Submission No 25

INQUIRY INTO ARTIFICIAL INTELLIGENCE (AI) IN NEW SOUTH WALES

Organisation: UNSW Allens Hub for Technology, Law & Innovation

Date Received: 20 October 2023

20 October 2023

Portfolio Committee No. 1 – Premier and Finance
Legislative Council, New South Wales
Via website: Artificial intelligence (AI) in New South Wales (nsw.gov.au)

Inquiry into Artificial Intelligence (AI) in New South Wales

About us

The **UNSW Allens Hub for Technology, Law and Innovation** ('UNSW Allens Hub') is an independent community of scholars based at UNSW Sydney. As a partnership between Allens and UNSW Law and Justice, the Hub aims to add depth to research on the diverse interactions among technology, law, and society. The partnership enriches academic and policy debates and drives considered reform of law and practice through engagement with the legal profession, the judiciary, government, industry, civil society and the broader community. More information about the UNSW Allens Hub can be found at http://www.allenshub.unsw.edu.au/.

The **UNSW Business School Regulatory Laboratory** ('RegLab') is a community of researchers examining regulation and governance in the UNSW Business School. Reg Lab is a transdisciplinary lab examining the challenges faced by regulators and the regulated in the context of rapidly changing business models. It has a focus on the networked industries sector and data driven innovation. It is jointly funded by the UNSW Business School and external partners (primarily, Google but supplemented by research funding by the Commonwealth).

We have joined forces for the purposes of preparing this submission.

About this Submission

We are grateful for the opportunity to make a submission to the Inquiry into AI in New South Wales. Our submission reflects our views as researchers; they are not an institutional position. This submission can be made public.

We focus on the topics in the Terms of Reference which best fit with our collective expertise, namely:

- "whether current laws regarding AI in New South Wales that regulate privacy, data security, surveillance, anti-discrimination, consumer, intellectual property and workplace protections, amongst others are fit for purpose";
- "the measures other jurisdictions, both international and domestic, are adopting in regard to the adaption to and regulation of AI"; and

- "the effectiveness of the NSW Government's policy response to AI including the Artificial Intelligence Assurance Framework".

Our main points relate to:

- Whether there is a need to define artificial intelligence as such, and the tendency of any such definition to obsolesce (with consequences for using the concept in law or regulatory instruments).
- Proposals for law reform (including the possibility of a law reform commission review) in the areas of data protection, privacy, discrimination law, consumer protection, intellectual property and administrative law.
- Potential amendments to the NSW AI Assurance Framework to improve the assessment of privacy risks.
- Support for standards development.
- The need for guidelines for government procurement of AI systems.
- The need for better co-ordination for policy development across government.
- The advantages of the Swiss over the EU approach to law reform in response to AI.
- The importance of different approaches to public and private sector uses of AI, particularly in the context of accountability and transparency, and some suggestions for the government's own commitment to responsible AI.
- The possibility of general 'generic' laws being supplemented by technology-specific guidance, perhaps pointing to international standards.
- Some suggested contexts in which AI 'bans' should be considered.
- A proposed redirection of focus from seeking 'trust' to ensuring 'trustworthiness'.
- Some suggested characteristics for a risk-based approach.

1. Should "artificial intelligence" be defined and, if so, how?

There is no single or optimal definition of "artificial intelligence". Whether any particular definition is appropriate depends on the purpose and context of the definition exercise. One might, for example, have one definition for describing a university course on artificial intelligence and another when defining the scope of regulation. It is the latter that is most crucial here. In other words, defining artificial intelligence *in government* means asking whether there are particular risks and harms associated with a particular kind of system and, if so, how that kind of system ought to be described for the purposes of legal and regulatory instruments.

That kind of exercise is less useful here because the kinds of harms commonly associated with artificial intelligence are rarely limited to a particular type of technology. Risks associated with lack of accountability can be found in the use of systems relying on explicit programming; a good example is Robodebt. Conversely, problems associated with data-driven decision-making, such as discrimination and unfairness, need not involve an engineered system at all. There are plenty of examples where organisations have relied on traditional statistics or even stereotypes to justify decisions that have disparate impact on certain groups.

The "define a technology and regulate it" approach is only useful where the problem being dealt with aligns with the specific technology. That can sometimes occur, for example with the decision of many governments to prohibit human reproductive cloning. Australia did not prohibit the existence of human clones; that would have involved the sacrifice of one of each identical twin pair. Rather, we prohibited a particular technological means of producing human clones because of ethical concerns around those techniques. That set of prohibited techniques thus needed to be defined.

Similarly, a definition of artificial intelligence is only useful in a legal or regulatory context where it aligns with the problem that is being addressed.

Some of the problems commonly associated with "artificial intelligence" do not in fact depend on the use of "AI systems". Fake photographs, for example, are an old problem; so-called deep fakes are simply an example of creative technologies running ahead of detection tools. Misinformation and disinformation can be authored by humans and propagated through networks and, while artificial intelligence can accelerate generation and target propagation, it is not a necessary ingredient. Encouraging people to self-harm can be done at scale using explicit programming, say outputting "go kill yourself" whenever particular words are used in the input. Inaccuracies can be propagated at scale with explicit programming as demonstrated by Robodebt. Bias is as evident in some statistical techniques as in some machine learning techniques. A badly programmed expert system can yield false answers just like Chat GPT. Those procuring any complex system need a degree of transparency as to how it operates; this problem is not unique to artificial intelligence and, even where information about a system could be communicated, most organisations choose to rely on trade secrets or commercial-in-confidence arrangements.

