Ms Linda Przhedetsky
University of Technology Sydney

Dr Zofia Bednarz
University of Sydney and ARC Centre of Excellence on Automated Decision-Making and Society

Thank you for the opportunity to participate in the consultation on the Discussion Paper *Safe and Responsible AI in Australia* (June 2023). This submission responds to the Discussion Paper by bringing to your attention a book chapter we co-authored in the forthcoming collection *Money, Power and AI: Automated Banks and Automated States* (edited by Zofia Bednarz and Monika Zalnieriute, Cambridge University Press 2023). Our chapter, 'AI Opacity in Financial Industry and How to Break It', contains findings relevant to questions posed in the Discussion Paper. We note that this submission is made in our personal capacities and not as representatives of any of the organisations with which we are associated.

We attach the full manuscript of our chapter to this submission. We would like to also draw your attention to some particular matters raised in the following questions asked by the Discussion Paper:

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

The problem with definitions in the Discussion Paper is that they are used to delimit the proposed scope of the regulation of AI. As we argue below, 'AI' as such is not an adequate regulatory target. It would be more beneficial to identify the harms we want to prevent now and in the future, and then decide if we need AI (or any other technology / social phenomenon)-specific regulation to address these harms. We expand on this idea in our response to Question 2.

Here, let us consider the following example. A recent class action lawsuit against a US insurer State Farm alleges that the insurer has discriminated against Black homeowners in how their claims have been considered, resulting in their claims taking longer to be approved than White homeowners' claims.[1] The insurer has used a third-party fraud detection algorithm that flags claims which need to be investigated further. It appears that this more detailed investigation has made Black homeowners' claims take much longer to be approved than those of White applicants. The tools for predictive fraud modelling the State Farm insurer has been using are provided by Duck Creek Technologies, an analytics company, which in turn relies on third-party systems, more specifically the FRISS software that delivers real-time fraud assessment on every claim, assigning a risk score, known in the industry

---

[1] Jacqueline Huskey v State Farm Fire & Casualty Company Class Action Complaint, (Illinois District Court, 2022), https://www.classaction.org/media/huskey-v-state-farm-fire-and-casualty-company.pdf.

as the FRISS-score. This score is based on both internal insurers' data, as well as external data, and applies techniques such as AI text mining - for example, to detect and make assessments of language patterns. The plaintiffs allege that the system is biased against Black homeowners. If this is found to be the case, then anti-discrimination laws will apply. At its heart, the issue in this case is discrimination. The issue would be the same had the assessment been made by the insurer's employees who discriminated against Black insureds, rather than discriminatory decisions being made by an automated system. Two considerations logically follow:

(1) Why did the use of AI models result in harm? Would having AI-focused regulation have prevented such discrimination (as allegedly occurred in this case), or any similar instances of discrimination? Why anti-discrimination laws did not sufficiently address the problem? It boils down to addressing the gaps in the legal and regulatory framework applicable (see our response to Q2 below).

(2) If having AI-focused regulation would solve the problems identified, then the following questions arise: how do we define 'AI' as a regulatory target, and would this definition be suitable in other contexts? We believe that it is not AI itself that is the problem here, but rather certain behaviours, such as, in the example above, State Farm relying on third-party solutions (which can, but do not necessarily use AI) without making sure they don't lead to discriminatory outcomes.

If we need an 'AI-specific' regulation, 'AI' needs to be sufficiently broadly defined, so as to include all the current automated decision-making systems, advanced analytics models, models generating text or images, and other systems and algorithms that may not exist yet or may not yet be widely used. We therefore propose to use a more neutral term of 'algorithmic systems', as we do in this submission, to refer to a broad range of computer systems used by private and public entities in their operations.

