The consultations on Discussion Paper: Supporting Safe and Responsible AI in Australia, provide a welcomed opportunity to contribute to the development of safe, responsible and ethical adoption of artificial intelligence (AI) in Australia. Informed by my own scholarly research and that of other learned academics who focus on the impact of AI on workers, work and employment, the following response first provides some general comments, and then responses to the specific questions posed in the Discussion Paper.

Firstly, the risk-based approach to regulating AI is supported in principle. The "Possible elements" outlined in Attachment C of the Discussion paper are also supported, however I note the following potential limitations in relation to the elements noted below.

Explanations: while this idea, in principle, supports greater transparency and therefore warrants exploration, it may also pose tensions between commercial-in-confidence information and transparent decision-making. As AI becomes more sophisticated, the factors that informed an outcome may not be easily understood or known to users and outcomes may be variable based on a number of factors. ChatGPT provides a case example of variability in outputs but also the innate difficulty in understanding the basis on which outputs were derived in an open-sourced AI. More consultation with AI developers would help to ensure 'explanations' support trust and fairness alongside innovation.

Training: Similarly, it may not be practically feasible nor cost-effective to provide all employees of an organisation with training in the design of AI. To ensure this is a useful and not onerous element for organisations, this element would need to be sufficiently broad and flexible, and perhaps balanced with standards or guidelines to support training. It should also be clear if the onus for the provision of training would be placed on the developers/vendors of the AI or on the organisations that purchase, subscribe or use AI technologies.

Monitoring and documentation: Monitoring would only be effective if organisations were also required to report, either to some form of regulatory body or publicly. In other cases where monitoring and reporting is a requirement (for example safety, gender equity, compliance), organisational size is often a factor in determining the required approach to reporting. In the case of AI, perhaps the extent of use, or the level of AI-risk would provide a more relevant reporting threshold.

Of the options outlined in the Discussion Paper, the proposed EU model appears to provide a sound framework, however new regulation focussed solely on AI would likely be insufficient on its own. To reduce the risk of overlap between AI regulation and existing regulations and minimise the complexity this could present organisations, it would also be important to ensure that existing legislation adequately encompasses AI-based decisions and ADM, and clearly delineates responsibility for those decisions (AI developer or user organisation) and potential avenues of recourse for those who experience adverse outcomes. This would also involve supporting existing regulatory bodies with the power (and required training) to address any potential breaches.

Any regulation (or guidelines or standards) would need to account for the role of all actors in the use of AI. For example, studies have already uncovered bias and discrimination in AI-enabled applications for hiring and employee evaluation (Köchling and Wehner, 2020; Tambe et al. 2021; Tursunbayeva et al., 2022), yet it is not clear who is responsible for that bias – the developer of the AI, the organisation who used the AI, the AI itself, or all parties. This may well be dependent upon who constructs or co-constructs the algorithm and/or provides the training data, and studies have shown that a close working relationship between developers and employing organisations is necessary to mitigate potential bias (Charlwood and Guenole, 2022; Langer and König, 2021). Additionally, candidates and the hiring organisation may not be aware of these in-built biases or the biases may be based on structural inequalities which have not been well considered (Köchling and Wehner, 2020; Ore and Sposato, 2022; Ozkazanc-Pan, 2021; Woods et al., 2020). For candidates, the responsible parties may be less important than the avenue for redressing the bias/discrimination, but for any kind of punitive or compensatory approach, responsibility is a key question, which requires further consideration.

The remainder of this submission provides a response to select questions posed in the Discussion Paper.

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

Response: While simplified for a reader, the definitions provided in the discussion paper are limited. In particular defining AI by 'predictive outputs' it potentially too narrow to account for outputs that are not predictive or encompass the current and future forms of AI that may require consideration. It must also be noted that any prescriptive definitions risk becoming outdated by technologies that have not yet emerged.

The OECD (2019)[1] and the EU's proposed AI Act (2021)[2] may provide potentially more useful definitions which focus on how AI operates and a broader range of potential outputs.

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?

Response: The Australian Government is already undertaking considerable work to address gaps in regulation with respect to AI. Specifically in relation to work and employment, it is pleasing to see the recent review of the Privacy Act which may go some way to addressing the risks AI presents not just for sharing of consumer data, but also to the privacy of employees. With this in mind I note that AI and ADM are also being deployed to manage employee performance, predict employee behaviour, and provide workplace surveillance, as well as manage workforces through the automated allocation and monitoring of task completion. Beyond privacy, such technologies pose risks to fair treatment of workers, potentially erroneous or unfair dismissals, and indirect discrimination. Like the Privacy Act, current workplace surveillance laws and other work-related regulations require reviews to ensure they are fit for purpose and can encompass automated decisions.