If regulators address these identified risks and harms only in contexts where "AI systems" are used, the resulting framework will not only be fragile in the event of ongoing technological change, it will fail to deal with problems we are already facing today.

Artificial intelligence – as commonly defined – rightly generates calls for regulatory action. But that action need not be technology-specific. The question to be answered is not "how do we regulate AI" but rather "how do we ensure that our laws operate appropriately and effectively to achieve policy objectives, including in contexts involving AI". There is no need to define "artificial intelligence" in order to address the issues associated with a diverse range of technological practices. Instead, many problems are better addressed through a program to reform and update privacy and discrimination legislation, consumer law, administrative law, and so forth, so that they operate to achieve their goals when applied to current practices associated with the broad frame of artificial intelligence. Some specific suggestions in this regard are set out in the following section.

Further analysis of the challenges of technology-specific approaches to regulation can be found in:

> Bennett Moses LK, 2017, 'Regulating in the Face of Socio-Technical Change', in Brownsword R; Yeung K; Scotford E (ed.), Oxford Handbook of the Law and Regulation of Technology, Oxford University Press, Oxford
>
> Bennett Moses LK, 2013, 'How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target', Law, Innovation and Technology, 5, pp. 1 - 20, http://dx.doi.org/10.5235/17579961.5.1.1

2. What potential risks from AI are not covered by existing regulatory approaches in Australia, and what regulatory action could mitigate these risks?

There are a number of potential risks from artificial intelligence, or particular kinds of artificial intelligence, that are not well captured by existing law. This section includes some examples, but a full audit of the laws that apply in New South Wales and their application in this area would be required to ensure a more comprehensive list.

A. Privacy / data protection law

In some jurisdictions, privacy and data protection regulation is already having a significant impact on the extent to which organisations may use personal information in AI-related activities, including

"the secondary use, disclosure and retention of existing personal data for the purpose of constructing training sets (including disputes over anonymisation), and the use of personal data collected by AI systems for new training sets (including consent issues)".[1]

In Australia, existing federal and state privacy or data protection laws regulate some aspects of the handling of personal information inherent in the design and operation of many AI systems. The *Privacy Act 1988* (Cth) (*Privacy Act*) generally applies to the handling of personal information by businesses operating in Australia with an annual turnover of at least $3 million and Commonwealth government agencies.

New South Wales has also passed privacy legislation that applies to state government agencies and some private sector activities, such as the handling of health information. Gaps in the *Privacy and Personal Information Protection Act 1998* (NSW) ('PPIPA') are discussed further below after a discussion of privacy law reform developments at the federal level.

It is widely accepted that the federal *Privacy Act* is outdated and inadequate to protect Australians' privacy and guard against the serious harms caused by privacy infringements. Throughout the Privacy Act Review conducted by the Commonwealth Attorney General's Department from 2020 to 2022, numerous submissions emphasised the need for urgent reform of these laws. The privacy risks introduced by certain AI systems and the widespread adoption of AI applications increase the urgency of proposed reforms. For example, the *Privacy Act* should be amended to:

- Clarify and expand the definition of "personal information",[2] bringing it into line with international best practice in data protection regulation. This would, for example, aid in ensuring that organisations design systems to appropriately guard against new types of privacy attacks on ML models, such as model inversion attacks and membership inference attacks;
- Introduce a "fair and reasonable" test for dealing with personal information (in addition to consent requirements) to ensure that uses of personal information in AI systems, inter alia, are in keeping with the individuals' reasonable expectations and do not unduly harm the individuals concerned. This recognises the obstacles to genuine consent posed by organisations' control of choice architecture, and consumers' severely limited ability to understand data practices and their consequences, including in the context of personal information used in AI-related activities;
- Bring "small businesses" within the scope of the legislation, acknowledging that the size of the business does not reduce the privacy harms it may create;
- Update the definition of "consent" to mean "voluntary, informed, current, specific, and an unambiguous indication through clear action". Consent to use of personal information for additional purposes, including AI-related activities, should not be found to exist where the organisation makes supply of a product or service conditional upon the individual providing consent for such extra purposes;

1 Graham Greenleaf, 'The "Brussels Effect" of the EU's "AI Act" on Data Privacy Outside Europe' (2021) 171 *Privacy Laws & Business International Report* 3-7.
2 See further Katharine Kemp, 'Ending the Fictions in Modern Data Practices: Submission in Response to the Privacy Act Review Report' (Submission, 31 March 2023) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4521070 2-3, on the appropriate definition of 'personal information'.

- Provide individuals with a direct right of action, such that they can bring proceedings to protect their personal information without depending on the Office of the Australian Information Commissioner to make a determination in respect of the complaint, often after extended delays; and
- Require organisations – including private sector organisations – to undertake a Privacy Impact Assessment (PIA) before undertaking any activity with high privacy risks. This is likely to include, for example, dealing with personal information in systems which may determine access to employment or education, or enjoyment of essential public or private services or benefits, or which potentially cause significant detriment to physical or emotional wellbeing.