The second point that needs to be made about definitions is that some important concepts are not defined in the discussion paper. An example is the 'social scoring', which the paper deems to be a 'high-risk' activity. First, linking to the point we are making above (however we define social scoring), we need to ask why only AI-based social scoring needs to be banned, and not all the types of social scoring? Second, certain common practices, such as:

(1) screening of tenants in the rental market, which we discuss in response to Q20, or

(2) decisions made by banks or insurers to determine customers' eligibility for their products, that we discuss in the paper attached,

are in fact examples of social scoring, or at least are not far removed from the concept of social scoring. Therefore, any terms used need to be carefully defined, which can only be done after the identification of specific harms that the Australian Government is seeking to prevent.

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

As outlined in section *3 Domestic and international landscape*, the potential risks of AI are currently governed by a range of general and sector-specific regulations. Our response to this question focuses on the application of technology neutral regulation including consumer safety, negligence and anti-discrimination laws, which apply to situations where algorithmic systems are used. Our view is that existing technology neutral regulation can serve as a solid foundation for mitigating harms and providing redress in situations where algorithmic systems have initiated or exacerbated harms that could have occurred without the use of technologies (albeit, in some cases these harms may have occurred to a different extent). A neat example can be located in copyright: 'Copying digital music is still a breach of copyright - the language of the statute still applies'.[2]

In theory, technology neutral laws should be able to be adapted to a range of situations in which technologies are used now, and into the future. By focusing on the prevention and remediation of undesirable outcomes, these laws remain flexible as technologies evolve, or are applied to new use cases. As explained by Professor Bennett-Moses, prioritising a technology neutral approach in the development and application of regulation means that we are better placed to 'protect values and minimise harm in light of an evolving socio-technical landscape rather than simply asking how technology ought to be regulated.'[3] It is however, important to emphasise that while it is anticipated that technology neutral laws will apply in a range of situations where algorithmic systems result in adverse impacts, many existing laws and their application to algorithmic systems have not been extensively tested in Australian courts. This means that there is little judicial guidance on how existing laws apply in contexts where algorithmic systems are involved in causing or amplifying harms. As algorithmic systems continue to evolve, regular review of existing legislation is required to ensure that it remains suitable within an evolving socio-technical context.

Despite emphasising the merits of applying existing technology neutral legislation to instances where algorithmic systems cause harm, the authors of this submission acknowledge that specific technologies, or specific applications of technologies, can present unique challenges that are not dealt with in existing regulation. In certain instances (for example, where there is a moral objection to a technology), technology-specific regulation may be appropriate. For example, facial recognition technologies pose unique challenges that are outlined in the Human Technology Institute's *Facial Recognition: Towards a Model Law* report.[4] The report recommends that facial recognition technologies require the introduction of legislation that imposes new legal obligations on persons developing and deploying or using facial recognition technologies in Australia, and to those developing facial recognition technologies outside of Australia if their technology is marketed or offered in Australia. As this example demonstrates, the development of technology-specific regulation has a role to play in preventing and remedying harms, and must be designed in ways that are both specific enough to address challenges brought about by particular features or categories of

---

[2] Moses, L.B. (2013), 'How to think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target', *Law, Innovation and Technology*, 5(1), 1-20, 6.

[3] Moses, L.B. (2013), 18.

[4] Davis, N., Perry, L., and Santow, E. (2022), *Facial recognition technology: towards a modern law,* University of Technology, Sydney.

technologies (for example, biometric surveillance technologies), yet comprehensive enough to remain relevant as technologies evolve.

In the next section, we will outline clear challenges in the design, application, and enforcement of existing technology neutral legislation, particularly as it applies to the private sector. The authors recommend that the Australian Government prioritises addressing these challenges to ensure that the goals of existing regulations can be realised in contexts where algorithmic systems are used.