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

Response: As noted in the Discussion paper, regulation can be slow to adapt to new advances in AI. A multi-pronged approach that encompasses updates to existing regulations, introduces new regulation in areas not covered by existing legislation, and also provides principles and standards to assist AI developers and vendors, and adopting organisations to develop their own policies to mitigate risks within a specific (technology, use, organisation or industry) context.

Alongside a risk-based regulatory framework, Voluntary Standards for AI developers and vendors could provide a useful non-regulatory framework to guide global AI companies on expected practice in Australia, without stifling innovation. This is particularly the case for AI that affects work or employment. For example, the Fair Conduct and Accountability Standards for digital labour platforms adopted by the Victorian Government specifies practices that digital labour platforms are encouraged to adopt. The Standards encourage transparency and complement/are consistent with existing workplace laws, yet also provide information and an avenue for advice for digital platform workers that previously did not exist.

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

Response: Broadly, for public benefit, it would be desirable to have consistency in approaches applied across public and private sectors, hence a risk-based, rather than a sector-based approach would be more desirable where the risk-based approach provides stricter controls on all organisations dealing with high-risk situations and highly sensitive data.

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

---

[1] For more information refer: OECD, "Recommendation of the Council on Artificial Intelligence" (2019) Available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[2] For more information refer: https://artificialintelligenceact.eu/the-act/

Response: As recent incidences have demonstrated, the Australian Government needs to support the responsible use of AI and ADM with policies, training and education for relevant employees on the use, benefits, and potential harms, and most importantly provide mechanisms for both employees and consumers to raise possible issues or concerns and avenues to seek redress where potential harm has been caused. These mechanisms could include proactive audits of AI in use and complaint mechanisms for both consumers/general public and employees. Any complaint mechanisms should not impose a burden of proof on the complainant.

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

Response: Initiatives should not be framed as or aimed at improving public trust but rather initiatives should be aimed at addressing the issues of lack of transparency in AI/ADM, and the currently limited avenues for redress or support when an individual is negatively affected by AI/ADM. Public trust will be increased by prioritising fairness and transparency in AI deployment through improved regulation and standards, regular public reporting, and by equipping existing (and potentially new) regulatory bodies to hold AI developers and users to account. The intention of this consultation process provides a strong starting point.

References

Charlwood, A. and Guenole, N. (2022), “Can HR adapt to the paradoxes of artificial intelligence?”, Human Resource Management Journal, Vol. 32 No. 4, pp. 729-742. https://doi.org/10.1111/1748-8583.12433

Köchling, A. and Wehner, M.C. (2020), “Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development”, Business Research, Vol. 13 No. 3, pp. 795-848. https://doi.org/10.1007/s40685-020-00134-w

Langer, M. and König, C.J. (2021), “Introducing a multi-stakeholder perspective on opacity, transparency and strategies to reduce opacity in algorithm-based human resource management”, Human Resource Management Review,Vol. 33 No. 1, https://doi.org/10.1016/j.hrmr.2021.100881

Ore, O. and Sposato, M. (2022), “Opportunities and risks of artificial intelligence in recruitment and selection”, International Journal of Organizational Analysis, Vol. 30 No. 6, pp. 1771-1782. https://doi.org/10.1108/IJOA-07-2020-2291

Ozkazanc-Pan, B. (2021), “Diversity and future of work: inequality abound or opportunities for all?”, Management Decision, Vol. 59 No. 11, pp. 2645-2659. https://doi.org/10.1108/MD-02-2019-0244

Tambe, P., Cappelli, P. and Yakubovich, V. (2019), “Artificial intelligence in human resources management: challenges and path forward”, California Management Review, Vol. 61 No. 4, pp. 15-42. https://doi.org/10.1177/0008125619867910

Tursunbayeva, A., Pagliari, C., Di Lauro, S., & Antonelli, G. (2022), “The ethics of people analytics: risks, opportunities and recommendations”, Personnel Review, Vol. 51 No. 3, pp. 900-921. https://doi.org/10.1108/PR-12-2019-0680

Woods, S.A., Ahmed, S., Nikolaou, I., Costa, A.C. and Anderson, N.R. (2020), “Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges”, European Journal of Work and Organizational Psychology, Vol. 29 No. 1, pp. 64-77. https://doi.org/10.1080/1359432X.2019.1681401