On 28 September 2023, the Commonwealth Attorney General announced the government's proposal to enact wide-ranging amendments to the *Privacy Act* in the Government's Response to the Privacy Act Review.[3] Draft exposure legislation is expected in 2024. According to the Government's Response, the proposed amendments are likely to include amendments similar to those listed above, including clarification and expansion of the definition of "personal information"; introduction of a "fair and reasonable test" for dealing with personal information; removal of the "small business" exemption; updating the definition of "consent" to mean "voluntary, informed, current, specific and unambiguous" (although not apparently requiring "clear action"); providing individuals with a direct right of action under the legislation (although only after first complaining to the OAIC); and requiring APP entities to undertake PIAs before the commencement of high-risk activities. Further consultation is contemplated on some of these issues

The Government's Response also agreed that there should be an obligation on organisations to include notices in respect of automated decision making in their privacy policies and to respond to individuals' requests for meaningful information about how automated decisions with legal or similarly significant effect are made. However, in our view, such proposals for mere "notice" or "transparency" are inadequate to address the issues associated with automated decision making, often because of their narrow scope (eg, confining them to fully automated systems or to artificial intelligence systems) but also because transparency alone is not sufficient. We discuss transparency further in section 9 below.

The federal *Privacy Act* currently applies to businesses with an annual turnover of at least $3 million operating in Australia, including such businesses in New South Wales. The *Privacy Act* needs to be updated to meet the challenges presented by AI systems and other technological advances, as outlined above. The New South Wales privacy legislation requires similar reforms. For example, the PPIPA: uses an outdated definition of "personal information" that does not clearly include information that allows an individual to be "singled out" without use of their name or traditional identifiers; does not include a "fair and reasonable test" for dealing with personal information; does not define "consent"; and appears to contemplate the possibility of both express and implied consent.

The NSW AI Assurance Framework, while useful in assisting government agencies to analyse and mitigate risk in the contexts to which it applies, does not currently compensate for these gaps in the legislation and adds some further difficulties in the assessment of privacy risks. For example, in assessing "Risk factors for individuals or communities", the Assurance Framework asks users to consider the risks of "Unauthorised use of health or sensitive personal information". It is not clear why the Assurance Framework does not also raise the risk of unauthorised use of "personal

[3] Australian Government, *Government Response: Privacy Act Review Report* (September 2023) https://www.ag.gov.au/sites/default/files/2023-09/government-response-privacy-act-review-report.PDF.

information", when the obligations under PPIPA relate to "personal information" and are not confined to the narrow categories of information to which "special restrictions" apply. Risks in respect of personal information extend to financial information and information about one's family, for example,4 and unauthorised collection of personal information through inferences or unauthorised disclosure of personal information, which may be an unintended outcome of data use in AI systems. These issues could perhaps be managed by amending the Assurance Framework to recognise that unauthorised collection, use or disclosure of personal information is a risk factor for individuals and communities.

Further, in our view, the AI Assurance Framework's approach to "Sensitive data considerations for AI projects" requires amendment to properly evaluate privacy risks. For example, the question is not whether the AI system uses information on "Religious individuals" or "Racially or ethically diverse individuals" or "Individuals with political opinions" or "Gender and/or sexually diverse individuals" (as the Framework suggests) but whether the system collects information about religious views, race, ethnicity, or political opinions of any individual who is reasonably identifiable. The fact that a reasonably identifiable individual is an atheist, heterosexual of European descent is still sensitive information or information in respect of which section 19 of PPIPA imposes special restrictions.5 This might be resolved by amending this part of the Assurance Framework to refer to the relevant type of information in respect of *any* reasonably identifiable individual, rather than suggesting that only "religious individuals" or "sexually diverse individuals", for example, need to be considered.

B. Discrimination law

We made a separate submission to the NSW Anti-Discrimination review urging it to focus, inter alia, on the question of the impact of AI (and similar) systems for decision-making on the effectiveness of anti-discrimination law. Currently, the draft terms of reference do not refer specifically to this critical issue.

Data-driven influencing, whether using machine learning or statistical techniques, is based on the idea that if we can understand empirical connections between variables, we can predict other variables. When these variables involve human behaviour and result in decisions that affect those humans, fairness and anti-discrimination principles are critical. Currently, discrimination law protects against discrimination on the basis of protected attributes in a range of contexts but does not protect against many examples of "algorithmic bias" because the laws were written at a time when the primary concern was human animus and cognitive limits rather than bad (machine learning) models.

Thus, in the context of machine learning, discrimination law does not operate as effectively as it might. Organisations may well seek to avoid direct discrimination by removing variables without eliminating disparate impact. Complex machine learning algorithms do not necessarily set a "requirement or condition" of (say) being male, rather they factor in correlates with being male among many other variables in ways that influence the outputs, and hence the decisions. Organisations will often also be able to avoid accusations of indirect discrimination by relying on the reasonableness test – to the extent the system sets a "requirement or condition", that it is reasonable to use it where it is generally useful in, say, filtering job applications. The primary problem is that discrimination law does not currently require any testing of black-boxed systems.

4 Which are not designated as "sensitive", or subject to "special restrictions" under PPIPA s 19.
5 The association between cohort size and risk in this section of the Assurance Framework also appears to be inverted.

It would be desirable to reform discrimination legislation so that the need for testing for discriminatory outcomes when a decision affecting a human is made in part or entirely on the basis of data driven inference are laid out more clearly. Guidance for such testing can be found in both international standards and the work of organisations such as the Gradient Institute. Testing may also be required in the context of generative AI and search, to ensure that people do not, for example, only "see" white males in professional roles. Legal changes could reduce the incentive to avoid direct discrimination by deleting variables, which restricts the ability to test for disparate impact. If done well, such requirements would not only apply to artificial intelligence or machine learning but to any potentially discriminatory data-driven process.