- **The Australian Government should fill gaps in existing legislation, and amend existing legislation to provide clarity on its application to algorithmic systems**

While this submission does not seek to provide a comprehensive list of legislative gaps arising from the application of existing regulation to algorithmic systems, the authors seek to emphasise that existing gaps need to be urgently addressed. Further, we stress that many of these gaps have already been identified. For example, the *Human Rights and Technology Final Report* produced by the Australian Human Rights Commission recommends:

> 'The Australian Government should introduce legislation that provides a rebuttable presumption that, where a corporation or other legal person is responsible for making a decision, that legal person is legally liable for the decision regardless of how it is made, including where the decision is automated or is made using artificial intelligence.'[5]

While the existing body of law concerning liability is clear in certain contexts such as in the financial services sector, the Australian Human Rights Commission suggests that there would be benefit in 'reform that addresses any situations of ambiguity regarding liability for AI-informed decision making'.[6]

There are also instances where existing legislation should be amended. Sheard's work on automated hiring systems explores the application of existing legislation to these technologies.[7] Sheard offers an example in which an employer publishes a digital job advertisement on a social media platform and uses micro-targeting tools to select 'targeting criteria' that enables them to exclude or limit persons with certain protected attributes (for example, age), from viewing the advertisement. Sheard explains that an instance in which targeting a job advertisement to individuals below the age of 50 years would mean that 'unfavourable treatment has occurred because of the protected attribute of age and the employer's conduct is therefore unlawful'.[8] Aside from the challenges of opacity (it would be very difficult for someone to know or prove that the advertisement was targeted to individuals under the age of 50 unless they had access to the employer's advertising account) and the use of proxy attributes (for example, features including postcode that may not be protected by anti-discrimination legislation, but may have a significant correlation with a protected attribute such as race or ethnicity), Sheard explains:

---

[5] Australian Human Rights Commission (2021), *Human Rights and Technology Final Report*, Sydney, 78.
[6] Ibid. 80.
[7] Sheard, N. (2022), 'Employment discrimination by algorithm: can anyone be held accountable?', *UNSW Law Journal,* 45(2), 617-648.
[8] Ibid. 633.

3

'As such discrimination is covert, it would not be captured by the advertising offence provisions in federal and state acts as the advertisement is not published or displayed in any way which indicates an intention to engage in discrimination.'[9]

This is an example that illustrates how existing language contained within anti-discrimination legislation may limit the application of the law in contexts where advertising is conducted differently than it would have been when anti-discrimination statutes were introduced. Consequently, relevant legislation must be amended to capture instances of discrimination such as the one outlined above.

Gaps in existing legislation - whether best addressed through the introduction of new legislation or the modification of existing legislation - will continue to become apparent as we learn about the impacts of algorithmic systems. As such, the Australian Government must commit not only to plugging existing gaps, but to paying special attention to addressing new gaps as they arise to ensure Australia's regulatory frameworks remain suitable.

- **The Australian government should provide regulatory guidance, education and support to developers, procurers, and users of algorithmic systems**

Brownsword offers the concept of 'regulatory disconnection', which helps to describe a mismatch, 'where the law is not fully connected to the technology.'[10] Regulatory disconnection can occur when it is unclear how the law applies to an emerging technology, or when technologies have evolved so much that their capabilities exceed those contemplated when existing regulation was developed. When there is a regulatory disconnection, regulatees can find themselves in a situation where they are uncertain as to what regulation they should follow, and how. As Brownsword explains:

'This violates one of the fundamental principles of the Rule of Law, namely, that regulators should make it clear to regulatees what the regulatory position is. Faced with such uncertainty, how are regulatees to interpret the regulatory void? Should they assume that their desired acts are permitted or, in the absence of clear authorization, should they treat their acts as prohibited?'

It is clear that many users of AI lack guidance on their legal obligations within Australia. The Human Technology Institute found that 64% of Australian corporate leaders and strategic decision makers were already using or planning to use AI systems in their operations.[11] Despite widespread use of AI, only 10% of corporate leaders indicated that their organisation possessed an AI strategy, and less than half of respondents (46%) claimed that their organisation undertook a risk assessment for their organisation's use of AI.[12] This shows that considerable work needs to be done to educate businesses about their legal obligations, and how they apply in contexts where algorithmic systems are used.

