

Supporting responsible AI: discussion paper University of Newcastle response



Definitions

The University of Newcastle welcomes the opportunity to provide feedback. Our University has several experts working in the broader area of Artificial Intelligence (AI), including in machine learning (ML), smart robots, neural information processing, digital health, pattern recognition, and law, specifically in the areas of the intersection of technology and law. Our experts' collated feedback is reflected in the responses to select questions.

Question 1: Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?

The definition of 'AI' under the subheading 'Technologies' is appropriate given the complexity of defining this terminology and its basis on the ISO definition. The definition is however restricted because it refers to human-defined objectives. However, a fully autonomous AI system could develop its own objectives or follow the objectives of other AI systems.

The definition of ML could be rephrased to refer to algorithms, methods, or the field, rather than 'patterns'. University of Newcastle researchers advise they sometimes use the following two short definitions, although they are also restricted:

The field of AI tries to make computers do things that, when done by people, are described as having indicated intelligence. [Brooks, R. (1991). New approaches to robotics. Science, 253:1227–1232.]

ML addresses the study of computer algorithms that improve automatically through experience. [inspired by Tom Mitchell]

For the purposes of the report, a general definition of AI alone may be inadequate because the field of AI comprises many different techniques with different properties where safe and responsible use may require different approaches and regulations. To discuss possible risks and associated ethical or legal implications of the various forms of AI it may be useful to distinguish between different types of AI.

Hence, we suggest distinguishing several categories of AI as below.

- **Traditional AI (TrAI):** ML, data science, and optimization techniques: Logic-based AI, decision trees, support vector machines, traditional neural networks, evolutionary algorithms, and many other ML and statistical techniques for classification, regression, and clustering as well as techniques for optimisation and search.
- **Traditional deep learning (DL):** including adversarial learning using models that can be trained, also via transfer learning, and applied by everyone with basic computing skills and standard hardware. We teach how to use DL at our universities around the world (Zhang et al., 2021).
- **New data-driven generative methods such as large language models (LLM):** ChatGPT, Hyper Clova X, PaLM 2 (Ghahramani, 2023) and others (see e.g. Table 5 in Goldstein et al. (2023)).

- **Self-taught models (ST):** where the agents learn in self-play without the use of humangenerated data: Alpha-go Zero (Silver et al., 2017), AlphaZero, MuZero, Alphastar (Vinyals et al., 2019).
- **Embodied AI systems (E):** such as autonomous cars, Tesla bot, autonomous drones, and some relevant IoT or Industry 4.0 systems.
- **Future brain-like systems (BL):** that could be suitable combinations of any of the above

In addition to “ML’ and ‘generative AI models’, we suggest that ‘natural language processing’ (NLP) might be included as AI because arguably there are nuanced differences between the two. The increasing uptake of chatbots across different industries, including the University of Newcastle’s ‘know the law’ app to support international students, are primarily based on NLP rather than ML.

Potential gaps in approaches

Question 2: What potential risks from AI are not covered by Australia’s existing regulatory approaches? Do you have suggestions for possible regulatory action to mitigate these risks?

One issue that is not covered relates to neurotechnology developments. This is not necessarily well covered (or even considered) as AI systems and tools are perceived to be separate from humans (rather than potentially integrated).

Urbina et al. (2022) discovered that readily available ML software can be used to design biochemical weapons. They also provide some mitigation strategies (Table 1 of Urbina et al., 2023).

Urbina, F., Lentzos, F., Invernizzi, C., and Ekins, S. (2022). Dual use of artificial-intelligence powered drug discovery. *Nature Machine Intelligence*, 4:189–191.

Urbina, F., Lentzos, F., Invernizzi, C., and Ekins, S. (2023). Preventing ai from creating biochemical threats. *Journal of Chemical Information and Modeling*, 63(3):691–694

Federated learning could be made a requirement if the data is sensitive.

Question 3: Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.