C. Administrative Law

Some time ago, there was a proposal for the Australian Law Reform Commission to look at reform of administrative law in light of automated decision-making and artificial intelligence. The reasons for doing this have only expanded since. Currently, the legislative provisions in Australia that authorise the use of computers in administrative decision-making are extremely simple and broad in nature, where they exist at all. They often authorise the use of computers to make decisions on behalf of the ultimate decision maker and deem the decision to be one that decision maker has made. In some cases at the Commonwealth level, there are explicit provisions around certifications as to whether the system is "functioning correctly". However, such requirements are poorly worded for the case of an AI system that is likely to optimise against a particular rate of accuracy rather than function correctly in every case. The provisions need to be more nuanced to recognise the distinction between a program that does not meet specifications and a program that makes mistakes.

Requirements around transparency and accountability for systems used in government decision-making are also critical alongside contestability. Procurement rules, for example, should prohibit government departments from agreeing to terms that require confidentiality as to crucial elements of system operation when systems are used in decisions affecting humans. System requirements should include the ability to generate explanations that mirror the requirements already existing in administrative law for decisions to be accompanied by adequate reasons.[6] In other words, legislation should not simply deem decisions to have been made but also specify requirements around issues like transparency, explainability, and sufficient evaluation and testing. As per the earlier point around definitions, these requirements need not be limited to AI systems but should extend to any situation where a system's output is deemed to be the decision of an authorised decision-maker.

One example of this at the federal level is the *Therapeutic Goods Act 1989* (Cth), where section 7C(1) provides that, 'The Secretary may arrange for the use, under the Secretary's control, of computer programs for any purposes for which the Secretary may make decisions under this Act or the regulations' and section 7C(2) provides that, 'A decision made by the operation of a computer program under such an arrangement is taken to be a decision made by the Secretary.'

There are numerous Commonwealth statutes which include the phrase, 'may arrange for use of computer programs to make decisions'. While we are not aware of New South Wales statutes that deem a system's output to be the decision of an authorised decision-maker, any such existing or contemplated legislative provision is likely to be particularly problematic in the context of the Royal Commission into the Robodebt Scheme.

[6] See also Lyria Bennett Moses and Edward Santow, 'Accountability in the Age of Artificial Intelligence: A Right to Reasons' (2020) 94 ALJ 829.

D. Consumer law

D.1 Consumer protection in financial services

UNSW Allens Hub Senior Research Fellow, Dr Kayleen Manwaring, has recently been involved in a significant research project investigating potential harms to consumers arising from the growing use of AI-related applications in financial services (particularly insurance) and how Australia's current laws apply to these harms. The research project found that these harms range across a number of subject areas, such as discrimination, privacy breaches, digital consumer manipulation and financial exclusion. Although the legal analysis concentrated on gaps in Commonwealth insurance regulation, the research on harms is also relevant to State-based insurance schemes.

The most relevant outputs of the project relating to this inquiry are:

> Bednarz Z; Manwaring K, 2022, 'Hidden depths: the effects of extrinsic data collection on consumer insurance contracts', 45 (July) *Computer Law and Security Review: the International Journal of Technology Law and Practice* 105667 http://dx.doi.org/10.1016/j.clsr.2022.105667
>
> Bednarz Z; Manwaring K, 2021, 'Keeping the (good) faith: implications of emerging technologies for consumer insurance contracts', 43(4) *The Sydney Law Review*, 455, http://www5.austlii.edu.au/au/journals/SydLawRw/2021/20.html
>
> Bednarz Z; Manwaring K, 2021, 'Insurance, Artificial Intelligence and Big Data: can provisions of Chapter 7 Corporations Act help address regulatory challenges brought about by new technologies?', 36 *Australian Journal of Corporate Law* 216

D.2 Digital consumer manipulation

UNSW Allens Hub members have completed significant research on exploitative and manipulative conduct by digital platforms and others providing digital services.[7] There is growing concern by scholars,[8] practitioners,[9] think tanks[10] and industry commentators[11] that the increase in electronic

[7] Katharine Kemp, 'Concealed data practices and competition law: why privacy matters' (2020) 16(2-3) European Competition Journal 628; Kayleen Manwaring, 'Will emerging information technologies outpace consumer protection law? The case of digital consumer manipulation' (2018) 26(2) Competition and Consumer Law Journal 141.

[8] Ryan Calo, 'Digital Market Manipulation' (2014) 82 George Washington Law Review 995; Eliza Mik, 'The Erosion of Autonomy in Online Consumer Transactions' (2016) 8 Law, Innovation and Technology 1; Natali Helberger, 'Profiling and Targeting Consumers in the Internet of Things: A New Challenge for Consumer Law' in Reiner Schulze and Dirk Staudenmayer (eds), *Digital Revolution: Challenges for Contract Law in Practice* (Hart Publishing 2016); Nancy S Kim, 'Two Alternate Visions of Contract Law in 2025' (2014) 52 Duquesne Law Review 303; Anthony Nadler and Lee McGuigan, 'An Impulse to Exploit: The Behavioral Turn in Data-Driven Marketing' (2018) 35 Critical Studies in Media Communication 151; Damian Clifford, 'Citizen-Consumers in a Personalised Galaxy: Emotion Influenced Decision-Making – A True Path to the Dark Side?' (CiTiP Working Paper 31/2017, KU Leuven Centre for IT & IP Law, submitted 15 September 2017), <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3037425> accessed 30 April 2018.

[9] James Halliday and Rebekah Lam, 'Internet of Things: Just Hype or the Next Big Thing? Part II' (2016) 34 Communications Law Bulletin 4.7.

[10] Wolfie Christl, Corporate Surveillance in Everyday Life: How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions (A Report by Cracked Labs, Vienna, June 2017).

