THE LAW SOCIETY OF NEW SOUTH WALES
youngLAWYERS

Submission on the Department of Industry, Innovation and Science's Discussion Paper - *Artificial Intelligence: Australia's Ethics Framework*

June 2019

Artificial Intelligence
Strategic Policy Division
Department of Industry, Innovation and Science
GPO Box 2013, Canberra, ACT, 2601
artificial.intelligence@industry.gov.au

Contact: **Jennifer Windsor**
President, NSW Young Lawyers

Ashleigh Fehrenbach
Chair, NSW Young Lawyers Communications, Entertainment and Technology Committee

Managing Editors:	Olivia Irvine, Alexandra de Zwart
Assistant Editor:	Gemma Valpiani
Contributors:	Ellen Brown, Shaheen Hoosen, Olivia Irvine, Eva Lu, Ravi Nayyar, Juanita Truong, Jaimie Wolbers, Ekaterina Zotova, Alexandra de Zwart

THE LAW SOCIETY OF NEW SOUTH WALES
youngLAWYERS

The NSW Young Lawyers Communications, Entertainment and Technology Committee make the following submission in response to the *Artificial Intelligence: Australia's Ethics Framework* Discussion Paper.

NSW Young Lawyers

NSW Young Lawyers is a division of The Law Society of New South Wales. NSW Young Lawyers supports practitioners in their professional and career development in numerous ways, including by encouraging active participation in its 15 separate committees, each dedicated to particular areas of practice. Membership is automatic for all NSW lawyers (solicitors and barristers) under 36 years and/or in their first five years of practice, as well as law students. NSW Young Lawyers currently has over 15,000 members.

The Communications, Entertainment and Technology Law Committee of NSW Young Lawyers aims to serve the interests of lawyers, law students and other members of the community concerned with areas of law relating to information and communication technology (including technology affecting legal practice), intellectual property, advertising and consumer protection, confidential information and privacy, entertainment, and the media. As innovation inevitably challenges custom, the CET Committee promotes forward thinking, particularly about the shape of the law and the legal profession.

Overview

The NSW Young Lawyers Communications, Entertainment and Technology Law Committee (**the Committee**) welcomes the opportunity to comment on Australia's Artificial Intelligence (**AI**) Framework Discussion Paper (**Discussion Paper**) on behalf of NSW Young Lawyers.

The Committee has responded to the selected questions outlined below, and have otherwise not made submissions on the remaining questions. The Committee has outlined considerations that it recommends the Department of Industry, Innovation and Science (**the Department**) take into account when reviewing these issues. The Committee hopes that these considerations provide helpful guidance to the Department in conducting this review.

In responding, the Committee has often taken a legal lens or posed key legal questions that need to be considered by the Department in creating any ethical framework for AI in the future.

While the Committee broadly agrees with the Core Principles and Toolkit items listed in the Discussion Paper, it does recommend refinement/clarification in some areas, as well as further complementary principles and considerations to be included.

Question 1: Are the principles put forward in the discussion paper the right ones? Is anything missing?

1. The Committee considers that the following changes and additions will deliver a more effective and comprehensive AI ethics framework, and will help the Department deliver on its intent of helping a wide range of stakeholders to ethically design and apply AI technologies.

2. First, the descriptions of Principle 3: Regulatory and legal compliance, Principle 4: Privacy protection and Principle 6: Transparency and explainability should be refined to more comprehensively reflect Australian laws and values. Secondly, the framework should include separate principles for "information security", "consistency", "complementarity", and "diversity and inclusion".

Principle 3: Regulatory and legal compliance

Human Rights

3. The Committee considers that human rights obligations should be explicitly incorporated into Principle 3. While this principle broadly incorporates such obligations, it is worded generally and does not expressly encourage compliance with human rights obligations in the same way as other similar frameworks, such as Google's AI Principles,[1] and the Institute of Electrical and Electronics Engineers' principles for Ethically Aligned Design.[2]

4. The Committee recommends that organisations that are responsible for developing and implementing AI technologies broadly comply with the *United Nations Guiding Principles on Business and Human Rights*[3] (as the direct coding of these obligations might not be feasible in some instances).

5. The United Nations Guiding Principles provide methods that States can employ to prevent gross human rights abuses by business enterprises, including:

   a. Engaging with business enterprises at an early stage to help them identify, prevent and mitigate the human rights-related risks of their activities and business relationships;

   b. Providing adequate assistance to business enterprises to assess and address the heightened risks of abuses, paying special attention to both gender-based and sexual violence;

   c. Denying access to public support and services for a business enterprise that is involved with gross human rights abuses and refuses to cooperate in addressing the situation; and

[1] Google 'will not design or deploy AI in … technologies whose purpose contravenes widely accepted principles of international law and human rights' Artificial Intelligence at Google: Our Principles', *Google AI* (Web Page) <https://ai.google/principles/>.

[2] IEEE seeks to '[e]mbody the highest ideals of human beneficence within human rights' by stating in its first General Principle that AI systems 'shall be created and operated to respect, promote, and protect internationally recognized human rights'. The IEEE states that '[w]hile the direct coding of human rights in A/IS [(autonomous and intelligent systems)] may be difficult or impossible based on contextual use, newer guidelines from The United Nations provide methods to pragmatically implement human rights ideals within business or corporate contexts that could be adapted for engineers and technologists' Institute of Electrical and Electronics Engineers, *Ethically Aligned Design* (IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 1 April 2019) 19.

[3] *United Nations Guiding Principles on Business and Human Rights*, UN Doc HR/PUB/11/04 (2011).

d. Ensuring that their current policies, legislation, regulations and enforcement measures are effective in addressing the risk of business involvement in gross human rights abuses.[4]

6. The Committee proposes that regulating businesses according the United Nations' Guiding Principles sets a practical goal for the enforcement of human rights obligations, and this should be outlined within Australia's AI ethics framework.

Australia's treaty obligations

7. Explicitly incorporating human rights into Australia's AI ethics framework would help ensure that Australia meets its obligations under international law. Australia is party to seven key human rights treaties as outlined in the Discussion Paper. In addition, Australia is party to several Optional Protocols, which mainly implement individual complaint mechanisms, and support the *United Nations Declaration on the Rights of Indigenous Peoples*.[5] AI systems should be developed and operated in accordance with Australia's obligations under these treaties, and, as a means of fulfilling Principle 1, should generate net-benefits by promoting human rights.

8. The Parliamentary Joint Committee on Human Rights notes that, '[t]he growing capacities for technology to be used to collect, store, access, match and share information has a range of potential human rights implications'.[6] The concerns in this area mainly relate to 'respect for informational privacy'; however, privacy concerns overlap with a range of other human rights.[7]

9. Regarding 'the matching and sharing of facial images and biometric data',[8] the Committee's submission in response to the Australian Human Rights Commission's (**AHRC**) *Human Rights and Technology Issues Paper*[9] highlights the effects that proposals such as these may have on the rights to equality and non-discrimination, and the right to privacy in relation to the data of individuals collected from various sources.

10. While uncommon in Australia at present, throughout the world there are numerous instances of AI technologies being used in ways that violate human rights, including:

    a. The persecution of Uighur people in China using machine learning tools to identify suspects by aggregating and reconciling data points,[10] violating the rights to privacy and freedom from discrimination; and

[4] Human Rights Council, *Report of the Special Representative of the Secretary General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie*, GA Res/17/31, UN GAOR, 17th sess, Agenda item 3, UN Doc A/HRC/17/31 (2 March 2011) 10-1.

[5] GA Res 61/295, UN Doc A/RES/61/295 (2 October 2007, adopted 13 September 2007).

[6] Parliamentary Joint Committee on Human Rights, Parliament of Australia, *Annual Report 2018* (12 February 2019) 16 [3.10].

[7] Ibid 17 [3.12].

[8] Ibid 17 [3.11].

[9] NSW Young Lawyers Communications Entertainment and Technology Committee, Submission to AHRC, *Australian Human Rights Commission Human Rights and Technology Issues Paper* (Submission, October 2018) 6.

[10] Meredith Whittaker et al, 'AI Now Report 2018', *AI Now Institute* (December 2018) 13.

b. The use of Automated Decision Systems in criminal court procedures, such as COMPAS in the United States,[11] potentially violating the right to a fair hearing and trial,[12] other criminal process rights,[13] and equality before the law.[14]

Principle 4: Privacy Protection

11. The Committee submits that a robust understanding of privacy law is essential to developing an ethical framework for AI. The Committee considers that in developing an Ethics Framework, the Department should take into account a more complete view of privacy laws in Australia and their application to AI than that set out in the Discussion Paper.

12. Principle 4 strongly focuses on the importance on “confidentiality” and “security” of “private data”. However, it should focus on privacy protection and adequate data governance. The Committee, in this analysis, draws from the European Commission's *Ethics Guidelines for Trustworthy AI*,[15] which covers:

    a. Privacy and data protection – ensuring the lawful collection of data initially from the user, as well as data generated about the user over the course of their interaction with the system, and that the data collected about the user will not be used to unlawfully or unfairly discriminate against them; and

    b. Adequate data governance – ensuring the quality and integrity of the data used, its relevance in light of the domain in which the systems will be deployed, its access protocols, and the capability to handle data in a manner that protects privacy.

13. The Committee recommends the Department consult further with privacy regulators and practitioners to refine Principle 4, as well as the underpinning contextual discussion.

