination, information-sharing, sharing of best practices and support among the Commonwealth member countries in the context of the adoption of AI systems for judicial decision-making in criminal justice systems, cybercrime and AI regulation in general. There is a longstanding recognition that co-operation among Commonwealth member countries is essential in the digital arena,[75] and joining forces may facilitate the policy coherence, joint initiatives and interoperability necessary for interaction among Commonwealth member countries, for example in the context of criminal justice, not least in the context of emerging technologies such as AI and constantly evolving fields such as cybercrime.

Such work can also help address inequality of resources and degrees of development among member countries. In this latter respect, Law Ministers may wish to consider and adopt shared training and training resources, as well as enhanced digital literacy programmes, for example for courts but also for the legal communities more broadly.[76]

7.16 The need for realistic expectations

Where these structures, and the proposed framework with its numerous principles, are adopted, Commonwealth member countries are well placed to enjoy the benefits that AI systems may bring for judicial decision-making in criminal justice systems as a tool to address cybercrime, while at the same time being in a position to manage the risks and challenges involved. Nevertheless, for the foreseeable future, the adoption of AI systems for judicial decision-making in criminal justice systems is likely to provide a support role rather than autonomous decision-making. This is appropriate and discussions must proceed with realistic expectations and an acute awareness of the difference between marketing samples and products ready for safe and compliant, rights-respecting, transparent implementation. Quite a lot of AI is presented as 'almost' perfect, with the

[74] Australian Human Rights Commission (2021) *Human Rights & Technology Final Report* (pp. 127–135).

[75] See the Commonwealth Cyber Declaration 2018.

[76] For example, CEPEJ recommends that 'Generally speaking, when any artificial intelligence-based information system is implemented there should be computer literacy programmes for users and debates involving professionals from the justice system' (CEPEV, 2018, 'European Ethical Charter', p. 12). The emphasis on increasing digital literacy is also found in Commonwealth initiatives such as the Commonwealth Cyber Declaration 2018, stating that 'digital literacy can be a powerful catalyst for economic empowerment and inclusion, and commit to take steps towards expanding digital access and digital inclusion for all communities without discrimination and regardless of gender, race, ethnicity, age, geographic location or language.'

promise of perfection around the proverbial corner. However, we must remain alert to the difference between near perfect and actually perfect; progress to the point of near perfection is no guarantee for ever reaching the stage of actual perfection.

Further, there are at least three other considerations that must be flagged briefly so as to at least bring them to the attention of Law Ministers. First, given the centrality of data for the proper operation of AI systems, it is necessary to engage with the supply, quality and protection of the required data. Many countries have started emphasising the value of 'open data' but such arrangements need detailed regulation and thoughtful approaches;[77] at the same time, several countries are adopting 'data localisation' measures.[78] Second, steps must be put in place to address the risk of 'AI systems overreliance' developing in judicial decision-making. Law Ministers may wish to consider and adopt specific administrative tools for monitoring and eliminating such overreliance. Third, Law Ministers may wish to consider and adopt steps aimed at avoiding oversimplification of the law stemming from pressures to make it easier for AI systems to work with the law. Put simply, law is painted with all the colours of the spectrum and, while legal certainty and clarity are valuable goals, we must avoid making the law black and white just to cater to machines applying it.

7.17 Recalling the difference between 'can' and 'should'

In the end, there are two overarching key questions with which Law Ministers must engage. The first is whether AI systems *can* play a role in judicial decision-making in criminal justice systems to address cybercrime. Engaging with this question requires account to be taken of matters such as technical limitations (including limitations imposed by the extent to which law in general – as opposed to specific aspects of law – can be reduced into code), compliance limitations (including the extent to which our current law caters to AI systems playing a role) and the constantly developing AI technology. It is also necessary to consider the complex relationship between technology and law. For example, if benefits may be gained from AI systems without undermining fundamental values, there may be instances where laws and procedures could be amended to better accommodate such AI systems. In other words, law and technology must be approached as a system.

The second question is whether AI systems *should* play a role in judicial decision-making in criminal justice systems to address cybercrime. This is the more important question, and it is a question that benefits from multistakeholder input and public debate.

[77] See further: Gloria Gonzalez Fuster, G. (2020) 'Artificial Intelligence and Law Enforcement' - Impact on Fundamental Rights, (July 2020) https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf, at (pp. 28-–29).

[78] See further Svantesson, D. (2020) 'Data Localisation Trends and Challenges: Considerations for the Review of the Privacy Guidelines'. Digital Economy Paper 301. Paris: OECD.

While obviously related, the ‘can’ question and the ‘should’ question are best approached separately. Failure to do so would create a risk that the answer to the question of whether AI *should* play a role in judicial decision-making in criminal justice systems to address cybercrime is overshadowed by the excitement and hype often associated with the question of whether AI *can* play such a role.