[11] For example, Yael Grauer, 'Dark Patterns Are Designed to Trick You (And They're All Over the Web)' (arsTECHNICA, 28 July 2016) <http://arstechnica.com/security/2016/07/dark-patterns-are-designed-to-trick-you-and-theyre-all-over-the-web/> accessed 1 May 2018.

marketing and transactions, and the vast amount of data exposed to public scrutiny by ecommerce, social media, Internet-connected devices and environments and other online activities, may grant marketers a significantly increased capacity to predict consumer behaviour, and use data and behavioural research to exploit the biases, emotions and vulnerabilities of consumers.[12]

The ability of commercial entities to manipulate or exploit consumers is greatly enhanced by the use of AI-related technologies, such as machine learning. AI technologies such as machine learning are at the forefront of the significant amount of data analysis and inferencing required to predict the behaviour of consumers in any number of situations, and to be able to target them in real-time in specific ways, in particular emotional states, in such locations and at the times when manipulation is most likely to be successful.

These practices have been called 'digital consumer manipulation', that is:

> *the use of personalised consumer data collected, processed and/or disseminated by digital technologies, combined with insights from behavioural research, to exploit consumers' cognitive biases, emotions and/or individual vulnerabilities for commercial benefit*[13]

The commercial benefit firms may gain from such techniques include inducing disadvantageous purchases of products or services, extracting more personal information from consumers than is needed for the transaction, and engaging in unjustifiable price discrimination.

One example can be seen in a financial services context. Digital consumer manipulation in this industry often takes the form of 'margin optimisation', a 'process where firms adapt the margins they aim to earn on individual consumers'.[14] Even with most commercial entities' practice of concealing their data-driven business practices where they can, some external evidence exists that EU, UK and US insurance firms, when setting prices, look at a consumer's willingness to pay based on their personal characteristics gained from the insights that external data provides.[15] Machine learning models and algorithms can be used to create inferences of price sensitivity and propensity for switching, based for example on the analysis of consumers' moment-to-moment behaviour on a website or app controlled by the financial firm, the time an individual spends reading terms and conditions, or websites visited before applying to the financial services provider.

Another common example of digital consumer manipulation is so-called 'dark patterns', the design of user interfaces (such as ecommerce websites) to take advantage of certain behavioural biases. Behavioural biases are well-known psychological biases that can be exploited to make it difficult for consumers to select their actual preferences, or to manipulate consumers into taking certain actions that benefit the interface owner rather than the consumer. They are commonly used to manipulate consumers into paying for goods and services they do not need or want, or disclosing personal information that is unnecessary for the transaction and is used by the receiver for their own commercial purposes, or on-sold to third parties. Willis, in her seminal 2020 paper 'Deception by

---

[12] Calo, 'Digital Market Manipulation' (n 8) 995ff; Kim, 'Two Alternate Visions of Contract Law in 2025' (n 8) 312; Helberger, 'Profiling and Targeting Consumers in the Internet of Things: A New Challenge for Consumer Law' (n 8) 140–61; Mik, 'The Erosion of Autonomy in Online Consumer Transactions' (n 8) 1ff; Halliday and Lam, 'Internet of Things: Just Hype or the Next Big Thing? Part II' (n 9) 7.

[13] Kayleen Manwaring, 'Surfing the third wave of computing: Consumer Contracting with eObjects in Australia' (PhD Thesis, University of New South Wales, 2019) 202.

[14] Financial Conduct Authority UK, 'General Insurance Pricing Practices: Interim Report' (Market Study MS18/1.2, October 2019) 21.

[15] Ibid; European Insurance and Occupational Pensions Authority (EIOPA), 'Big Data Analytics in Motor and Health Insurance: A Thematic Review' (Report, 2019) 12, 39.

Design' details how machine learning and 'creative artificial intelligence' are used to optimise the effectiveness of the design and execution of dark patterns. Consumer experimentation can be executed much more quickly and at far greater scale with the use of AI, and website design can be both created and *personalised* by AI applications for micro-segments of consumers in response to the learnings from behavioural experimentation.[16]

Amazon has recently been targeted by the US Fair Trade Commission (FTC) for its use of these 'dark patterns'.[17] The FTC argues that these digital consumer manipulation techniques constitute unfair or misleading conduct in breach of section 5 of the Federal Trade Commission Act. Consumer advocates, and the ACCC in its Digital Platform Services Inquiry, have recently identified them as serious issues for Australian consumers. Some of these 'dark patterns', while harmful to consumers, are not currently captured by the Australian Consumer Law (ACL) (which is part of the law of each State and Territory by virtue of an "application law" enacted in the relevant jurisdiction).[18] Patterns that are not misleading or deceptive (in breach of ss 18 and 29 of the ACL) can be unfairly manipulative in other ways, not currently prohibited under the ACL in the absence of an unfair trading practices prohibition.

Commentators have also raised the possibility of dark patterns fuelled by other features of AI. For example, AI applications or features designed to persuade consumers to:

- believe that a particular sound, text, picture, video, or any sort of media is real/authentic when it was AI-generated (*false appearance*)
- believe that a human is interacting with them when it's an AI-based system (*impersonation*).[19]

Harms from digital consumer manipulation that have attracted condemnation include its potential to:

- impair consumer choice and autonomy;[20]
- create or exacerbate information asymmetry;[21]
- unfairly disadvantage consumers;[22]

---

[16] Lauren E Willis, 'Deception by Design', 34(1) Harvard Journal of Law and Technology 115.