Privacy laws in Australia

14. The *Privacy Act 1988* (Cth) (**Privacy Act**), and the Australian Privacy Principles (**APPs**), are not the only privacy laws in Australia that apply to organisations working in AI. The Discussion Paper does not refer to any state or territory privacy laws that may apply, for example, to public universities or hospitals employing AI systems. While the Discussion Paper focuses on the Privacy Act in its analysis, the Committee considers that it would also be beneficial to draw attention to these state and territory privacy laws for completeness, so that readers are aware the breadth of privacy legislation in Australia, and how the proposed framework would need to interact with the various

[11] See Dawson D et al, *Artificial Intelligence: Australia’s Ethics Framework* (Data61 CSIRO, Australia, Discussion Paper, 2019) 40-1 [5.2.1].

[12] *International Covenant on Civil and Political Rights* opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) art 14(1).

[13] Ibid art 14(2)-(7).

[14] Ibid art 26.

[15] European Commission Independent High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (Guidelines, 8 April 2019) 17 <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

limbs of that legislation. A full list of the applicable legislation is available on the Office of the Australian Information Commissioner's website.[16]

15. The Committee also notes that the European Union General Data Protection Regulation (**GDPR**) may apply to organisations in Australia employing AI systems that have an establishment in the European Union (**EU**), or offer their goods or services to, or monitor the behaviour of, people in the EU.[17]

The definition of "Personal Information"

16. Principle 4 uses the term "private data". Personal information does not need to be "private" for it to be protected under privacy laws. The Committee recommends that the term "personal information" (as defined in section 6 of the Privacy Act) should be used for Principle 4.

17. Further, the Discussion Paper often uses the word "sensitive" to describe personal data, as well as the term "sensitive data". This could be confusing as "sensitive information" is a subset of personal information and is defined under the Privacy Act.[18] Higher standards apply under the Privacy Act when sensitive information is collected, used or disclosed, and using "sensitive data" could confuse organisations as to whether these higher standards apply in the AI context. If the Ethical Framework intends to set out different obligations for sensitive information used in AI systems, "sensitive information" should be defined.

18. Protecting the consent process is not fundamental to the protection of privacy. This is because consent is not the sole mechanism by which the collection, use or disclosure of personal information may be lawfully authorised under the Privacy Act. In fact, consent is often the exception for collection, use or disclosure of personal information under the Privacy Act.[19] The Committee recommends that any ethical framework should emphasise the requirement for the lawful collection, use and disclosure of personal information, being the restrictions or limitations on collection, use and disclosure of personal information as set out in the relevant privacy laws. The Ethical Framework should also focus on how these restrictions and limitations apply to personal information collected initially from the user, as well as data generated about the user over the course of their interaction with the AI system.

Protecting privacy is more than confidentiality and security

19. Protecting privacy involves more than the obligations outlined in Principle 4.[20] Invasions of privacy could arise from the unlawful collection or use of personal information to develop an AI system, even where there was no data breach. The Committee recommends that the Ethical Framework set out the restrictions and limitations on the use and disclosure of personal information for secondary purposes, and the potential difficulties obtaining consent in the AI context.

[16] 'Other Privacy Jurisdictions', *Office of the Australian Information Commissioner* <https://www.oaic.gov.au/privacy-law/other-privacy-jurisdictions>.

[17] *EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ 2016 L 119/1, art 3.

[18] *Privacy Act 1988* (Cth) s 6.

[19] Please see Australian Privacy Principles 3.3(a), 6.1(a), 7.3(b), 7.4 and 8.2(b).

[20] Dawson (n 11) 6.

20. The Committee also recommends that the Ethical Framework set out the importance of data quality, and access to data. Transparency of algorithms can mitigate harmful privacy risks, and visibility into the data used in automated decision making can prevent skewed data input, thereby preventing the generation of biased datasets.

21. Privacy Impact Assessments may also provide a means by which the privacy impacts of AI can be assessed. Safeguards, such as data minimisation and purpose limitation, should also be implemented to prevent the unauthorised collection, use and disclosure of personal information.

Principle 6: Transparency and Explainability

22. For the reasons set out below, the Committee considers that the description of Principle 6 should convey the importance of the following values:

    a. auditability and intelligibility of an AI-enabled process operation; and

    b. public engagement and education of stakeholders about the nature of AI generally, particularly in relation to applications with the capacity to affect their lives.

a. Auditability and intelligibility

23. The Committee considers that where algorithms make decisions that affect people, more is required than being “informed” of an algorithm's existence and the information used to make decisions, to ensure that affected people understand how these decisions are made. The inclusion of auditability and intelligibility in the description of Principle 6 would better encapsulate its true values and aims, and this is consistent with the Department’s concept of ‘AI for a fairer go’.

24. First, the Committee considers that Principle 6 should explicitly prioritise auditability of the steps taken by an AI system when a decision is made, except when this would reveal trade secrets or fail to explain why a decision was made. Auditability means that a process can be examined and verified as correct, which requires that a person can see what exactly that process entails. Auditability should be a priority, not an obligation, as full auditability might be difficult for some newer iterations of AI.[21] Furthermore, differing levels of explanation would be required according to stakeholders' technical proficiency.

25. While the Committee is concerned that such a policy may encourage a “black box”, and stresses its support for the fundamental right of procedural fairness, immediate calls for full auditability, or “full technical transparency”, of AI-enabled processes would be ill-advised at this time.[22] Further investigation is needed into whether such auditability is feasible and whether different levels of auditability are required depending on the significance of the decision in question; for example, with regards to safety and human rights.[23]

26. Secondly, Principle 6 should explicitly require that AI-enabled processes be intelligible to the relevant stakeholders, namely, impacted individuals and government regulators. The Committee

[21] Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?* (House of Lords Paper No 100, Session 2017-19) 36.
[22] Ibid 38.
[23] Ibid.

recommends that entities responsible for AI-enabled processes provide explanations that are understandable to the reasonable person in the position of the impacted individual at the time they were impacted by the AI-enabled process.

27. The Committee considers that Principle 6 should enshrine the value of intelligibility, given that the 'development of intelligible AI systems is a fundamental necessity if AI is to become an integral and trusted tool in our society'.[24] The Department should investigate what sort of standards for intelligibility, if any, should be adopted by developers of AI systems. This would be in accordance with the "industry standards" element of the proposed toolkit for ethical AI.[25]

28. The Committee also considers that legislating a "right to an explanation" is an important component in effectively regulating AI in Australia and ensuring that both public and private entities engage in considered and careful design of AI systems. Importantly, any "business rules" incorporated into a system should be understandable, and the system should be able to generate a comprehensive audit trail of the decision-making path. Whilst it is reassuring that the Department is taking this approach to automated decision making within an administrative decision context, similar obligations ought to be required of the private sector.

b. Education of all stakeholders and public engagement

29. Principle 6 should more explicitly refer to the values of education and public engagement to build public trust in AI and drive greater engagement in policy development. Members of the public should have a functional understanding of the different types of AI so that they can appreciate how these technologies can be positively used, and be aware of the associated ethical issues. This will encourage the development of AI projects in a context where the ethical and societal implications have been properly considered by a wider range of stakeholders.

30. The Committee considers that the development of Australia's framework for the practical application of ethical AI would benefit from analysing the UK's national approach for AI as outlined by the House of Lords Select Committee on Artificial Intelligence in *AI in the UK, Ready, Willing and Able*,[26] and adoption of the UK approach of the promotion of education alongside the development of AI.

31. The House of Lords' report proposes a cross-sector ethical code of conduct, known as the 'AI Code'.[27] The AI Code's fourth principle promotes education alongside AI. The House of Lords' report notes:

> At earlier stages of education, children need to be adequately prepared for working with, and using, AI. For a proportion, this will mean a thorough education in AI-related subjects, requiring adequate resourcing of the computing curriculum and support for teachers. For all children, the basic knowledge and understanding necessary to navigate an AI driven world will be essential. In particular, we recommend that the ethical design and use of technology becomes an integral part of the curriculum.[28]

32. Not only will education strengthen public trust in technology, it will also allow citizens to better understand the potential benefits and detriments of AI and engage with it in a positive way. As noted above, it will also help equip the next generation to be skilled up to enter the ever-changing job

[24] Ibid 40.
[25] Dawson (n 11) 8.
[26] Select Committee on Artificial Intelligence (n 21) 120.
[27] Ibid.
[28] Ibid 6.

market. The Committee considers education to be a key factor in allowing AI to flourish. By empowering the public to interrogate the intricacies of AI and the associated policy issues, the Department and its partner organisations (from both the public and private sectors) can facilitate a richer policy debate and wider public involvement in the policy process. In turn, the Committee considers that such education would help mitigate the sense of isolation and alienation felt by those members of society who consider themselves "left behind" by technological disruption, which is embodied by AI. Education and considered public engagement are essential to delivering ‘AI for a fairer go’,[29] not least since these values can unite diverse Australian communities by facilitating an informed understanding of AI and its policy issues.

33. Like the UK, the practical application of ethical AI can be implemented within existing regulation in respective sectors in Australia. At this stage, the proposed UK AI ethics framework is a short-term solution, which has the potential to influence future regulation. The UK has suggested that an AI-specific regulation is not appropriate at this stage.

34. While the Committee considers that it is a matter for the Department to determine what precise form this education and public engagement should take, it recommends that, as a minimum, this include:

    a. a greater focus on STEM subjects in school and university, including the ethical issues of computing and its applications; and

    b. a greater emphasis on digital literacy and the use of data across society.