The Government might take further steps to enhance the public awareness of AI, and its use in the public and private domains, with a view to triggering meaningful public debate on the associated benefits and risks. AI has impacted and will continue to impact each member of society, so it is important people understand what is happening and how their lives, legal rights (privacy, right to their bioinformation for example) and more may be affected. A coordinating body within the Government that takes clear lead during the public discourse could be helpful to achieving this understanding.

AI education in schools is also important, to teach basic principles of AI and ML to increase understanding and reduce fear and concerns. This can be combined with ethical training.

A more evolved ethical framework could also be helpful to guide decision making about AI. While the principles referred to are helpful, a focus on levels of ethical abstraction could provide further guidance. For example, see at *Appendix 1* a model developed for Judicial use of AI by Professor Tania Sourdin (See Sourdin, T ‘Judges, Technology and Artificial Intelligence’).

The Government might also consider supporting a National Institute for Data Forensics and AI Security to complement existing efforts in cyber security and support the development of AI tools that can detect inappropriate use of AI.

Universities could work with Government to establish interdisciplinary AI centres that provide fora for regular discussion of AI related issues. The centres should help to keep the community (academics, students, industry, school, etc) up to date with the fast developments of AI and allow for the discussion and implementation of responsible AI practices and ethical training.

AI should be considered in the current round of consultation to establish national research priorities.

Question 4: Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia.

There is support for a range of light touch, medium and high touch regulatory responses. A multi-disciplinary AI Council that includes academics, researchers and innovators will assist to address the challenges of categorization and formulating responses so that where self-regulatory or industry-regulatory responses are appropriate there is some reporting capacity and provision exists for review mechanisms (e.g. via guidelines – see for example guidelines at PS 139 - <https://download.asic.gov.au/media/5689909/rg139-published-13-june-2013-20200727.pdf>)

‘Sandbox’ arrangements could also be supported to enable and support innovation. The AI Council could usefully adopt a real ‘co design’ approach with a focus on 4 or 5 industries, Government or professional areas each year.

Responses suitable for Australia

Question 5: Are there any governance measures being taken or considered by other countries (including any not discussed in this paper) that are relevant, adaptable and desirable for Australia?

Hong Kong is a vibrant financial market where AI tools are increasingly used by the financial service industry and an innovation hub in the world. There are also around 100,000 Australians living and working in the city. As such it is important for the Australian Government to understand the relevant practices there which would impact the interests of one of the largest Australian communities abroad.

In addition, the Justice Sector requires special attention (as has been the case in Europe). How AI judges other forms of AI or replaces, nudges, or corrects human judges should be considered.

Target areas

Question 6: Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ?

An artificial delimitation between the approaches to public and private sector use of AI technologies might not be useful. Instead, approaches should be based on risk level assessment outcomes for each application of AI. In this regard, Box 4 in the discussion paper serves as a good start.

A quality culture rather than a compliance culture should be fostered and supported within Government. There are already additional review mechanisms that apply to Government – if these work correctly (e.g., via the new Administrative Appeals Tribunal Act), there should be no need to consider additional regulatory layers (see however the ethical abstraction approach referred to above). The approach to triage as low, medium, and high risk, consistent with the European Union's (EU) AI Act is sensible, and will make best use of limited resources, i.e., limiting robust impact assessment where high risk approaches are considered (and judicial decision making should fall within this category).

Question 7: How can the Australian Government further support responsible AI practices in its own agencies?

Supporting legal and policy researchers based in universities to undertake policy related research work could help underpin a sound Governance framework for Australia. The Government so far has invested heavily on the technical side of AI research and could expand this in policy study. There should be a dedicated funding scheme for this type of important work.

Question 8: In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better? Please provide some examples.

From a legal perspective, clearly a great deal of AI that is 'off the shelf' can be useful and of low risk if used in particular ways. For example, law students can be encouraged to use CHATGPT and then critique it to pick up on 'hallucinations' and the issues with the creation of a simplistic narrative. Building a system that is law specific (which has already been done) might provide additional useful material for law students, lawyers, citizens, and organisations.