In addition to providing general guidance to businesses, special attention must be paid to educating micro, small and medium-sized enterprises (SMEs). The ABS reports that 97.5% of Australian

---

[9] Ibid. 632 (explanation contained in footnote 134).
[10] Brownsword, R. (2008). Rights, Regulation, and the Technological Revolution. Oxford University Press, 160.
[11] Solomon, L. and Davis, N. (2023), *The State of AI Governance in Australia*, University of Technology, Sydney, 12.
[12] Ibid. 29.

businesses can be considered small businesses (defined as those with fewer than 20 employees) while 2.3% of businesses fall within the medium business category (20-199 employees).[13] SMEs typically have fewer financial and human resources, and tend to be slower to adopt new information and communication technologies.[14] By contrast, larger businesses with higher turnover typically benefit from economies of scale, and can more readily adopt new technologies. These dynamics have been recognised internationally, prompting calls for governments to pay special attention to support SMEs as they develop, procure, and use algorithmic systems.[15]

The Australian Government has a critical role to play both in educating businesses about their legal obligations, and how these apply in contexts where AI systems are developed, procured, and used. Education must be supported with detailed, accessible, and practical guidance on how existing regulations apply.[16] An example of effective guidance is the document produced by the Australian Human Rights Commission in collaboration with the Actuaries Institute of Australia, *Guidance Resource: Artificial intelligence and discrimination in insurance pricing and underwriting.*[17] Resources such as this one need to be produced, updated, and publicised across a range of sectors to ensure that businesses using algorithmic systems can comply with the law. Resources of this nature provide businesses with practical guidance and increased regulatory certainty, which is often lacking in voluntary AI principles and frameworks offered across public and private sectors.

- **The Australian government should increase regulator powers, funding, education and resources to deal with the challenges catalysed and exacerbated by algorithmic systems**

As detailed in the authors' attached chapter, 'AI Opacity in Financial Industry and How to Break It', opacity facilitated by algorithmic systems is one of the greatest challenges posed by new technologies. In the context of tenancy application technologies (the focus of Przhedetsky's doctoral research), people applying for rental housing may not know how third party platforms use their personal data to score, rate and rank their applications. Applications such as Certn (a rental application platform available in the North American market) claim to use algorithms, including machine learning techniques to make assessments of rental applicants:

> 'Aside from an obvious red flag like "the guy that takes pictures with automatic rifles wearing a Nazi hat" with the right analysis tools utilizing AI and machine learning, someone's posts can give great insight into their behavior. Uncovering traits like willingness to pay and

---

[13] Australian Small Business and Family Enterprise Ombudsman (n.d.), *Contribution to Australian Business Numbers*, https://www.asbfeo.gov.au/contribution-australian-business-numbers.

[14] Organisation for Economic Cooperation and Development (2004), *ICT, E-Business and Small and Medium Enterprises,* OECD Digital Economy Papers No 86.

[15] See, e.g, House of Lords, Select Committee on Artificial Intelligence (2018), *AI in the UK: Ready, Willing and able?*, Technical report.

[16] The development of guidance (with a focus on compliance with federal anti-discrimination laws) is also recommended by the Australian Human Right Commission's *Human Rights and Technology Final Report* (see page 195): ' The Australian Government should resource the Australian Human Rights Commission to produce guidelines for government and non-government bodies on complying with federal anti-discrimination laws in the use of AI- informed decision making.'

[17] Australian Human Rights Commission (2022), *Guidance Resource: Artificial intelligence and discrimination in insurance pricing and underwriting*, Sydney.

qualities like agreeableness and emotion range that increase the probability of being a good neighbor.'[18]

The key concern here is that it is unclear which data sources are used to make an assessment of a renter's suitability, how these data sources are assessed, and whether the assessment is fair and legally compliant. Further, the opacity involved in the analysis means that it is extremely difficult for a rental applicant to seek redress if they believe that they have been assessed in a way that may not comply with existing regulation. Though the above example concerns a technology that does not operate within Australia, there is evidence to suggest that there is appetite for similar technologies in the Australian market. A 2022 investigation by The Guardian found:

> 'A patent application filed by Snug in 2018, which is still active, suggested that the company's intention was to collect information from users that included friend lists, social media networks and ratings on third-party platforms such as Airbnb and Uber, and to develop a kind of rental credit system.'[19]