Missing Principle: Information Security

35. The Committee considers that the core principle of "information security" should be included given its importance in generating public trust. Information security, alongside cyber resilience, encompasses the protection of information against unauthorised access and use,[30] and the ability to withstand and recover from cyber-attacks.[31]

36. Despite its intersection with Principle 3 and Principle 4, the Committee considers that this should be a separate principle. Information security risks for all persons, be they public or private, individuals or businesses, continue to grow,[32] necessitating the creation of a dedicated core principle for information security in the AI context.

37. This accords with section 2, ‘Technical robustness and safety’, of the pilot version of the ‘Trustworthy AI Assessment List’ proposed by the EU's Independent High-Level Expert Group on Artificial Intelligence.[33] Section 2 explicitly calls for relevant actors to consider ‘different types and natures of vulnerabilities, such as… cyber-attacks’, and considers ‘whether security or network problems such as cybersecurity hazards could pose safety risks or damage due to unintentional behaviour of the AI system’.[34] Similarly, the National Science and Technology Council warns that ‘AI systems also have

[29] Ibid 6.

[30] *Oxford English Dictionary* (online at 2 May 2019) ‘information security’.

[31] Australian Securities and Investments Commission, *Cyber Resilience: Health Check* (Report No 429, March 2015) 4-5

[32] See, for example, Suzanne Widup et al, *2018 Verizon Data Breach Investigations Report: 11th Edition* (Research Report, April 2018) 4; Europol, *Internet Organised Crime Threat Assessment* (Report, 2018) 7-9.

[33] European Commission Independent High-Level Expert Group on Artificial Intelligence (n 15) 27.

[34] Ibid.

their own cybersecurity needs'.[35] The Pentagon's 2018 AI Strategy includes an increased 'focus on defensive cybersecurity of hardware and software platforms as a precondition for secure uses of AI'.[36]

38. The Committee seeks to draw the Department's attention to three potential sources of information security risk in the AI context:

    a. Disruption of the operation of AI tools by modifying their underlying code in a harmful way;

    b. Breach of the systems used for training an algorithm to interfere with the training data and "fool" or "game" the algorithm through adversarial examples.[37] Adversarial examples are 'inputs to machine learning models that an attacker has intentionally designed to cause the model to make a mistake';[38] and

    c. Creation of stimuli to confuse the algorithm (which has been demonstrated recently in relation to autonomous vehicles).[39]

    Developers should implement information security controls against adversarial examples, such as adversarial training and defensive distillation.

39. The Committee also considers that the public and private sectors should, more generally, recognise the serious information security risks enabled, or otherwise facilitated, by AI before working together to formulate strategy to guard against those risks. The House of Lords Committee raised the issue of AI 'super-charging conventional cyber-attacks, and facilitating an entirely new scale of cyber-attack'.[40] Attackers could use AI to better anticipate and subvert information security controls on a larger scale than previously possible. An example is the use of AI to drive spear-phishing attacks, be it in better selecting targets, tailoring the content of the messages that are sent, or more efficiently sending the messages as part of a wider campaign.[41]

40. The House of Lords Committee cites a poll from the 2017 Black Hat conference showing that 62% of attendees believed there to be 'a high possibility that AI could be used by hackers for offensive purposes'.[42] The Committee agrees with the view of the House of Lords Committee that 'the potential for well-meaning AI research to be used by others to cause harm is significant', such that researchers and developers 'must be alive to the potential ethical implications of their work'.[43]

[35] Committee on Technology, Executive Office of the President, *Preparing for the Future of Artificial Intelligence* (Report, 12 October 2016) 36.

[36] Department of Defense, United States Government, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity* (12 February 2019) 15.

[37] Ian Goodfellow et al, 'Attacking Machine Learning with Adversarial Examples', *OpenAI* (Web Page, 24 February 2017) <https://openai.com/blog/adversarial-example-research/>; Select Committee on Artificial Intelligence (n 21) 98-9.

[38] Ibid.

[39] Goodfellow (n 37), citing Nicolas Papernot et al, *Practical Black-Box Attacks against Machine Learning* (Paper, 19 Mar 2017); Ariel Bogle, 'Hackers Tricked a Tesla, and It's A Sign of Things to Come', *ABC News* (Web Page, 14 April 2019) <https://www.abc.net.au/news/science/2019-04-14/tesla-tencent-study-humans-are-trickable-so-are-computers/10994578>.

[40] Select Committee on Artificial Intelligence (n 21) 98.

[41] Ibid, citing Future of Humanity Institute, Submission No AIC0103, *AI in the UK: Ready, Willing and Able?* (11 October 2017) 3.

[42] The Cylance Team, 'Black Hat Attendees See AI as Double-Edged Sword', *ThreatVector* (Blog Post, 1 August 2017), cited in Select Committee on Artificial Intelligence (n 21) 98.

[43] Select Committee on Artificial Intelligence (n 21) 100.

41. The Committee reiterates the need for an AI ethics framework to stress that any AI-enabled information security controls are designed with threats to AI systems in mind, such as adversarial examples, or other interference with the data used to train and operate AI systems that could thwart their detection and prevention of information security risk. Information security teams must implement 'clear processes and mechanisms … by which AI applications carefully vet and sanitise their respective data supply chains', as well as 'mandatory third-party validation of AI systems' to test their effectiveness, particularly if the AI applications are foundational to the overall information security system.[44] Such validation would be synchronous with Tool No 7, 'Mechanisms for monitoring and improvement', of the Discussion Paper's proposed toolkit.

Missing Principle: Consistency

42. The Committee submits that "consistency" should be a core principle of an ethical framework for AI. AI systems must align, and be consistent, with community expectations, human values, social norms and customs.

43. The Committee recognises the intersection between the proposed core principle of "consistency", and other core principles in the Discussion Paper, including Principle 2: Do no harm, Principle 4: Privacy protection, Principle 5: Fairness, Principle 6: Transparency and explainability, and Principle 8: Accountability. However, the Committee is of the view that these principles do not adequately identify the range of human values that should be taken into consideration when developing and programming AI systems. As such, the Committee considers that the principles put forward in the Discussion Paper could better reflect the values of the Australian public.

44. The Committee stresses the need for AI systems to be aligned with community expectations and human values. These values must be critically considered, evaluated and articulated. If utility functions are not specified, AI systems may produce misguided and unintended results.

45. The Committee acknowledges that defining our values will be difficult. Although human beings hold many common goals, such as happiness, autonomy, security, knowledge, freedom, opportunities and resources),[45] values are subjective in nature and vary across the globe.[46] Even within cultures, there are various competing values, such as privacy and security, or values that could be 'formalized in many different ways mathematically or in computer code', such as the notion of fairness.[47]

46. In addition to the above, the Committee also recognises that human beings 'make decisions based on any number of contextual factors, including their experiences, memories, upbringing, and cultural norms. These factors allow us to have a fundamental understanding of "right and wrong" in a wide range of contexts'.[48] As AI systems do not have these types of experiences to draw upon,[49] it will be

[44] Ibid 100-1, quoting NCC Group, Submission No AIC0240, *AI in the UK: Ready, Willing and Able?* (20 December 2017) 8.
[45] James H Moor, 'Just Consequentialism and Computing' (1999) 1 *Ethics and Information Technology* 65, 66.
[46] Ariel Conn, 'How Do We Align Artificial Intelligence with Human Values', *Future of Life Institute* (3 February 2017) <https://futureoflife.org/2017/02/03/align-artificial-intelligence-with-human-values/>.
[47] Iyad Rahwan, 'Society-in-the-Loop: Programming the Algorithmic Social Contract' (2017) 20(1) *Ethics and Information Technology* 5, 10.
[48] IBM, *Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers & Developers* (September 2018) <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf>.
[49] Ibid.

difficult to encode these human values in a programming language.[50] However, notwithstanding the difficulties in defining what "our" values are, and in training AI systems to learn these values, there are mechanisms that can and should be employed in order to ensure that AI systems are consistent with human values.[51]

47. The Committee notes that one way of embedding these values into AI systems is through computational narrative intelligence. AI systems 'that can read and understand stories can learn the values held by the culture from which the stories originate'.[52] Organisations, however, should ensure that existing value biases perpetuated by dominant groups are not unconsciously programmed into algorithms. In this vein, AI system developers / designers should consider a value-sensitive design approach, which 'asserts that bias in computer systems pre-exists the system itself',[53] requiring the examination of a developer's own practices as '[t]his bias manifests during the operation of the systems due to feedback loops and dissonance between the system and our dynamic social and cultural contexts'.[54]

48. Values can also be embedded in AI systems through inverse reinforcement learning, which teaches human values through observation, feedback and rewards.[55] This is illustrated, for example, when a robot learns to walk. If the robot takes a large step and falls, this fall is considered negative feedback, and the robot adjusts its actions accordingly by taking a smaller step.[56] This example may be overly simplistic, however, in some instances, AI systems will be able to understand and recognise patterns of common biases.[57]

49. In light of the above analysis, the Committee submits that the principle of "consistency" should be a core principle of an ethical framework for AI. The failure of AI systems to be aligned with such values could produce unintended results and create social disruptions, undermining the many benefits that AI systems can and will provide.

Missing Principle: Complementarity

50. The Committee submits that the principle of "complementarity" should be a core principle of the AI ethics framework. Although AI has been defined as a system that works autonomously without explicit guidance from human beings,[58] AI must augment, not replace, 'the perception, cognition, and problem-solving abilities of people'.[59]

[50] Conn (n 46).

[51] Ibid.

[52] Mark O Riedl and Brent Harrison, 'Using Stories to Teach Human Values to Artificial Agents', *ScienceDaily* (12 February 2016) <https://www.sciencedaily.com/releases/2016/02/160212200239.htm>.

[53] Whittaker (n 10) 27.

[54] For an example, see the Amazon Candidate Job Case Study: Whittaker (n 10).

[55] Tucker Davey, 'Cognitive Biases and AI Value Alignment: An Interview with Owain Evans', *Future of Life Institute* (8 October 2018) <https://futureoflife.org/2018/10/08/cognitive-biases-ai-value-alignment-owain-evans/>.

[56] Bernard Marr, 'Artificial Intelligence: What's the Difference between Deep Learning and Reinforcement Learning?', *Forbes* (22 October 2018) <https://www.forbes.com/sites/bernardmarr/2018/10/22/artificial-intelligence-whats-the-difference-between-deep-learning-and-reinforcement-learning/#326c2a43271e>.

[57] Davey (n 55).

[58] Dawson (n 11) 14.

[59] 'About Us: Our Work (Thematic Pillars)', *Partnership on AI* <https://www.partnershiponai.org/about/#pillar-6>.

51. The Committee agrees that 'machines are better than humans at crunching numbers, memorizing, predicting, and executing precise moves'.[60] AI systems can also integrate and analyse unmanageable amounts of data,[61] and unlike human beings, such technology can work continuously day in and day out.[62] However, as discussed below, the Committee submits that AI must be complementary to human beings because, despite these superior qualities, AI has a number of fundamental limitations.

52. Although AI has several superior qualities, the logical, deterministic and analytical nature of AI is incapable of being applied directly to the very 'complex, unpredictable, emergent biological and social systems' that exist in our society.[63] These insights are human in nature, 'not physical or mathematical',[64] and as such, AI systems should not be developed with the aim of replacing human beings altogether but, rather, should be developed with the aim of augmenting our human abilities.

53. In light of the comparative advantages of human beings and AI, the Committee stresses the importance of AI being used in collaboration with, and to complement, human beings.

54. AI systems should 'provide access to real-time information; collect, curate, process and analyse data; and analyse sentiments and represent diverse interpretations'.[65] In this sense, the Committee agrees that new career pathways must be created for those who perform tasks that are being replaced by AI systems (this displacement is discussed further in question 4 below).[66]

55. As human beings and AI have different strengths and capabilities, each should be used collaboratively in order to compensate for the limitations of the other. This idea was highlighted in an empirical study that detected cancer. In this study, AI systems had a 7.5% error rate, and pathologists had a 3.5% error rate.[67] However, when combined, the error rate was drastically reduced to 0.5%.[68] These results suggest that the relationship between human beings and AI should be one of synergy and symbiosis, as the interaction between the two produces a combined effect greater than the sum of their separate parts.

56. Accordingly, the Committee submits that the principle of "complementarity" should be added to the Department's ethical framework for AI. As many insights are human in nature, rather than physical or mathematical, AI systems should not replace human beings altogether but should be used to augment our perception, cognition and problem-solving abilities.

[60] Amit M Joshi and Maude Lavanchy, 'Data Analytics & Artificial Intelligence: What it means for your Business and Society', *IMD* (April 2018) <https://www.imd.org/research-knowledge/articles/artificial-intelligence-real-world-impact-on-business-and-society/>.

[61] Mohammad Hossein Jarrahi, 'Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making' (2018) 61(4) *Business Horizons* 577.

[62] John O McGinnis and Russell G Pearce, 'The Great Disruption: How Machine Intelligence will Transform the Role of Lawyers in the Delivery of Legal Services' (2014) 82(6) *Fordham Law Review* 3041, 3041.

[63] Rick Robinson, '11 Reasons Computers Can't Understand or Solve Our Problems without Human Judgement', *The Urban Technologist* (7 September 2014) <https://theurbantechnologist.com/2014/09/07/11-reasons-computers-cant-understand-or-solve-our-problems-without-human-judgement/>.

[64] Ibid.

[65] Jarrahi (n 61).

[66] Dawson (n 11) 7.

[67] Dayong Wang et al, *Deep Learning Identifying Metastatic Breast Cancer* (18 June 2016) <https://arxiv.org/pdf/1606.05718.pdf>.

[68] Ibid.

Missing Principle: Diversity and Inclusion

57. The Committee considers that "diversity and inclusion" should be added to the principles proposed in the Discussion Paper. Although Principle 1 seeks to 'generate benefits for people that are greater than the costs',[69] and Principles 5 states that '[t]he development or use of the AI system must not result in unfair discrimination against individuals, communities or groups', [70] explicitly promoting the development and use of AI systems to strive for diversity and inclusion is essential to the design of an AI ethics framework:

> In a world marked by inequality, artificial intelligence should not end up reinforcing the problems of exclusion and the concentration of wealth and resources. With regards to AI, a policy of inclusion should thus fulfill a dual objective: ensuring that the development of this technology does not contribute to an increase in social and economic inequality; and using AI to help genuinely reduce these problems. Rather than undermining our individual paths in life and our welfare systems, AI's first priority should be to help promote our fundamental human rights, enhance social relations and reinforce solidarity. Diversity should also figure within these priorities.[71]

58. Principle 5 appears to meet the first of these objectives, but Principle 1 is unclear in the benefits it seeks to provide, as well as for whom the AI system must generate benefits.[72] The value of "inclusion" would complement Principle 5 to fulfil the second objective of using AI to genuinely reduce social and economic inequality by countering exclusion, and enabling every person to participate and make a meaningful contribution.[73] The Microsoft AI principles reflect this by stating that 'AI systems should treat all people fairly', and that 'AI systems should empower everyone and engage people'.[74]

59. Employing both Principle 5 and a principle of "diversity and inclusion" would mirror the aims of substantive equality, as reflected in the special measures, reasonable accommodation and affirmative action provisions in anti-discrimination legislation.[75]

60. Principle 5 addresses the procedure for developing AI systems, whereas diversity and inclusion aims to achieve egalitarian outcomes.[76] The goal should be to achieve these outcomes in both the effect of the AI system, and its environment, as '[p]atterns of cultural discrimination are often embedded in AI systems in complex and meaningful ways'.[77] CognitionX states that:

> Any prejudices and inequalities we have as a society can end up coded into our systems. One of the reliable ways we know can mitigate this risk is to have more diverse development teams in terms of specialisms, identities and experience. Particularly regarding gender, this is a huge challenge; few young women take up technology subjects and careers; just 16% of the graduates in computer studies are women and the figure is

[69] Dawson (n 11) 6.
[70] Ibid.
[71] Cédric Villani, *For a Meaningful Artificial Intelligence: Towards a French and European Strategy* (Report, 8 March 2018) 7.
[72] Dawson (n 11) 6.
[73] Villani (n 71) 7; Gilian Triggs, 'Social Inclusion and Human Rights in Australia' (Speech, Chain Reaction Foundation Breakfast Café, Sydney, 20 August 2013).
[74] 'Microsoft AI principles', *Microsoft* (Web Page, 2019) <https://www.microsoft.com/en-us/ai/our-approach-to-ai>.
[75] See, for example, *Equal Opportunity Act 2010* (Vic) s 12(1); *Disability Discrimination Act 1992* (Cth) s 5(2)(a); *Affirmative Action (Equal Opportunity for Women) Act 1986* (Cth).
[76] Hugh Collins, 'Discrimination, Equality and Social Inclusion' (2003) 66 *Modern Law Review* 16, 17.
[77] Whittaker (n 10) 39.

14% for engineering and technology. Nearly all of the 200-plus senior women in tech who responded to a recent survey had experienced sexist interactions.[78]

61. Actions that further diversity and inclusion outside of AI design and development should be encouraged to target inequalities that may become embedded as bias in the AI system. An example of diversity and inclusion in practice are the ‘more tangible steps to promote inclusion and diversity’ taken by NeurIPS to their conference in response to feedback.[79]

62. Similar actions would be encouraged and promoted in the AI context through the addition of "diversity and inclusion" as a core principle in the AI ethics framework.

[78] CognitionX, 'CognitionX - Written evidence (AIC0022)', *House of Lords Select Committee on Artificial Intelligence* (5 September 2017), [8.7].

[79] Neural Information Processing Systems, ‘NIPS Name Change’, *News Releases* (Web Page, October 2017) <https://nips.cc/Conferences/2018/News?article=2110>.

Question 4: Would the proposed tools enable you or your organisation to implement the core principles for ethical AI?

63. In responding to this question, the Committee has focused on:

    a. **Ethical Toolkit Item 8**: In addressing this question the Committee has looked at recourse mechanisms from two perspectives:

        i. <u>Recourse mechanisms and civil liability:</u> If there is an issue with AI / the use of AI – who is ultimately accountable / who do the aggrieved approach for remedies.

        ii. <u>Recourse mechanisms and broad industry outcomes:</u> Recourse mechanisms for workers who have been displaced by developments in AI.

    b. **Ethical Toolkit Items 2, 3 and 9**: In addressing these questions, the Committee has looked at ways to help ensure that the use of any AI systems adhere to ethical principles, Australian policies and legislation and strike a balance between managing risk and encouraging innovation. The Committee also addresses the need for consultation and for any approach to be consistent with human rights principles.

a. Ethical Toolkit Item 8

i. Recourse mechanisms and civil liability

64. The Committee submits that questions of ethics should not be considered without contemplating and determining key questions of law, and, in particular, where rights and liabilities fall in the AI space.

65. The Committee considers there to be two broad types of AI:

    a. Supervised Algorithms: Where the AI learns to predict the target through the inputs, which are pre-determined by a human, and through the correction of the outputs, which are also completed by a human. The learning stops when a human decides that the algorithm has identified the target a sufficient number of times and, hence, the level of the AI algorithm's performance is deemed acceptable.[80]

    b. Unsupervised Algorithms: Where AI learns to predict the target through the inputs, and the model, which are pre-determined by a human. However, a human has no ability to supervise how the algorithms learn from the model. The algorithms discover and present the structure in the data by relying on their own observations.[81]

66. In understanding and assessing the liability regimes applicable to an AI product, we consider that the extent of the interaction between the human and the AI algorithm will be a fundamental consideration.

AI and personhood

67. In order to impose any liability there must be a wrongdoer, whether it is a natural person or an entity that is legally considered to be a person, and deemed to have personhood. Philosopher Charles Taylor

[80] Jason Brownlee, 'Supervised and Unsupervised Machine Learning Algorithms', *Machine Learning Mastery* (Web Page, 16 March 2016) <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>.

[81] Ibid.

noted that, from the "naturalist" epistemological tradition, the term “personhood” may refer to any human (or non-human) agent who: (1) possesses continuous consciousness over time; and (2) is therefore capable of framing representations about the world, formulating plans and acting on them.[82]

68. In general, currently, it is not controversial that AI algorithms do not possess a level of consciousness or ability to frame representations about the world the same as human beings. As a result, current AI algorithms, however sophisticated, cannot be seen as having “natural” personhood and, therefore, cannot be liable directly for their decisions or actions.

69. Current case law does not address the issue of liability for the actions of AI products. However in the administrative law space in the case of *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79, the majority of the Full Court dismissed the appeal, agreeing with the decision of the Primary Judge that a computer-generated letter was not evidence of a "decision" of an authorised decision-maker at the ATO. This is because the Court concluded that both limbs of the test needed to be fulfilled; so while there was an “objective manifestation of the conclusion” (the letter), there was no 'mental process of reaching a conclusion'.[83] Accordingly, in this case, the Court held that no decision was made and the Deputy Commissioner was not bound to what was conveyed by the computer-generated letter. The Court noted the potential for this to cause administrative uncertainty[84]. Special leave to the High Court was refused.

70. The implications of this case are yet to be fully recognised. It does raise questions and considerations that will need to be addressed in the future. For example, will an algorithm be able to make binding decisions? Will shareholders and humans be considered as “shareholders” of algorithms? Or developers of algorithms as their directors? With the emergence of technology, these questions of law may need to be more thoroughly examined in the future.

Other Considerations

71. Agency: In determining potential questions of liability, it will be important to consider the applicability of the agency doctrine to AI algorithms. The attempt to attribute the agency relationship to AI algorithms and their owners, users, or designers reveals a multitude of issues, including:

    a. If the AI algorithm is an agent, can liability be attributed to the principal for AI’s conduct (and does this change depending on the supervised or unsupervised nature of the AI product)?

    b. Can the principal be found negligent in creating, selecting, controlling or supervising the AI algorithm (and does this change depending on the supervised or unsupervised nature of the AI product)?

    c. To what extent can the principal communicate the authority to the AI algorithm and be sure that the authority is understood (especially, in the case of the unsupervised AI algorithms)?

    d. Can a principal be held vicariously liable for the actions of the AI algorithm?

[82] Charles Taylor, 'The Concept of a Person' in *Volume 1: Human Agency and Language: Philosophical Papers* (Cambridge University Press, 1985) 97 – 114.

[83] *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79 [140].

[84] Ibid [152].

72. Foreseeability: In circumstances where humans program an unsupervised AI algorithm, which is designed to act in an unforeseeable way, could or should liability be attributed to the human designer if there is an issue? Theoretically, there is no causation between the breach of a duty of care by a human, and the harm caused by that breach. However, consumer expectations dictate that if there is an issue, there should be someone held accountable and from whom to seek a remedy. This case would be different if the initial inputs or the model were "faulty", or identified as a "safety defect" – in both these cases the liability and remedy is governed under the Australian Consumer Law.[85] This is a complex and novel category of duty of care. Further consideration may need to be given as to whether a degree of oversight is required (and the subsequent liability implications of this).

73. Striking a balance: While it will be important to have clear regimes in place to hold parties accountable in the unfortunate circumstance of an issue, it will also be important to have a regime that continues to promote innovation – otherwise, this could cause an "AI winter",[86] meaning it may reduce interest from developers and investors in unsupervised algorithms, which will impede important technical advancements, and potentially reduce Australia's ability to compete on a world stage.

ii. Recourse mechanisms and broad industry outcomes

74. The Committee considers that the "recourse mechanisms" should consider individualised justice for those who may be negatively impacted by the use of AI-enabled processes in their industry. This should include engagement with specific individuals who will lose their jobs by being "replaced" with AI-enabled processes. This engagement could include creating a system of compensation by way of finance, or the provision of upskilling.

75. Scope of Potential displacement of workers: The Committee makes this recommendation in light of the way the broader economy is likely to be transformed by AI technologies. The encroachment of AI into particular industries has the ability to trigger a 'move from declining occupations to growing and, in some cases, new occupations'.[87] The McKinsey Global Institute (**McKinsey**) suggests that up to 14% of the global workforce will have to switch occupations,[88] and found that roughly half of the activities across 800 occupations could be automated.[89] This will disproportionately affect 'physical activities in predictable, structured environments, [and] data collection and processing'.[90] The same research found that up to 30% of the global workforce could be 'displaced by automation in the period 2016-30'.[91]

76. Upskilling: Any proposed recourse mechanism should work harmoniously with government strategies to address dislocation and, ideally create a mechanism to mitigate the effects of, 'significant workplace transitions'.[92] Such programs (government or industry funded) might include provision for helping

[85] *Competition and Consumer Act 2010* (Cth) sch 2.
[86] Neil Mehta, and Murthy V Devarakonda, 'Machine Learning, Natural Language Programming, and Electronic Health Records: The Next Step in the Artificial Intelligence Journey?' (2018) 141 (6) *The Journal of Allergy and Clinical Immunology* 2019, 2019.
[87] McKinsey Global Institute, *AI, Automation, and the Future of Work: Ten Things to Solve for* (Briefing Note, June 2018) 1.
[88] Ibid.
[89] Ibid 2.
[90] Ibid.
[91] Ibid 3.
[92] Ibid 1.

workers acquire new skills, such as programming, and adapt to work alongside machines in certain occupations.[93]

77. Compensation considerations: A more recognisable approach to compensation or welfare would rely on payment or income stream to restore individuals and their families to their economic position prior to industrial disruption by AI. Long term, however, there is an imperative for industries and governments to acknowledge reskilling in future policy design. The World Economic Forum states that 'by 2022, no less than 54% of all employees will require significant re-skilling and upskilling. Of these, about 35% are expected to require additional training of up to six months, 9% will require reskilling lasting six to 12 months, while 10% will require additional skills training of more than a year'.[94]

78. The Committee suggests that the Department, in consultation with private and industry bodies, and the Productivity Commission, identify areas of employment and industry sectors in which human personnel are likely to be replaced with AI-enabled processes. A report such as this would allow the Department, and the affected industries, to formulate strategies for re-skilling and/or potential compensation avenues (including engaging in presumably contentious debates on how much financial burden (if any) should be borne by employers, and what role the government could / should play). The Committee also acknowledges that a balance needs to be struck here to ensure businesses are not disincentivised to adopt new technologies that may help efficiencies.

79. Other considerations for the Department: The Committee recognises the Australian Government's experience with structural changes in the labour market due to technological innovation. As part of this recourse mechanism, apart from industrial changes, government sectors may need to plan for the real potential that 'occupational mix shifts', encouraged by AI and automation, will exacerbate income inequality. It is predicted that there will be significant demand for 'high-skill medical and tech or other professionals', which are typically high-wage jobs, and a decline in demand for human personnel in automatable work.[95] However, whilst McKinsey expects the demand for many jobs will increase, including for teachers and nursing aides, those jobs typically have lower wage structures.[96] This creates a risk 'that automation could exacerbate wage polarization, income inequality, and the lack of income advancement … stoking social, and political tensions'.[97]

80. The Committee recommends flexibility in any policy approach regarding potentially compensating affected individuals under the recourse mechanisms of the future, given the 'large uncertainties about the likely new technologies and their precise relationship to tasks', which makes it 'difficult to make precise predictions as to which jobs will see a fall in demand and the scale of new job creation'.[98]

b. Ethical Toolkits Items 2, 3 and 9: Considerations underlying AI regulation and risk management

81. In order to ensure that the use of AI systems adheres to ethical principles and Australian policies and legislation, as well as to help classify and manage risk, the Committee submits that the Department

[93] Ibid.

[94] Till Alexander Leopold, Vesselina Ratcheva and Saadia Zahidi, *The Future of Jobs Report: 2018* (World Economic Forum, Insight Report, 20 July 2018) ix.

[95] McKinsey Global Institute (n 87) 3.

[96] Ibid.

[97] Ibid.

[98] British Academy for the Humanities and Social Sciences and the Royal Society, *The Impact of Artificial Intelligence on Work* (Report, 11 September 2018) 4.

should consider the creation of a regulatory and government agency (the **AI Regulator**) as a central advisory body with a high level of expertise in AI-related technologies. If appropriate, the Department may consider increasing the mandate of an existing agency if appropriate.