[17] Federal Trade Commission, 'FTC Takes Action Against Amazon for Enrolling Consumers in Amazon Prime Without Consent and Sabotaging Their Attempts to Cancel' (Media Release, 21 June 2023), https://www.ftc.gov/news-events/news/press-releases/2023/06/ftc-takes-action-against-amazon-enrolling-consumers-amazon-prime-without-consent-sabotaging-their.

[18] In New South Wales, the *Fair Trading Act 1987* (NSW) ss 27-28 makes the *Competition and Consumer Act 2010* (Cth), Sch 2 (which contains the ACL) apply as a law of New South Wales as the *Australian Consumer Law (NSW)*.

[19] Luiza Jarovsky, 'Dark patterns in AI: Privacy implications' (blog, 22 March 2023) https://www.theprivacywhisperer.com/p/dark-patterns-in-ai-privacy-implications

[20] Mik, 'The Erosion of Autonomy in Online Consumer Transactions' (n 8); Calo, 'Digital Market Manipulation' (n 8); Helberger, 'Profiling and Targeting Consumers in the Internet of Things: A New Challenge for Consumer Law' (n 8); Marijn Sax, Natali Helberger and Nadine Bol, 'Health as a Means Towards Profitable Ends: mHealth Apps, User Autonomy, and Unfair Commercial Practices' (2018) 41 Journal of Consumer Policy 103.

[21] Donald Bergh and others, 'Information Asymmetry in Management Research: Past Accomplishments and Future Opportunities' (2019) 45 Journal of Management 122, 123.

[22] See also Nir Eyal and Ryan Hoover, Hooked: How to Build Habit-Forming Products (Portfolio/Penguin 2014).

- violate privacy;[23]
- compromise the dignity of consumers;[24] and
- hinder or distort competition.[25]

As the use of machine learning techniques in data analytics increases, and transparency decreases, the likelihood of disadvantages for consumers and other data subjects is likely to increase. The new activities now made possible by hyper-personalised profiling, algorithmic microtargeting of marketing campaigns, and the growth of new data collectors and marketing media via connected devices and environments may lead to an opaqueness unprecedented in the consumer space: in other words, a mass inability to know our own minds.

The current Treasury Consultation Regulation Impact Statement 'Protecting consumers from unfair trading practices' (Aug 2023) considers some of these issues. We would encourage legal reform in these areas, particularly the introduction of a prohibition on unfair trading (comprised of both general and specific provisions), which has significant potential to close the gap on the regulation of undesirable consumer manipulation.

E. Intellectual property law

There are significant issues and uncertainties in how different areas of intellectual property law will be interpreted in response to situations enabled by the use of AI, particularly generative AI. For example:

- currently in Australia's intellectual property law, AI cannot hold the status of an 'author' of a copyright work[26] or the 'inventor' of a patent, [27] and consequently some AI-generated works may not be able to be the subject of IP rights;
- mass reproduction of copyright works to build training datasets for large language models has likely occurred and will continue to occur even when technically a breach of copyright law at the point of copying, enabled by corporate secrecy and the large profits to be gained.[28] In some jurisdictions, this conduct is may be allowed by data mining or similar exceptions in copyright law, and in others (like Australia) it is likely prohibited;
- while the US 'fair use' exception *may* allow for freely available LLM outputs in general[29] (although this is currently the subject of class action litigation in the states),[30] there is a risk

[23] Tal Z Zarsky, 'Privacy and Manipulation in the Digital Age' (2019) 20 Theoretical Inquiries in Law 157, 175; Cass Sunstein, 'Fifty Shades of Manipulation' (2016) 1 Journal of Marketing Behaviour 213, 239.
[24] Ibid; Helberger, 'Profiling and Targeting Consumers in the Internet of Things: A New Challenge for Consumer Law' (n 8).
[25] Maurice E Stucke and Ariel Ezrachi, 'How Digital Assistants Can Harm Our Economy, Privacy, and Democracy' (2017) 32 Berkeley Technology Law Journal 1239, 1256–70; Calo, 'Digital Market Manipulation' (n 8) 1026.
[26] Arts Law Centre of Australia 'Artificial Intelligence (AI) and Copyright' *Information Sheet*, https://www.artslaw.com.au/information-sheet/artificial-intelligence-ai-and-copyright/; *Telstra Corp Ltd v Phone Directories Co Pty Ltd* [2010] FCAFC 149.
[27] Alexandra George and Toby Walsh, 'Artificial Intelligence is breaking patent law' *Nature*, May 2022, 605(7911):616.
[28] See eg the Getty Images lawsuit in the US, Christopher J Valente, Michael J Stortz, Amy Wong, Peter Soskin, Michael W Meredith, 'Recent Trends in Generative Artificial Intelligence Litigation in the United States' *K&L Gates Hub*, 5 Sep 2023, https://www.klgates.com/Recent-Trends-in-Generative-Artificial-Intelligence-Litigation-in-the-United-States-9-5-2023
[29] Matthew Sag, 'Copyright Safety for Generative AI' (2023) 61 *Houston Law Review* 2 (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4438593
[30] Valente et al (n 28).

that the limitations of the contrasting Australian 'fair dealing' exception may be leveraged by rightsholder litigants to prevent access to this technology by Australians; and

- how 'moral rights' are to be enforced[31] in the light of large language model inputs and outputs.

This list is by no means exhaustive, and several other issues may arise. We note that the NSW government has no control over intellectual property law, but it needs to be aware of the potential for the interference in desired outcomes. This is a particularly difficult area, as competing interests do need to be finely balanced, in incentivising innovation by providing economic and moral rights to creators, without unduly discouraging innovation by downstream innovators or productivity gains by users. While some are rightly concerned about the livelihood of creative authors and artists being displaced, as well as a degeneration of our cultural footprint,[32] others have highlighted the dangers of market concentration of LLM productivity tools in the hands of large digital platforms,[33] or effects on research, search engines and interoperability of old and new technology.[34] Ultimately, both Australian creators and Australian users may be detrimentally affected in ways that do not suit government policy goals and community expectations, and therefore we would recommend the NSW government to urge significant consultation on necessary Commonwealth law reform in this area.

3. Further non-regulatory initiatives Australian governments could implement to support responsible AI practices in Australia

Governments should support participation by Standards Australia in international standards development. In particular, financial support would help ensure that not for profit sectors (including consumer groups and privacy advocates) have the ability to participate meaningfully in national and international meetings.

Governments should also develop procurement requirements that ensure core administrative values (fairness, accountability, etc) are factored into decisions as to which system to procure and the terms under which that occurs, including in respect to the ability to disclose important information about how systems operate. AI procurement guidelines are likely to have a significantly greater impact than AI ethical principles which lack meaningful consequences for non-compliance.

4. Coordination of AI governance across government

Significant work has been done on this topic by ANU's Tech Policy Design Centre.[35] We agree with the importance of greater co-ordination across departments and units. This would enable, for example, a clear plan for staging law reform rather than, as occurs now, many consultations cutting

[31] Rita Matulionyte, 'The (Forgotten) moral rights in the age of AI' *Kluwer Copyright Blog*, Feb 7 2022, https://copyrightblog.kluweriplaw.com/2022/02/07/the-forgotten-moral-rights-in-the-age-of-ai/

[32] Dilan Thampapillai, 'Books 3 has revealed thousands of pirated Australian books. In the age of AI, is copyright law still fit for purpose?' *The Conversation,* 29 September 2023, https://theconversation.com/books-3-has-revealed-thousands-of-pirated-australian-books-in-the-age-of-ai-is-copyright-law-still-fit-for-purpose-214637

[33] Katharine Trendacosta & Cory Doctorow, 'AI Art Generators and the Online Image Market' *EFF* 3 April 2023, https://www.eff.org/deeplinks/2023/04/ai-art-generators-and-online-image-market

[34] Kit Walsh, 'How We Think About Copyright and AI Art' *EFF* 3 April 2023, https://www.eff.org/deeplinks/2023/04/how-we-think-about-copyright-and-ai-art-0

[35] Johanna Weaver et al, *Tending the Tech Ecosystem* (May 2022), https://techpolicydesign.au/wp-content/uploads/2022/11/Web_TPDC_Publication_NO.1_2022-3.pdf.

across each other at the same time. The scope for such a co-ordination function would need to be flexible - whether "digital", "AI" or some other phrase is the most appropriate will likely change over time as technology evolves. Alternatively, a broad term like "emerging technologies" could be used, which will then allow for shifts as different technologies come to the fore.

5. Relevant governance measures being taken or considered by other countries

At present, the most well-known and advanced regulation of AI is likely to be found in the European Union, where the AI Act is nearing completion. The EU AI Act defines "artificial intelligence" and takes a risk-based approach to regulating it, with varying obligations on providers and users depending on whether the risks are classified as Minimal, Limited, High, or Unacceptable. While the EU AI Act undoubtedly has merits, it also suffers from the "define the technology and regulate it" approach explained in section 1 above.

Australia could perhaps look to some of the thinking in Switzerland, which encourages a distinct approach to that operating in the EU. In particular, the position paper argues:[36]

> The challenges posed by algorithmic systems are manifold and often have a new dimension or quality, but they are not unique to such systems. Therefore, these challenges should not be covered by a general "AI law" or an "algorithm law". Instead, a combination of general and sector-specific standards is appropriate. The focus here is on the selective adaptation of existing laws.

This is similar to our proposed approach, as explained in response to Section 1.

6. Should different approaches apply to public and private sector use of AI technologies?

In many circumstances, different approaches will apply to the public and private sector, just as they do in other domains. For example, public sector decision-makers have to give reasons for (most) decisions, whereas, in the private sector, 'reasons for decision' (say, on pricing) are rarely required. Private firms are accountable to the market in a very different sense to the way in which government is accountable to citizens.

Context is relevant more generally in determining the best approach. For example, if a private organisation set up an "AI dating site" with minimal transparency ("meet your mystery dream match"), that product will succeed or fail in the market but need not involve heavy handed government regulation around mandated explanations as to the "reasons" for particular matches. On the contrary, legislation authorising Ministers or other decision-makers to rely on systems to make decisions on their behalf should require such systems to have a degree of transparency. This is in line with broader policies around the benefits of "sunlight" in government.

One way in which the New South Wales Government could demonstrate "best practice" in the use of AI technologies would be to adopt a position of a "model user", analogous to "model litigant" obligations. We note that the AI Assurance Framework is a positive step in this direction.

7. Support for responsible AI practices in government agencies

There are a variety of things the New South Wales Government can do to support responsible AI practices in its own agencies:

[36] A Swiss Position Paper can be found at https://algorithmwatch.ch/en/position-paper-legal-framework-for-ai/.

- Provide education and training alongside clear expectations
- Reviewing and updating internal policies, including the NSW Government AI Strategy, AI Ethics Framework and AI Assurance Framework such as outlined in respect of privacy / data protection risks under section 2(A) above.
- Ensure that legislation that authorises reliance on systems for decision-making include provisions setting requirements for such systems (in line with what are currently unenforceable ethical principles) or otherwise requiring compliance with particular frameworks (like the AI Assurance Framework).
- Implement recommendations flowing from the Robodebt Royal Commission

8. Generic vs technology-specific solutions to the risks of AI

As explained in section 1 above, we believe that problem-specific solutions are preferable to technology-specific ones. There are some circumstances where the two align - in other words, the problem relates directly to the use of particular technology. Examples include the use of biometrics for mass identification and the use of automated weapons that make 'kill' decisions.

In other cases, problem-specific, principles-based legislation can be supplemented by subordinate legislation or guidance that explains how general principles apply to particular technological contexts. Where appropriate international standards are available, guidelines can point to standards compliance with which would constitute compliance with particular legal requirements.

9. The role of transparency in addressing potential AI risks

Transparency is a concept that many people are agitating for, but the crucial questions are *what* is rendered transparent, *to whom*, *how* and in which contexts. A driver of an automated vehicle does not need a continuous output from an automated vehicle explaining the logic behind a particular automated decision to steer slightly left to stay in a lane. Rather, they want to know that the car has been evaluated (overall) as safe. On the other hand, the public should be able to find out the logic behind government systems that make decisions affecting them, the nature and quality of training data used, the testing and evaluation of systems that has been conducted (and the results of such), the assumptions on which a system relies, and so forth. Mandating uniform transparency requirements across sectors and contexts would not be helpful in almost all cases. An exception is the proposal (across sectors and contexts) to prohibit misleading uses of AI and automated systems. People should have a right to know when they are interacting with a machine rather than a human (unless they voluntarily relinquish that right for a specific activity, for example in the context of AI research). Similarly, there should be transparency about the involvement of AI in content-generation, so that (for example), an AI-generated image is labelled as such rather than represented as a human artwork.

This does not mean that it is not important to have transparency in particular sectors. One example is the use of AI systems in policing, including facial recognition, predictive policing, and data-driven risk classifications.37 There should be greater public transparency about the use of such systems (along with justifications, privacy assessments, and so forth). Not everything can be made public,

37 Lyria Bennett Moses, 'Oversight of Police Intelligence: A Complex Web, but Is It Enough?' (2023) 60 *Osgoode Hall Law Journal* 289, http://dx.doi.org/10.60082/2817-5069.3892

and operational secrecy is sometimes important, but is often invoked inappropriately to protect controversial programs from scrutiny.38

One way in which governments can provide signalling as to best practice, would be to include the model cards[39] (where applicable) used by the Commonwealth in its use of generative AI. Similarly, a requirement in public sector procurement that model cards are a mandatory part of supply of generative AI products and services would assist with transparency.

A model card is a human-readable document that provides critical information about a machine learning model. It is used to help people understand how the model works, its limitations, and its potential biases.

Model cards usually include the following minimum information:

(a) **Model name and version**: This information helps to identify the model and to track its development over time;
(b) **Model type**: This information describes the type of machine learning model, such as a neural network, large language model, decision tree, or support vector machine;
(c) **Model inputs and outputs**: This information describes the types of data that the model can take as input and the types of data that it produces as output;
(d) **Model training data**: This information describes the data that was used to train the model. This information can be used to assess the model's performance on different types of data;
(e) **Model evaluation metrics**: This information describes how the model was evaluated. This information can be used to assess the model's performance on different tasks; and
(f) **Known limitations and biases**: This information describes any known limitations or biases in the model. This information can be used to help users interpret the model's results and to make informed decisions about its use.

10. Prohibition of high-risk AI applications

There are a variety of contexts in which high-risk AI should be prohibited (or subject to strong restrictions). Ultimately, where the use of AI as the primary decision-maker in a process that would otherwise require rigorous and nuanced human input may result in significant harm or a burden on human rights, AI should be banned. Examples of these contexts may include where lethal force is used in police or military operations, and in formal dispute resolution which should continue to rely on human judges and juries.

11. Trust vs trustworthiness

Public trust in AI should not be sought as an end in and of itself. It is crucial that the public remain appropriately sceptical about computer systems with which they interact so that they take appropriate measures to protect their privacy and challenge illegal decisions. We want the public to be aware not only of the benefits of AI, but also of its limitations. What the government should focus

38 Lyria Bennett Moses and Louis de Koker, 'Open Secrets: Balancing Operational Secrecy and Transparency in the Collection and Use of Data by National Security and Law Enforcement Agencies' (2017) 41 Melbourne University Law Review 530.

39 Margaret Mitchell et al, 'Model Cards for Model Reporting' (2019) Proceedings on the Conference for Fairness, Accountability, and Transparency https://dl.acm.org/doi/10.1145/3287560.3287596.

on is what is commonly referred to as *trustworthiness* - making the systems better so that the public can have confidence in their deployment. The "model user" approach set out above would also assist trustworthiness.

12. Self-regulation vs co-designed regulation

Recent experience in the provision of self-regulatory codes in the mis/dis-information area suggests that industry developed codes will be problematic. Specifically, the problems are likely to be driven by a combination of lengthy delay in code development and information asymmetries. The often-employed approach of self-regulatory codes which become mandatory if they are breached is that there needs to be a regulator in place to monitor such a breach. On the other hand, a co-designed set of codes which are mandatory would work to lessen information asymmetries during the regulatory co-design process. It is also important that the potential regulators are involved. International standards may also be appropriate for local adoption, particularly when involvement of a diverse range of organisations and interests are represented.

The authors of this submission are able to assist in the regulatory co-design process as facilitators; one is also an active participant in standards development for AI.

Yours sincerely,

Lyria Bennett Moses, Katharine Kemp, Annabelle Lee, Kayleen Manwaring, Rob Nicholls