If a technology like the one described above were to operate in the Australian market, it would be incredibly difficult for consumers and regulators to identify and prove if and when regulations had been breached. For this reason, the authors suggest that the Australian Government make the following changes, as recommended in the Australian Human Rights Commission's *Human Rights and Technology Final Report*:

> **Recommendation 10:** The Australian government should introduce legislation to require that any affected individual is notified when a corporation or other legal person materially uses AI in a decision-making process that affects the legal, or similarly significant rights of the individual.[20]

> **Recommendation 13:** The Australian Government should introduce legislation to provide that where a court, or regulatory, oversight or dispute resolution body, has power to order the production of information or other material from a corporation or other legal person:

>> (a) for the avoidance of doubt, the person must comply with this order even where the person uses a form of technology, such as artificial intelligence, that makes it difficult to comply with the order

---

[18] Certn (n.d.), *How to screen applicants beyond the credit report*,
https://blog.certn.co/how-to-screen-applicants-beyond-the-credit-report-part-2
[19] Convery, S. (2022, November 17), 'Imperfect match: Australian renters in the dark over use of data by tech company Snug', *The Guardian: Australian Edition.*
https://www.theguardian.com/australia-news/2022/nov/17/imperfect-match-australian-renters-in-the-dark-over-use-of-data-by-tech-company-snug
[20]Australian Human Rights Commission (2021), *Human Rights and Technology Final Report*, Sydney, 194.
Note, the authors also support a similar notification being required when artificial intelligence is materially used in administrative decisions (Recommendation 3 of the report) and the clarification that persons have legal entitlements to receive reasons for administrative decisions (Recommendation 6).

(b) if the person fails to comply with the order because of the technology the person uses, the body may draw an adverse inference about the decision-making process or other related matters.

The Australian Government should urgently act on the above recommendations to create an environment where consumers are empowered to know when they are significantly impacted by the use of algorithmic systems in decision-making processes (the status quo relies on consumers to try and identify if and when they have been impacted) and to strengthen regulators' and courts' abilities to identify instances in which the law has been breached, to identify the extent of these breaches, to issue appropriate penalties, and to provide consumers with meaningful redress.

The authors note that these are just the first steps in empowering regulators and courts to function effectively in a rapidly changing socio-technical context. Any improvement to the status quo requires improved regulator education to increase regulators' AI readiness, particularly in addressing 'limitations in knowledge and skills, insufficient coordination between regulators, issues of leadership and management of organisational and attitudinal change, and resource constraints.'[21] Regulators and courts must be educated about how algorithmic systems function, as well as the risks and harms that may result from their use.

Further, appropriate funding must be allocated to build regulators' capacity and carry out enforcement action, the need for which is increased considerably through algorithmic technologies' abilities to scale harms (for example, the rollout of a discriminatory automated decision-making algorithm across a system used by thousands of people would have greater, more immediate impacts than a situation in which humans were to make each individual decision on a case-by-case basis).

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

    a) where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?

    b) mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

As explained further in our paper: any discussion of transparency relating to algorithmic systems must first recognise that organisations that use algorithmic systems to profit from the opacity of their decision-making processes. While in our paper we focus on private organisations, and more specifically financial firms, similar consideration can be applied to government agencies and public organisations more generally. The problem lies not really with the use of AI tools, and the issue of technical opacity of AI models (often referred to as 'black boxes'), but rather business processes of corporations, and AI-focused regulation alone is unlikely to succeed in breaking the opacity. In our paper we show how the current broad legal and regulatory frameworks often enhance 'AI opacity' in two main ways:

---

[21] Aitken, M., Leslie, D., Ostmann, F., Pratt, J., Margetts, H., and Dorobantu, C. (2022), *Common Regulatory Capcity for AI*, The Alan Turing Institute, 5.
Although this report focuses on the UK context, the issues outlined are common among regulators worldwide.