82. At a minimum, the AI Regulator should consider:

    a. analysis and recommendations regarding desirable legislation for AI-related technologies;

    b. assistance to other Australian government bodies and organisations in compliance with ethical and legal policies and regulations; and

    c. risk assessment of the AI products.

83. Consultation with relevant parties and key stakeholders in developing these policies and procedures is also paramount.

Considerations for identification of risk

84. The Committee proposes that the assessment of the potential risks posed by AI-related technology could be undertaken in two stages: self-assessment, and then authorisation from the AI Regulator. This approach is based on the authorisation process used by the ACCC in assessing applications for merger and non-merger authorisation under the *Competition and Consumer Act 2010* (Cth).

85. Self-Assessment: The developer of the AI-product would need to identify the potential risks posed by their AI-product and assess the seriousness of those risks against legal and ethical criteria, including the algorithm's accuracy, unwanted algorithmic bias, algorithmic fairness and discriminatory attitudes. The developer would also need to assess whether the AI-product could pose a risk (i.e. consideration of 'unknown unknowns').

86. The regulator would need to determine whether this self-assessment could be conducted with the aid of computer programs specialising in reviewing and assessing AI-products, such as IBM's 'AI Fairness 360'. Consideration may be given to creating a 'white list' of computer programs vetted by the regulator that could assist with this task. This would also assist in providing consistency across the board.

87. Seeking Authorisation from the Regulator: If the potential risks were sufficiently high (according to a threshold developed by the AI Regulator), or identified as unknown unknowns, the developer would need to seek *authorisation* for their AI-product. The AI Regulator would assess the potential risks, benefits and detriments to the public, and whether the benefits would outweigh the risks. As noted below, it would be beneficial if the other stakeholders, and the public in general, were involved in the review process through public consultation (where appropriate).

88. The Department may consider that the authorisation process should be open and transparent. For example, through maintaining a public registry of all authorisation applications. This should be subject to exceptions, for example, the maintenance of privileged and confidential material, and maintaining the confidentiality of potential patent applications or trade secrets.

89. Striking the right balance: The degree of regulation will need to be carefully considered in order to manage risk, and not unduly stifle creativity and innovation.

Developing consultation: aligning the consultation process with human rights

90. The Committee considers that a broad approach to consultation would be more closely aligned with Principle 5 of the Discussion Paper by recognising the interests of both direct and indirect stakeholders, as well as establishing accordance with human rights obligations.

91. The Committee proposes that the AHRC should be empowered to provide public and community consultation functions to enable a broad range of key stakeholders – determined by the AHRC – to participate in a public consultation process in relation to human rights and AI technology. In relation to their functions under a Human Rights Act, the AHRC states:

> The Commission's current statutory functions include promoting understanding, acceptance and public discussion of human rights in Australia. The Commission has substantial expertise and experience in this area and is ready to play a leading role in engaging the Australian community on the content and effect of a Human Rights Act.[99]

92. In relation to consultation, the AHRC proposes 'holding public forums'.[100] The inclusion of affected individuals and communities on a large scale similarly accords with the rationale behind the *Convention on the Rights of Persons with Disabilities*, which seeks to ensure 'participation of all key stakeholders'. [101] Thus, a consultation process should possess the means to identify all key stakeholders by employing a broader approach that integrates a public consultation process.

93. However, a significant concern in public consultation is the identification of key stakeholders. AI Now states that:

> Approaches to fairness and bias must take into account both allocative and representational harms, and those that debate the definitions of fairness and bias must recognize and give voice to the individuals and communities most affected. Any formulation of fairness that excludes impacted populations and the institutional context in which a system is deployed is too limited.*102*

94. Although identifying the 'individuals and communities most affected' may be achieved with a public forum process, Oxford University's Future of Humanity Institute proposes a method for identifying, and consulting, key affected individuals and communities by using 'scenario based surveys, and studying particular groups who have been exposed to instances of phenomena (such as employment shocks, or new forms of surveillance) that could later affect larger populations'.[103]

95. The Committee considers that empowering the AHRC to conduct public consultation, such as by holding public forums, and identifying and studying groups affected by AI systems, would provide a base for integrating key stakeholders in the application of AI systems who may not otherwise have been identified.

[99] Australian Human Rights Commission, Submission to National Human Rights Consultation, *Australian Human Rights Commission Submission* (Submission, June 2009) 74 (citations omitted).

[100] Ibid.

[101] Faraaz Mahomad, Michael Ashley Stein and Vikram Patel, 'Involuntary Mental Health Treatment in the Era of the United Nations Convention on the Rights of Persons with Disabilities' (2018) 15(10) *PLoS Med* 1, 1.

[102] Whittaker (n 10) 28.

[103] Allan Dafoe, *AI Governance: A Research Agenda* (Future of Humanity Institute, University of Oxford, 27 August 2018) 39.

Question 7: Are there additional ethical issues related to AI that have not been raised in the discussion paper? What are they and why are they important?

96. The Committee considers that there are a further two ethical issues related to AI that have not been raised in the Discussion Paper:

    a. Police use of AI in "predictive policing"; and

    b. Issues relating to copyright law.

    The significance of these issues is outlined below.

a. Police use of AI in "predictive policing"

Policing, policing organisations and public consent

97. Policing is the 'attempt to maintain security through surveillance and the threat of sanctioning'.[104] It is 'arguably a necessity in any social order', but is conducted 'by a number of different processes and institutional arrangements'.[105] In that regard, the Committee defines policing organisations as 'the police', that is, the 'specialized body of people given the primary formal responsibility for legitimate force'.[106] This demarcation is necessary, given the multifaceted, dynamic nature of policing, with Reiner considering that 'a state-organised specialist "police" organization of the modern kind is only one example of policing'.[107]

98. The police are unique – they are 'specialist repositories for the state's monopolization of legitimate force' in their territory.[108] The police, as a core component of the criminal justice system, have a highly complex function, given the different types of actors therein and the multifaceted nature of the environment in which they operate.[109] The police must take actions that can have serious implications for victims of crime, offenders, and third parties, and must ascertain how much data they need to 'reduce uncertainty about the crime environment' before taking those actions.[110] The level of discretion afforded to the police in making these decisions adds to the complexity of their role.[111]

Police use of technology

99. Just as policing organisations 'can have a variety of shifting forms', policing itself can be conducted through a variety of 'different processes and institutional arrangements'.[112] These processes and arrangements include the technologies employed by policing organisations to execute their

[104] Robert Reiner, *The Politics of the Police* (Oxford University Press, 3rd ed, 2000) 3, citing Steven Spitzer, 'Security and Control in Capitalist Societies: The Fetishism of Security and the Secret Thereof' in John Lowman, Robert J. Menzies and T. S. Palys (eds), *Transcarceration: Essays in the Theory of Social Control* (Gower, 1987); at 3, citing Clifford D. Shearing, 'The Relationship between Public and Private Policing' in Michael Tonry and Norval Morris (eds), *Modern Policing* (Chicago University Press, 1992).
[105] Reiner (n 104) 1-2.
[106] Ibid 6-7.
[107] Ibid 1-2.
[108] Ibid 6.
[109] Manuel A. Utset, 'Digital Surveillance and Preventive Policing' (2017) 49 *Connecticut Law Review* 1455, 1464.
[110] Ibid 1464, 1466.
[111] Ibid 1466.
[112] Reiner (n 104) 1-2.

functions. Modern policing has become increasingly data-rich. There is a greater volume of human activity occurring in cyberspace (criminal and non-criminal), and the 'expansion of tracking and sensing technologies (including natural language processing and image recognition) is exponentially increasing the volume and accessibility of information on human behaviour'.[113] Joh uses automatic licence plate readers (systems using 'cameras mounted on patrol cars or at fixed locations and data analytics to identify' car licence plates) as an example of the data-rich nature of modern policing.[114] These systems 'can read up to fifty license plates per second, and typically record the date, time, and GPS location of every scanned plate'.[115]

100. Police use of AI is enhanced by their growing data collection capabilities, which has encouraged the creation of new technologies for police to properly leverage the 'data abundance' of their work.[116] Police use of big data technologies has led to the development of 'a variety of tools and methods for acquiring, storing, and processing large data sets to extract useful knowledge'.[117] When applied to big datasets of police information, AI, through enhanced analytics capabilities, enables the police to proactively detect and monitor emerging threats, and thus devise risk-based strategies to combat them.[118] Analytics such as these are 'promoted as the greatest potential benefit of using big data technologies by enabling 'improved predictive analysis',[119] in addition to automating 'routine processing and the generation of insight on a vast range of policing problems'.[120] Therefore, policing has been considered an ideal use case for AI: it is 'an information-based activity' and its effectiveness is dependent on 'large quantities of information, or data, on human behavior, collected from a variety of sources'.[121] The predicted dependence by the police on AI likely stems from the fact that, 'in many criminal cases, there is already simply too much data for the traditional officers to capture and assess'.[122] Ferguson explicitly states that 'the future of policing will be driven by data'.[123]

[113] Deloitte, *Policing 4.0: Deciding the Future of Policing in the UK* (Report, 18 September 2018) 10.

[114] Elizabeth E. Joh, 'The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing' (2016) 10 *Harvard Law & Policy Review* 15, 22, citing New York State Division of Criminal Justice Services, New York State, *Suggested Guidelines: Operation of License Plate Reader Technology* (2011) 11.

[115] Ibid, citing New York State Division of Criminal Justice Services (n 114) 7.

[116] Andrew Guthrie Ferguson, 'Illuminating Black Data Policing' (2018) 15 *Ohio State Journal of Criminal Law* 503, 507-8; Lydia Bennett Moses and Janet Chan, 'Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools' (2014) 37(2) *University of New South Wales Journal* 643, 658-9, 663; Deloitte (n 113) 10.

[117] Utset (n 109) 1455.

[118] Joh (n 114) 16-17; Walter L. Perry et al, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (RAND Corporation, Report, 14 October 2013) xiv; International Criminal Police Organisation and United Nations Interregional Crime and Justice Research, *Artificial Intelligence and Robotics for Law Enforcement* (Report, 21 March 2019) 3; Elizabeth E. Joh, 'Policing by Numbers: Big Data and the Fourth Amendment' (2014) 89 *Washington Law Review* 35, 48.

[119] Alana Maurushat, 'BD Use by Law Enforcement and Intelligence in the National Security Space: Perceived Benefits, Risks And Challenges' (2016) 21 *Media and Arts Law Review* 229, 236, citing Stephen L Morgan and Christopher Winship, *Counterfactuals and Causal Inference: Analytical Methods for Social Research* (Cambridge University Press, 2007).

[120] Deloitte (n 113) 37.

[121] International Criminal Police Organisation and United Nations Interregional Crime and Justice Research (n 118) 3.

[122] Ibid.

[123] Ferguson (n 116) 503.

Predictive policing

101. Predictive policing cements policing organisations' reliance on more powerful information technology systems and analytics capabilities.[124] Considering its reliance on seemingly objective police data, predictive policing can reinforce the credibility of police institutions in the eyes of the public.[125]

102. Perry et. al. define predictive policing as 'the application of analytical techniques — particularly quantitative techniques — to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions'.[126] There are different types of predictive policing, largely clustering around forecasting: the location of a crime, such as hot spot analysis, risk terrain analysis and statistical regression; and the time of a crime, namely temporal and spatiotemporal methods.[127] While the police have used statistical and geospatial analysis for intelligence-led policing for decades (such as in the New York Police Department's CompStat system)[128] to forecast crime levels, the distinguishing characteristic of predictive policing is the use of AI to analyse large police datasets, and thus predict the likely occurrence of crime.[129] Since prediction has always been inherent to the work of the police, predictive policing represents more of 'a shift in tools than strategy'.[130]

103. There are many (potential) use cases, such as predicting potential offenders based on criminal histories, and identifying groups that are likely to become victims of crime.[131] Several trials in the UK and United States have used predictive geospatial tools to, in most cases, forecast the location of crime more effectively versus incumbent methods.[132] The Los Angeles Police Department has deployed predictive analytics to forecast gang violence and augment the LAPD's monitoring of crime in real time.[133] The Santa Cruz Police Department followed in 2011 by deploying a computer algorithm, after analysing car and home burglaries over an eight year period, to predict the time and location of crime, and to inform officers of certain locations warranting investigation.[134]

[124] Perry et al (n 118) 2.

[125] Andrew Guthrie Ferguson, 'Policing Predictive Policing' (2017) 94 *Washington University Law Review* 1109, 1114.

[126] Perry et al (n 118) 1-2.

[127] Ibid 17, 19.

[128] Joh (n 114) 43.

[129] Perry et al (n 118) 2; Joh (n 114) 44.

[130] Ferguson (n 116) 1123.

[131] Perry et al (n 118) xvi-xvii.

[132] Alexander Babuta, Marion Oswald and Christine Rinik, *Machine Learning Algorithms and Police Decision-Making Legal, Ethical and Regulatory Challenges* (Royal United Services Institute, Whitehall Report No 3-18, September 2018) 3, citing Shane D Johnson et al., *Prospective Crime Mapping in Operational Context: Final Report* (Home Office, Home Office Online Report No 19/07, 2007); at 3, citing Beth Pearsall, 'Predictive Policing: The Future of Law Enforcement?' [2010] (266) *National Institute of Justice Journal* 16; at 3, citing Jennifer Bachner, *Predictive Policing: Preventing Crime with Data and Analytics* (IBM Center for The Business of Government, Report, 2013); at 3, citing Perry et al (n 118).

[133] Perry et al (n 118) 4.

[134] Ferguson (n 116) 1112, citing Stephen Baxter and Santa Cruz Sentinel, 'Modest Gains in First Six Months of Santa Cruz's Predictive Police Program', *Santa Cruz Sentinel* (News article, 26 February 2012) <https://www.santacruzsentinel.com/2012/02/26/modest-gains-in-first-six-months-of-santa-cruzs-predictive-police-program/>.

Concerns about use of AI as a tool of predictive policing

104. Despite the benefits discussed above, the literature discloses a number of concerns about the use of AI for predictive policing, suggesting potential implications for the police's retention of public consent.

105. Firstly, there are concerns about predictive policing encouraging mass surveillance, which is seen generally 'in the literature as acts of domination'.[135] In particular, police usage of big data has raised concern.[136] In turn, the connotation of an amorphous AI algorithm grading citizens' behaviour, based on large amounts of data collected through extensive surveillance in a clandestine fashion, creates the potential for public discontent in relation to the use of AI for predictive policing. The greater surveillance discretion (particularly 'by allowing the identification of large numbers of suspicious activities and people by sifting through large quantities of digitized data') afforded to police accentuates that potential.[137] The level of concern about this police discretion is so great that San Francisco became the first American city to ban use of facial recognition technologies by its local policing organisations, and requires those organisations to disclose inventories of surveillance technology.[138]

106. Secondly, the use of AI in predictive policing may exacerbate reported shortcomings.[139] For example, there are concerns that AI could exacerbate issues of racial bias in policing. This is because the development and implementation of AI technologies 'can have discriminatory impacts'.[140] The use by predictive policing systems of data generated though biased police practice can perpetuate that bias in targeting the same, potentially disadvantaged, communities. There are dangers associated with 'overreliance on unaccountable and potentially biased data to address sensitive issues like public safety', not least because police data is an incomplete repository of relevant crime information.[141] Moreover, the issue of low transparency associated with the deployment of AI as a predictive policing tool needs to be addressed.[142] As a technological concept, AI has the ability to further the lack of transparency in policing due to its complexity and perception as (potentially) a "black box",[143] which is accentuated by frequent changes to the algorithms.[144]

107. The Committee considers that the factors outlined above may contribute to the erosion of public trust in the police. That erosion is particularly likely if the police overlook direct engagement with citizens

[135] Maurushat (n 119) 250-1.
[136] Ibid 250, citing Robert Chalmers, 'Orwell or All Well? The Rise of Surveillance Culture' (2005) 30(6) *Alternative Law Journal* 258.
[137] Joh (n 114) 19.
[138] Kari Paul, 'San Francisco Is First US City to Ban Police Use of Facial Recognition Tech', *The Guardian* (online at 15 May 2019) <https://www.theguardian.com/us-news/2019/may/14/san-francisco-facial-recognition-police-ban>.
[139] Ferguson (n 116) 524.
[140] Ibid 517.
[141] Rashida Richardson, Jason M. Schultz and Kate Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice' (2019) 94 *New York University Law Review* 192, 224, 225-6; Elizabeth E. Joh, 'Artificial Intelligence and Policing: First Questions' (2018) 41 *Seattle University Law Review* 1139, 1142.
[142] Joh (n 114) 38; Ferguson (n 116) 524.
[143] Ferguson (n 116) 524; Wolfgang Schulz et al, *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications* (Council of Europe, Study No DGI(2017) 12, March 2018) 38.
[144] Schulz et al (n 143) 38.

in favour of greater use of analytics,[145] deepening the gulf between law enforcement agencies and the public.[146]

108. The Department needs to consider key questions, including how police deployment of AI can champion fairness, accountability, transparency and explainability.[147] The stakes are high given that inaccurate algorithm predictions used by police could cause ‘wrongful stops, arrests, and unjustified force’, thereby undermining public trust in the police and police legitimacy as an institution.[148]

b. Copyright

109. The reliance of the proposed framework on existing law means that, in the context of copyright law, the AI ethics framework may have to negotiate the divide between industry and legal understanding of AI technology. Furthermore, the Discussion Paper fails to address the ethical implications of AI in the copyright context, when copyright protection and copyright infringement have significant legal and moral dimensions. There are several questions to be raised about the ability of copyright law to police AI data inputs, or the creative outputs of AI systems, as well as the need for recognition of how the law understands existing computer programs.

Does copyright law protect AI programs?

110. In *Data Access Corporation v Powerflex Services Pty Ltd*, the majority of the High Court stated that 'it is impossible to overemphasise the importance of the fact that a computer has no "intelligence" to execute instructions over and beyond the simple logical functions which are hard wired into its circuits'.[149]

111. Copyright law approaches the protection of coding or, more specifically, the persons who originate code, by considering code as a ‘literary work’.[150] Case law on the infringement of copyright in code considers that a breach will be determined by how "substantial" a part (meaning how large a portion considering the originality of the content, and the essentiality to the program)[151] from the entirety of the work is identically reproduced.[152] Individual functions within the code could also be works capable of being infringed. For example, in *Data Access Corporation v Powerflex Services Pty Ltd*, a compression table was found to be a unique work.[153] However, smaller parts of the work, such as macros, were not considered to be separate works.[154]

112. This is because code is understood in copyright law to be a “set of directions”, which is intended 'directly, or after conversion to another language, code or notation … to cause a [computer] to

[145] Maurushat (n 119) 251, citing Judy Putt, *Community Policing in Australia* (Australian Institute of Criminology, Research and Public Policy Series No 111, 2010); at 251, citing Sean Richmond, ‘National Security Debate Misses Big Picture of “Balanced” Response’, *The Conversation* (Opinion Article, 25 February 2015) <https://theconversation.com/national-security-debate-misses-big-picture-of-balanced-response-37923>.
[146] International Criminal Police Organisation and United Nations Interregional Crime and Justice Research (n 118) 12.
[147] Ibid 13.
[148] Joh (n 118) 1143.
[149] *Data Access Corporation v Powerflex Services Pty Ltd* (1999) 166 ALR 228, 238 [36].
[150] *Copyright Act 1968* (Cth) s 47AB.
[151] *Data Access Corporation v Powerflex Services Pty Ltd* (1999) 166 ALR 228, 249 [84].
[152] Ibid 254 [110].
[153] Ibid 255 [123].
[154] Ibid 252 [100].

perform a particular function'.[155] In this way, there is no protection of expression in the same way as for other media: copyright will only protect against the direct copying of code because code has been defined by its functionality, and copyright does not recognise functionality.[156]

113. The existing capacity of copyright law to protect the original creation of code is an open question. Noting that there is, in some cases, a strong component of user input, and learning on the part of the AI program, the extent to which the underlying code that makes up the program is materially protected would seem to be debatable for some AI projects.

114. By considering code as a literary text,[157] there are in-built limitations stemming from a lack of technological neutrality. This characterisation shapes what can be considered original, what constitutes breach and, therefore, what is protected by copyright law. The Committee acknowledges that there are limits to the scope of rights that copyright law protects in every medium of creation, and that copyright law is by no means the only source of legal redress, however, the lack of technological neutrality inherent to the rules of copyright with respect to the characterisation of new technology in an old media framework should be a significant consideration when adapting the AI ethics framework to copyright law. For example, the Full Court of the Australian Federal Court in *National Rugby League v Singtel Optus*[158] discussed how the principle of technological neutrality could be applied to section 111 of the *Copyright Act 1968* (Cth) with respect to cloud recording devices. It was noted that the courts are unable to 'construct [their] own idea of desirable policy' without some existing indication from the legislation.[159] As such, despite a specific provision allowing for the recording of broadcasts for personal viewing at a later, more convenient time, it was found that the cloud model was not understood in the construction of the provision and was not protected by the section. Any possible gaps in construction, which as a result of insufficient technological neutrality will not be overcome in the courts by anything short of legislative change. For the purpose of this analysis, the Committee has not considered whether the code could be considered a "books of the company" under the Corporations Act.

Authorship of AI generated works

115. The Committee notes the complex questions surrounding the ownership and authorship of works generated by AI systems. Prior to the advent of AI technology, computers were seen as tools, like a paint brush and canvas that allowed artists to create artistic works. These works of art were protected by copyright law, just as a painting would be, because they met the definition of originality, which generally requires a human author.[160] However, AI has changed the role of computers in the artistic process from tools to creators. AI programs can make decisions in the creative process without human involvement, generating works of art. Whilst this art may be based on the examples input, and parameters set, by programmers, it is the computer program itself that generates the artwork via

[155] Rod Evenden, 'Copyright Protection of Computer Programs in Australia', (2001) 43 *Computers and Law* 24, 25; *Data Access Corporation v Powerflex Services Pty Ltd* (1999) 166 ALR 228, 236-7 [25].
[156] Ibid.
[157] *Copyright Act 1968* (Cth) s 47AB.
[158] *National Rugby League Investments Pty Ltd v Singtel Optus Pty Ltd* [2012] FCAFC 59.
[159] *National Rugby League Investments Pty Ltd v Singtel Optus Pty Ltd* [2012] FCAFC 59 [97].
[160] Andres Guadamuz, 'Artificial intelligence and copyright', [2017] (October) *WIPO Magazine* 5 <https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html>.

a "neural network". AI is, and has been, used to generate news articles, poems, paintings, music, games, books and musicals. If authored by a human, these artworks would all be protected by copyright law. However, as the author is a computer program, there is no such copyright protection afforded, meaning the artworks could be freely used by anyone.

116. The purpose of copyright protection is to incentivise the creation of new and innovative works of art:

> The exclusive economic rights granted to copyright owners promote creativity and innovation. Copyright enables creators to profit from their work. It protects creators from 'free-riding' or unauthorised exploitation by others which would undermine the incentive to create and invest in new works wanted by the public.[161]

117. Without copyright protection for AI generated works, the people trying to make money from those works (who invested the money in creating the works and designed the AI program) would be faced with the prospect that infinite copies of the work could be used without payment or attribution. The Committee is of the opinion that this state of affairs does not accurately balance the competing interests at the heart of copyright law: whilst the public interest in the free use of artistic works is advanced, Guadamuz notes that this 'may well have a chilling effect on investment in automated systems'.[162]

118. Two options have been widely advanced to deal with copyright in AI generated works: 1) to deny copyright protection altogether; or 2) to attribute authorship of the works to the developer/s of the AI program.

119. In Australia, it appears as if the first option presently prevails. In *Acohs Pty Ltd v Ucorp Pty Ltd*, a work generated with computer intervention was not copyrightable as it was not produced by a human.[163] Similar decisions have been made around the world.[164]

120. However, in many jurisdictions, such as New Zealand and the UK, option two has been applied so that authorship is ascribed to the programmer.[165] In the UK, section 9(3) of the *Copyright, Designs and Patents Act 1988* gives authorship of a computer-generated artistic work to 'the person by whom the arrangements necessary for the creation of the work are undertaken'.[166] A "computer-generated work" is defined as a work 'generated by [a] computer in circumstances such that there is no human author of the work'.[167]

121. The Committee considers that the UK approach (option two) is to be preferred as it recognises the work and financial investment required to create an AI program that is sophisticated enough to generate artistic works.

122. The flow on question from this approach is whether the programmer or user of the program is the "person making the arrangements" for the work to be generated. The Committee suggests that this

[161] Explanatory Memorandum, Copyright Amendment Bill 2006 (Cth) 5.
[162] Guadamuz (n 160) 5.
[163] *Acohs Pty Ltd v Ucorp Pty Ltd* (2012) 287 ALR 403, 413 [57].
[164] See, for example, *Feist Publications v Rural Telephone Service Company, Inc.* 499 U.S. 340 (1991) in the USA and *C-5/08 Infopaq International A/S v Danske Dagbaldes Forening* (2009) in the EU.
[165] *Copyright Act 1994* (NZ) s 5(2)(a); *Copyright, Designs and Patents Act 1988* (UK) s 9(3). Similar legislation applies in other jurisdictions, including Hong Kong (SAR), India, Ireland, and South Africa.
[166] *Copyright, Designs and Patents Act 1988* (UK) s 9(3).
[167] Ibid s 178.

question should be determined on a case by case basis, with consideration given to the input of each party, and whether that input was artistic in nature, or involved the contribution of skill and labour of an artistic kind.[168] In some circumstances, a user may do nothing more than press a button that causes an AI system to create a work. This can hardly be considered as sufficient input to warrant copyright protection for the user in the final creation over that of the programmer who created the system. However, in some circumstances, the user may invest significant artistic skill and labour into "making the arrangements". For example, the user may use artistic skill in the selection of "training data" they input into the AI system (i.e. a specific selection of romance novels to train an AI system to generate a romance novel in a particular style and of a particular quality). In these cases, authorship of the AI generated work should be determined by balancing the input of the developer and user, and the level of artistic skill and labour involved in each of their roles in the process.

123. It is important to incentivise individuals and organisations to invest in technology and development by ensuring they receive a return on that investment. In turn, whilst it may limit public access to the arts until copyright protection expires for a particular work, the promotion of investment and development through affording copyright protection to the "human authors" of AI generated works will stimulate innovation that will benefit the public.

Copyright infringement by AI systems

124. The Committee also notes the issue of copyright infringement when training AI systems. For example, is it an infringement of copyright to use an artist's song lyrics to train an AI program to create its own song?[169] Furthermore, does that AI generated song infringe the original artist's copyright? Developers and operators of AI systems 'should not only document the creative process when selecting and inputting the underlying art', but 'should also consider evaluating the resulting AI-work to determine whether it is sufficiently transformative before releasing it to the public to mitigate any potential claims of infringement'.[170] However, whilst developers and operators of AI systems must be constantly aware of the implications of copyright law in an AI context, the ambiguity and legal uncertainty surrounding AI systems and AI generated works makes this task exceedingly difficult.

Concluding Comments

NSW Young Lawyers and the Committee thank you for the opportunity to make this submission. If you have any queries or require further submissions please contact the undersigned at your convenience.

[168] See *Nova Productions v Mazooma Games* [2007] EWCA Civ 219.

[169] As was the case with Robbie Barrat in the USA, who used Kayne West lyrics to train an AI system to create its own rap song.

[170] Sarah Ligon, 'AI Can Create Art, but Can It Own Copyright in It, or Infringe?', *Lexis Practice Advisor Journal* (28 February 2019) <https://www.lexisnexis.com/lexis-practice-advisor/the-journal/b/lpa/posts/ai-can-create-art-but-can-it-own-copyright-in-it-or-infringe>.

THE LAW SOCIETY OF NEW SOUTH WALES
youngLAWYERS

Contact:

Jennifer Windsor

President

NSW Young Lawyers

Email: president@younglawyers.com.au

Alternate Contact:

A. Fehrnbach

Ashleigh Fehrenbach

Chair

NSW Young Lawyers Communications, Entertainment and Technology Committee

Email: cet.chair@younglawyers.com.au