Question 9: Given the importance of transparency across the AI lifecycle, please share your thoughts on:

- a. **where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?**
- b. **mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.**

Expert AI systems could assist to mitigate risk.

Explainability is a complex topic area. The American Bar Association's Resolution 112 involving AI and ethics (2019) urges courts and lawyers to address emerging ethical and legal issues related to the use of AI in law, including through the explainability of automated decisions made by AI.¹ As outlined by Deeks, explainable AI can 'foster trust between humans and the system, identify cases in which the system appears to be biased or unfair, and bolster our own knowledge of how the world works.' At the same time, however, explainable AI can be costly to build and may decrease the algorithm's accuracy.²

Nevertheless, some commentators have questioned whether AI can ever be truly explainable. Burrell, for instance, has argued that because humans' reason differently to machines, they cannot always understand or interpret the interactions among data and algorithms, regardless of whether

¹ See Legal Talk Network, 'The Intersection of Ethics and Artificial Intelligence' (Podcast, 30 January 2020) <<https://legaltalknetwork.com/podcasts/digital-detectives/2020/01/the-intersection-of-ethics-and-artificial-intelligence/>>.

² Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 *Columbia Law Review* 1829, 1834.

they are suitably trained. One reason for this is the process the ML system goes through in refining its results and adjusting the 'weight' accorded to a multitude of variables.³

Surden argues that while some ML techniques such as decision trees produce answers that are easy to understand and inspect, neural-network and deep learning approaches can be extremely difficult, if not impossible, for humans to understand, including the programmers who created them.⁴ Deeks has also recognised this problem, noting that an explainable AI approach which simply reveals the source code for the ML model will rarely be satisfactory as most people will be unable to understand the code.

Nevertheless, Deeks has outlined two alternative approaches to explainable AI: a model-centric approach and a subject-centric approach. The former attempts to explain the whole model through, for example, revealing the creator's intentions or the parameters specified by the creators. The latter focuses on the model's performance in a particular case, and might, for instance, provide the subject of a decision with information about the characteristics of individuals who received similar decisions.⁵ Coglianese and Lehr have also challenged the notion that complex AI processes can never be completely explainable, noting:

'analysts can, and do, possess full knowledge of algorithms' inner workings, and they can mathematically explain how these algorithms optimize their objective functions. What they lack is simply an interpretive ability to describe this optimization in conventional, intuitive terms.'⁶

Further, the European Commission (2020) released a White Paper report touching on the explainability of AI outcomes. The report identifies a 'need to build bridges between disciplines that currently work separately, such as ML or deep learning and symbolic approaches.'⁷ ML and deep learning are 'characterized by limited interpretability, the need for a large volume of data to train the models and learn through correlation.'⁸ Symbolic approaches 'are where rules are created through human intervention.'⁹ The report suggests that 'combining symbolic reasoning with deep neural networks may help us improve explainability of AI outcomes.'¹⁰

³ Jenna Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms' (2016) 3(1) *Big Data & Society* 1.

⁴ Harry Surden, 'The Ethics of Artificial Intelligence in Law: Basic Questions' (Draft paper for the Oxford Handbook of Ethics of AI, 2020).

⁵ Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 *Columbia Law Review* 1829, 1835–7.

⁶ Cary Coglianese and David Lehr, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 *The Georgetown Law Journal* 1147, 1206–7.

⁷ European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust* (White Paper, 2020) 5.

⁸ European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust* (White Paper, 2020) 5.

⁹ European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust* (White Paper, 2020) 5.

¹⁰ European Commission, *White Paper on Artificial Intelligence - A European approach to excellence and trust* (White Paper, 2020) 5.

Question 10: Do you have suggestions for:

- a. **Whether any high-risk AI applications or technologies should be banned completely?**
- b. **Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?**

Banning certain high-risk activities may be appropriate under certain circumstances.

From a legal perspective, there are some concerns about neuro tech. What happens if a company goes 'bust' – what impact might there be if neurotech includes embedded materials (micro bots or other)? Other concerns are about forms of AI replacing judges (so that AI then is effectively judging humans and other forms of AI) which is particularly relevant in a common law as innovation will be stifled and because of significant democratic risks (amongst others).

Question 11: What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?

A good step would be for the government to develop 'friendly' chat bots, responsive and accurate tech approaches and have humans in the background to ensure that there are humans 'on' the loop as well as 'in' the loop. There needs to be greater attention paid to the need to support vulnerable people, those who have historically had less exposure to advanced technology and to education. Universities, as anchor institutions in their communities, can assist, including via AI Knowledge Centres located at the universities and curricula, to provide education to a broader community audience, and support for industry.

Risk-based approaches

Question 14: Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?

A risk-based approach of four risk categories as proposed by the EU AI Act seems to be sensible. Potentially a more detailed scale of 6-7 categories or a matrix approach could be developed and considered. This may allow for more streamlined assessments.

Question 15: What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

Benefits:

- Sensitive applications or methods could be analysed, monitored and restrictions could be applied.
- Companies or universities working on desirable high-risk applications could receive dedicated support.
- Some misuse of AI could be prevented.

Limitations:

- Inappropriate use of AI techniques may still occur.

Question 16: Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?

While potentially a consequence of applying a risk-based approach, placing additional restrictions on start-up companies could use additional resources and prevent them from becoming successful. Instead, additional dedicated support and monitoring could be offered.

Question 17: What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?

There is support for the elements of attachment C. However, there are some weaknesses, e.g., it may be challenging to have enough human experts and assessors available to implement all elements.

If a large language model is trained to indoctrinate people and, e.g., manipulates an election outcome, then this may only be detectable by careful analysis using sophisticated ML tools. Just keeping 'a human in the loop' or imposing some regulations may not be sufficient.

Question 18: How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?

It must be recognised that much of the risk is not due to AI techniques but due to human error, malpractice, misuse and the risks and dangers of existing applications, e.g., in the areas of biotechnology or arms trade. In most circumstances, new AI only amplifies the existing risks in these domains. Hence, the existing regulations and assessment processes in these high-risk application domains need to be updated to take the impact of AI into account.

Question 19: How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?

OpenAI has some strategies in place that could also be used or replicated on a national level. Global approaches and international collaboration are desirable, and universities are in a good position to facilitate this through the establishment of AI centres that then internationally can collaborate on these issues, in collaboration with government and industry.

Note:

This Table has been prepared by Tania Sourdin and appears in 'Judges, Technology and Artificial Intelligence' (2021) Edward Elgar. The Table extends framework approaches that have been developed in the medical arena which consider levels of abstraction that are relevant in the medical field see Jessica Morley, Caio CV Machado, Christopher Burr, Josh Cows, Indra Joshi, Mariarosaria Taddeo and Luciano Floridi, 'The Ethics of AI in Health Care: A Mapping Review' (2020) 260 *Social Science & Medicine* 113, 132. The author notes that the levels of abstraction and thematic ethical concerns are not discrete and may overlap.