(1) enabling corporate secrecy, for example through rules on trade secrets, exclusion of de-identified information from the scope of privacy protections, and, more specifically in the financial services context, allowing for opaque scoring systems such as credit scoring or insurance underwriting;

(2) incentivising the use of AI and other automated systems, effectively perpetuating the cycle of opacity in decision-making processes that affect consumers – rules on financial products governance (design and distribution obligations) and open banking (Consumer Data Right), provide useful examples that we discuss in more detail in the attached paper.

As the example of financial services we discuss shows, a more sector-specific approach (which of course could function alongside high-level AI law(s)), would provide better protections for individuals and the society at large, than an approach that simply focuses on regulating AI. It is not the use of algorithmic systems *per se* that increases opacity, but rather certain practices of organisations, which are often enabled, and at times incentivised, by the legal and regulatory framework. In particular, greater clarity and some adaptations of certain existing regulations would be useful, as we are arguing above. In the context of combatting the opacity of processes related to algorithmic systems of all sorts, for example, the following rules could be implemented (these are just some selected examples):

(1) AI models could be excluded from the application of trade secrets rules,

(2) the data minimisation principle could be made more stringent in the context of the Consumer Data Right,[22]

(3) consumers of financial products could have enhanced rights to explanation of credit scoring and insurance underwriting decisions.[23]

In addition, more effective enforcement of rules which are already in place (such as those regarding explanations of decisions made about consumers), and education of regulators and courts in relation to algorithmic systems and AI tools would be beneficial, and not necessarily dependent on the adoption of an AI-specific law.

- **Question 20:** Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to:
  a) public or private organisations or both
  b) developers or deployers or both

---

[22] On the data minimisation principle in the Consumer Data Right see e.g. Zofia Bednarz, Chris Dolman and Kimberlee Weatherall, 'Insurance Underwriting in an Open Data Era - Opportunities, Challenges and Uncertainties', presentation at the Actuaries Institute 2022 All-Actuaries Summit 2-4 May 2022, available at: https://actuaries.logicaldoc.cloud/download-ticket?ticketId=09c77750-aa90-4ba9-835e-280ae347487b.
[23] On the limited rights to explanation of insurance underwriting decisions see e.g. Zofia Bednarz and Kayleen Manwaring, 'Keeping the (Good) Faith: Implications of Emerging Technologies for Consumer Insurance Contracts' (2021) 43(4) *Sydney Law Review* 455, 480-481.

The authors emphasise that voluntary principles and frameworks are no substitute to enforceable regulation. Enforceable regulation, supported by well-resourced, well-funded, well-educated regulators with appropriate investigation powers and enforcement capabilities should be the Australian Government's first priority when improving Australia's regulatory landscape in light of significant technological developments and advances in algorithmic systems.

As demonstrated in the example of tenancy screening technologies (see authors' response to Question 2), algorithmic technologies are being used to score, rate and rank people to determine access to rental housing. We can see from other jurisdictions that unless the use of these technologies is governed by appropriate, enforceable regulation, businesses will exploit the capabilities of algorithmic systems at the expense of consumers:

> 'The [tenancy screening] reports can be created in a few seconds, using searches based on partial names or incomplete dates of birth. Tenants generally have no choice but to submit to the screenings and typically pay an application fee for the privilege. Automated reports are usually delivered to landlords without a human ever glancing at the results to see if they contain obvious mistakes, according to court records and interviews.'[24]

Przhedetsky's research is particularly concerned with the use of algorithmic systems in facilitating or limiting people's access to essential services, including housing. When people cannot access these services, they cannot meet their basic needs. Though this is just one example, it highlights that the use of opaque, potentially harmful algorithms in these markets has significant impacts and needs to be regulated as a priority, and only enforceable regulation can achieve this aim.

We hope that this brief submission, and attached paper are useful. We are more than happy to discuss the content and its implications in further detail.

Full paper starts on the next page

---

[24] Kirchner, L. and Goldstein, M. (2020, May 28), 'Access Denied: Faulty Automated Background Checks Freeze Out Renters', *The Markup,*
https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters