24 July 2023

Technology Strategy Branch

Department of Industry, Science and Resources

Canberra ACT 2600

By email: DigitalEconomy@industry.gov.au

Dear Sir/Madam

**Submission - Safe and responsible AI in Australia**

Thank you for the opportunity to make a submission on the Discussion Paper on **Safe and responsible AI in Australia.**

**About the submitter**

I am a practising lawyer and former academic who has practised and written in the fields of digital commerce, privacy and banking and financial services law for more than 30 years. Prior to my current practice, I was Crown Counsel to the Victorian Attorney-General and Premier’s office for 4.5 years, a partner with Clayton Utz in digital business and government and financial services law for 12 years and an Associate Professor of Law at the University of Melbourne Law School.

I am a Deputy Chair of the Digital Commerce Committee of the Law Council of Australia and have contributed to the Law Council’s submission. The Law Council has made a long submission to which I contributed.

I wish to make a short submission focusing only Question 11 in the Discussion Paper on the need for a basic transparency disclosure when AI technology is used. The purpose of this disclosure is to build awareness, transparency and trust for people using AI systems and to prompt them to ask more questions if appropriate, to promote additional voluntary disclosures by providers and users of AI systems and to reduce the risk of people being misled, including by over-reliance on, mistake or misinformation about AI systems or by deliberate fakes and scams which use AI.

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

**Ensuring basic transparency in the use of AI technology to build trust and reduce the risk of people being misled by fakes and scams**

1. AI systems will produce light, sound, images, video, text, automated movements or hardware or software outputs (e.g. execution of a financial trade or smart legal contract) and other phenomena (***AI artefacts***) which so closely resemble human artefacts that humans will find it very difficult if not impossible to distinguish AI artefacts from human artefacts. In some contexts, this will create a serious risk of humans relying on the AI artefact as if it was a human-produced artefact and trusting in the accuracy or completeness or reliability of the AI artefact as if it were a human produced artefact when that trust is not warranted.
2. For example, some of these artefacts will have an effect on a person’s social, educational, commercial or legal position (e.g. a determination or ranking made by an AI-informed system which affects a person’s social, (e.g. assessment of suitability or eligibility for a social opportunity), educational (eg. assessment or admissions decision in an educational setting), commercial (e.g. a renter score or credit rating) or legal (e.g. eligibility for benefits or a licence or penalty for assessed ineligibility for benefits).
3. In other examples, an AI-generated artefact such as or enhanced image or video which portrays situations or events that are not real or did not occur may be indistinguishable from an image or video of real events or situations and cause humans to believe and act as if the real event or situation occurred. An AI generated sound or voice which is designed to mimic a sound or human voice that occurs without AI generation may mislead the hearer into thinking the sound or voice is genuine and cause them to act as if it were genuine to their detriment (e.g. a voicemail from a family member asking for money or access codes). AI generated text is already raising issues of plagiarism and cheating in educational settings.
4. Excellent quality AI “fakes” have the potential (in some contexts, not all), to undermine human trust in AI technology and in a range of human interactions with AI technology as well as to facilitate scamming and deception as well as misplaced trust in reliability.
5. Scamming and deception are age old human practices and AI adds another means. But it is the excellence of the quality of the fakes which AI can produce, the difficulty of detection, and the potentially ubiquitous use of AI artefacts which exponentially increase the incidence of fakes and the risk of deception and justifies consideration now of requiring transparency disclosures in the use of AI in some contexts to reduce the risk of deception and loss of trust.
6. There are many existing civil and criminal laws which penalise deception and fraud. But these operate ex post facto and they assume that the deception can be detected and the perpetrator found and held to account by legal process. It is far from clear that AI fakes will be detected quickly or that the person or system which created or used the fake will be discoverable or, without adequate disclosure, able to be held to account. In addition, AI fakes may be used in broader propaganda and political causes rather than to cause economic harm to specific individuals.
7. The last two paragraphs have focussed on deliberately fake AI artefacts. But the need for disclosure of the involvement of AI is also present to manage the risk of attributing an incorrect degree of reliability to an AI process or artefact (apart altogether from questions of fraud). The involvement of AI in decision making should be disclosed in decisions affecting a person’s life (whether an assessment of eligibility or liability or a ranking) to ensure that users are aware that their lives have been affected by AI and can ask questions of the deployer of the AI system about its reliability.
8. To increase confidence in the use of AI systems, to promote appropriate levels of reliance and to enhance accountability for error or bias and reduce the risk of deception by AI fakes will require a multi-pronged approach – a combination of system design and testing, disclosures, regulation, technical means for detection and correction or error or bias, education of the public to and, if there is error or fraud, proactive correction obligations.
9. I urge the government to consider creating a ***basic transparency obligation*** on the creators and deployers of AI artefacts. This is not intended as a full regulatory response to AI but as a minimum regulatory step at this stage in the development of AI even if the government is minded to take more time to consider other regulatory steps. The ***basic transparency obligation*** is:
   1. to disclose, in most contexts of AI use, that an AI artefact (broadly defined as above and including a decision) has been created using AI and identify the AI system used;
   2. to ensure as far as possible that the disclosure is included with or associated with the content of the AI artefact and cannot readily be removed or disassociated form the content (e.g. a non-removable watermark from an image or video, or a spoken disclosure in a sound file or including the disclosure in a digital file of the artefact and requiring a digital signature of the entire digital file including the disclosure); and
   3. where possible, disclose the persons responsible for creating the AI artefact and (this will be more difficult) the persons responsible for disseminating the AI artefact after its creation.
10. Note that the ***basic transparency obligation*** is not dependent upon the use of personal information about a data subject. It is additional to and independent of any data privacy disclosure obligations.
11. I acknowledge that there are significant details to be worked through as to the form and detail of the disclosure and to what extent it can be made irremovable from the content. It also needs to be considered in which contexts it is not appropriate to require the disclosure. These will not necessarily be the same as “low risk” contexts where other AI regulation may not apply. The contexts where there is a significant risk of inappropriate reliance on or trust in an AI artefact or the risk of deception resulting in harm will not be the same as and the contexts in which the disclosure is mandated. We consider the government should make a commitment to a basic transparency obligation for the reasons stated and consult on the details.
12. I note that the EU AI Act contains a form of basic transparency obligation in Article 52 which the government should consider.
13. *Transparency obligations for certain AI systems*

*1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.*

*2. Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences.*

*3. Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful (‘deep fake’), shall disclose that the content has been artificially generated or manipulated.*

*However, the first subparagraph shall not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties.*

*4. Paragraphs 1, 2 and 3 shall not affect the requirements and obligations set out in Title III of this Regulation.*

Please contact me if you would like me to elaborate on this submission.

Kind regards



Mark Sneddon

Director

Sneddon Legal and Consulting Pty Ltd