Levels of Abstraction							
		Individual	Interpersonal	Group	Institutional	Sectoral	Societal
Thematic Ethical Considerations and Concerns	Technical Concerns (inaccurate, biased, distorted or inappropriate underlying algorithmic processes)	Erroneous or inappropriate judgment (or other legal decision, such as an administrative decision) could take place.	There could be a loss of trust in the judge/litigant or judge/advocate relationship and the justice system more broadly; de-humanisation of justice process. ⁱ	There could be an erroneous or inappropriate application of precedent to a particular group of people, cases or a particular area of law. ⁱⁱ	There could be an inappropriate use of funds or circumstances where resources are not directed to courts and organisations that are established to assist people in exercising their rights. This could, in turn, cause negative impacts on the most vulnerable. ⁱⁱⁱ	Judge AI and AI tools could be used in areas of law where there are clear expressed concerns ^{iv} or where precedent is required.	Poorer provision of justice services throughout society and either a magnification of the delays and backlogs that plague many legal systems ^v or a decline in trust in government and justice arrangements.
	Process, Outcome and Purpose Concerns (notions of fairness, privacy and security, transparency and explainability, and contestability and accountability)	Individuals may be unable to access or assess the information used by an AI Judge to make a decision, ^{vi} and their ability to appeal a decision could be lost or made more difficult. ^{vii} See also 'Sectoral'.	There could be an overreliance on AI systems and their ability to make the 'right' decision, leading to the usurpation of the judicial role. ^{viii}	Throughout their lifecycle, AI systems may not be built to be inclusive or accessible and could involve (or result in) discrimination against certain groups. ^{ix}	AI actors may not provide meaningful information to authorize external audits ^x or ensure stakeholders are aware of their 'interactions' with AI tools, including in the workplace. ^{xi}	Private data (such as social surveillance data) ^{xii} may be used in the making of judicial decisions. ^{xiii} Data protection may be inadequate or principle relating to 'open' justice may not be maintained.	Societal inequalities in process, outcome and purpose. ^{xiv}
	Human-Centred Concerns	The design and implementation	Preference may not be given to a	The systems may promote unfairness or	The autonomy, respect and	The systems may be replicated	Societal respect for justice

	(wellbeing and human-centred values)	on of AI tools and Judge AI could be fundamentally incompatible with human rights (such as due process rights or the right to appeal a decision). ^{xv} In addition, the tools may not support human wellbeing and dignity.	‘human-right-by-design approach’. Thus the ‘rules’ prohibiting direct or indirect discrimination between individuals, groups or a sector of society may not be ‘fully integrated’. ^{xvi}	injustice. This could, in turn, lead to a loss of human dignity and a reduction in societal wellbeing. See also ‘Interpersonal’.	independence of large-scale institutions including the courts may be lost as the development of Judge AI tools impacts on the role and function of the judiciary. ^{xvii}	across the sector with little regard for broader impacts on human wellbeing. See also ‘Interpersonal’.	process, democratic values, the rule of law and the judge’s independence in the decision making process (viz. the exercise of discretion) ^{xviii} may be lost. ^{xix}
--	--------------------------------------	---	---	---	---	---	--

ⁱ The ‘dehumanisation’ of justice processes as a result of the infiltration of AI tools or Judge AI has been raised as a point of concern in the literature, see e.g.: Tania Sourdin, Bin Li and Tony Burke, ‘Just Quick and Cheap? Civil Dispute Resolution and Technology’ (2019) 19 *Macquarie Law Journal* 17, 18; Michael Kirby, ‘The Future of Courts: Do They Have One?’ (1999) 8 *Journal of Judicial Administration* 185, 191; Belinda Smyth and Bruce Fehlberg, ‘Australian Post-Separation Parenting on the Smartphone: What’s ‘App-ening?’ (2019) 41(1) *Journal of Social Welfare and Family Law* 53, 62; Felicity Bell, ‘Family Law, Access to Justice, and Automation’ (2019) 19 *Macquarie Law Journal* 103, 131–132.

ⁱⁱ Indeed there is concern that using data from previous cases may lead to the making of incorrect decisions because many civil cases that proceed to judgment are likely to be anomalous or ‘aberrant’: Naomi Burstyn, Tania Sourdin, Chinthaka Liyanage, Bahadorreza Ofoghi and John Zeleznikow, ‘Using Technology to Discover More about the Justice System’ (2018) 44(1) *Rutgers Computer & Technology Law Journal* 1. Similarly, Surden notes that where an AI tool is asked to make a decision about a case where no like decisions exist or where the sample of ‘past examples’ are ‘too few’, then the AI ‘may not be able to detect patterns that are reliable predictors’: Harry Surden, ‘Machine Learning and Law’ (2014) 89 *Washington Law Review* 87, 105–106.

ⁱⁱⁱ This is of especial concern in the context of COVID-19, see e.g.: Law Council of Australia, *Principles for Facilitating Access to Justice for Marginalised and Vulnerable Groups as a Result of the COVID-19 Pandemic* (Report, May 2020).

^{iv} The literature reveals two principal areas where the use of AI is arguably unsuitable or inappropriate: criminal and family law. McKay, for example, argues that the criminal law setting is unsuitable for automation because it is a sector ‘which is focussed on human behaviours’: Carolyn McKay, ‘Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making’ (Research Paper No 19/67, University of Sydney Law School, November 2019). See also: MM Plesničar and K Šugman Stubbs, ‘Subjectivity, algorithms and the courtroom’, in Aleš Završnik (ed), *Big Data, Crime and Social Control* (Routledge, 2018); Monika Zalnieriute and Felicity Bell, ‘Technology and Judicial Role’, in Gabrielle Appleby and Andrew Lynch (eds), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (Cambridge University Press, forthcoming). Similarly, some commentators have argued that Judge AI is inappropriate for the family law domain by virtue of the high levels of emotion that are present in such matters: Felicity Bell, ‘Family Law, Access to Justice, and Automation’ (2019) 19 *Macquarie Law Journal* 103, 131–132. It is noted, however, that there are some persuasive arguments for the use of AI in family law, see e.g.: Melissa Conley Tyler and Mark McPherson, ‘Online Dispute Resolution and Family Disputes’ (2006) 12(2) *Journal of Family Studies* 165, 170; Tania Sourdin and Chinthaka Liyanage, ‘The Promise and Reality of Online Dispute Resolution in Australia’ in Mohamed S Abdel Wahab, Ethan Katsh and Daniel Rainey (eds), *Online Dispute Resolution: Theory and Practice a Treatise on Technology and Dispute Resolution* (Eleven International Publishing, 2012) 483.

^v For a Report on the backlogs and delays that plague the Australian family law system, see: Australian Law Reform Commission, *Family Law for the Future – An Inquiry into the Family Law System: Final Report* (Report No 135, March 2019).

^{vi} Ashley Deeks, 'The Judicial Demand for Explainable Artificial Intelligence' (2019) 119 *Columbia Law Review* 1829, 1832-1834.

^{vii} Organisation for Economic Cooperation and Development, *Recommendation of the Council on Artificial Intelligence* (OECD Legal Instruments 0449, May 2019) [1.3]; Monika Zalnieriute, Lyria Bennett Moses and George Williams, 'The Rule of Law and Automation of Government Decision-Making' (2019) 82(3) *Modern Law Review* 425, 449; Lord Sales, 'Algorithms, Artificial Intelligence and the Law' (Speech, Sir Henry Brooke Lecture for BAILII Freshfields Bruckhaus Deringer, London, 12 November 2019) 10.

^{viii} Deeks argues that one of the main concerns regarding the use or development of Judge AI is that an 'overreliance on machines' may be 'at the expense of human expertise': Ashley Deeks, 'High-Tech International Law' (2020) 88(3) *The George Washington Law Review* 574, 641. Indeed Chief Justice Bathurst of Australia has argued that 'judges have "explanatory" accountability in their obligation to provide open, public justice and reasons explaining their decisions, "content" accountability in terms of the appellate process and "probity" accountability in terms of their use of public resources': Thomas Bathurst, 'Who Judges the Judges, and How Should They be Judges?' (Speech, Law Society Opening of Law Term Dinner, 30 January 2019) 8 [24] cited in Pamela Stewart and Anita Stuhmcke, 'Judicial analytics and Australian courts: A call for national ethical guidelines' (2020) 45(2) *Alternative Law Journal* 82, 85. In circumstances where Judge AI is relied on to make judgments, the author notes that this form of human 'accountability' may be lost.

^{ix} *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment* (entered into force on 3-4 December 2018) 9; Asaf Tzachor et al, 'Artificial Intelligence in a Crisis Needs Ethics with Urgency' (2020) 2 *Nature Machine Intelligence* 365, 365.

^x *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment* (entered into force on 3-4 December 2018) 11.

^{xi} Organisation for Economic Cooperation and Development, *Recommendation of the Council on Artificial Intelligence* (OECD Legal Instruments 0449, May 2019) [1.3]. For discussion on the importance of AI explainability for external stakeholders, see: Umang Bhatt, McKane Andrus, Adrian Weller and Alice Xiang, 'Machine Learning Explainability for External Stakeholders' (Workshop Paper, ICML Workshop on Extending Explainable AI, 2020).

^{xii} Such issues have been the subject of much commentary in China, see e.g.: Ali Dukakis, 'China rolls out software surveillance for the COVID-19 pandemic, alarming human rights advocates', *ABC News* (Online, 14 April 2020) <<https://abcnews.go.com/International/china-rolls-software-surveillance-covid-19-pandemic-alarming/story?id=70131355>>; James Leibold, 'Surveillance in China's Xinjiang Region: Ethnic Sorting, Coercion, and Inducement' (2020) 29(121) *Journal of Contemporary China* 46; Larry Backer, 'And an Algorithm to Entangle them All? Social Credit, Data Driven Governance, and Legal Entanglement in Post-Law Legal Orders' (Research Paper No 05, Penn State Law, 1 January 2020) 14.

^{xiii} It has also been noted that this is of particular concern in the context of COVID-19. Tzachor et al, for example, have argued that 'there is growing concern that the use of AI and data in response to COVID-19 may compromise privacy and civil liberties by incentivizing the collection and processing of large amounts of data, which may often be private or personal': Asaf Tzachor et al, 'Artificial Intelligence in a Crisis Needs Ethics with Urgency' (2020) 2 *Nature Machine Intelligence* 365, 365.

^{xiv} See e.g. Will Bateman, 'Automated Discretionary Decision-Making in the Public Sector – Legal Dimensions' (Research Paper No 20.10, ANU College of Law, 2019) 6-15; Marshall McLuhan, 'The Medium is the Message' in Marshall McLuhan (ed), *Understanding Media: Extensions of Man* (MIT Press, 1994).

^{xv} *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment* (entered into force on 3-4 December 2018) 8; Monika Zalnieriute, Lyria Bennett Moses and George Williams, 'The Rule of Law and Automation of Government Decision-Making' (2019) 82(3) *Modern Law Review* 425, 449; Heather Scheiwe Kulp, 'Future Justice? Online Dispute Resolution and Access to Justice', *Just Court ADR* (Blog Post, 8 August 2011) <<http://blog.abouttsi.org/2011/policy/future-justice-online-dispute-resolution-and-access-to-justice/>>.

^{xvi} *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment* (entered into force on 3-4 December 2018) 8.

^{xvii} For example, Stewart and Stuhmcke note that ‘the research-scape concerning judicial behaviour in the US is today dominated by three principal commercial providers of legal analytic services’: Pamela Stewart and Anita Stuhmcke, ‘Judicial analytics and Australian courts: A call for national ethical guidelines’ (2020) 45(2) *Alternative Law Journal* 82, 83.

^{xviii} Will Bateman, ‘Automated Discretionary Decision-Making in the Public Sector – Legal Dimensions’ (Research Paper No 20.10, ANU College of Law, 2019) 3-5.

^{xix} For a discussion on how Judge AI may undermine the rule of law, see: Pamela Stewart and Anita Stuhmcke, ‘Judicial analytics and Australian courts: A call for national ethical guidelines’ (2020) 45(2) *Alternative Law Journal* 82, 85-86. Such aspects of human wellbeing were also identified as of especial importance in the Organisation for Economic Cooperation and Development (OECD)’s Recommendation of the Council on Artificial Intelligence: Organisation for Economic Cooperation and Development, *Recommendation of the Council on Artificial Intelligence* (OECD Legal Instruments 0449, May 2019) [1.2]. See also: *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment* (entered into force on 3-4 December 2018) 8.