Centre for Culture and Technology (CCAT)
Curtin University

CCAT submission in response to the Safe and Responsible AI in Australia Discussion Paper

July 2023

Acknowledgements: this submission has been prepared by Kai-Ti Kao, Katie Ellis, Eleanor Sandry, Mike Kent, Tama Leaver and Stuart Bender from the Centre for Culture and Technology, Curtin University.

1. Who we are

The Centre for Culture and Technology (CCAT) is a research Centre in the Faculty of Humanities at Curtin University. We partner with community, government, and industry to ***codesign socially just digital futures***. Our research focuses on how cultural practices are changing in relation to digital technologies and platforms in areas such as accessibility, health communications, intimacy, social life, popular culture, knowledge production, commerce, politics, and activism. At the heart of our work lies a commitment to investigating the opportunities and challenges that arise from the rapid integration of digital technologies and media platforms into everyday life and culture for specific social groups such as people with disability, Indigenous people, children and youth, and people with diverse sexualities and gender identities.

We are motivated by the proposition that the study of culture, with its emphasis on identity, meanings, relationships, power and values, needs to be better integrated with the study of media and digital technologies, including Artificial Intelligence (AI).

Within CCAT our research on AI recognises both opportunities and challenges related to the integration of AI with our existing communication strategies and social and cultural institutions. For example, we recognise that AI has great potential to foster the inclusion of people with disability in health, justice, education, and housing.[1] It is also an important tool within education, but not without its challenges.[2]

Our research calls attention to how AI have been developed with particular conceptions of "human" in mind. The AI field, from the technology industry to scientific research, has been dominated by White men and their outputs both reflect and privilege their experience of the world.[3] The resulting ways that AI function to exclude, marginalise, and discriminate against those who do not fit this mould has been well-documented.[4]

While AI is defined as "a collection of interrelated technologies used to solve problems autonomously and perform tasks to achieve defined objectives without explicit guidance from a human being"[5] human interaction is essential to the current discussion and way forward for AI governance.[6] It is through this human interaction that we ensure "the design of AI, its inputs, outputs and regulatory framework do not preclude entire subsets of the population from experiencing its benefits".[7] This is essential to establishing trust amongst the Australian population.

2. Summary

Our submission response to the Safe and Responsible AI in Australia Discussion Paper can be summarised across five key themes:

1. **Literacy**
    - A focus on AI literacy and public education can help build greater trust in AI; currently there is confusion over what AI can do and too many assumptions made about the potential of AI.
    - A focus on digital literacy is therefore also important in helping to dispel much of the mystique that currently surrounds AI. Digital literacy can help us recognise that many of the concerns and issues raised in relation to public engagement with AI are not tied to specific technologies but are more deeply embedded in society and have not been adequately managed in the past.

2. **Risk**
    - There is an assumption within this Discussion Paper that the risks related to AI have already been identified or are foreseeable. It is important that we recognise that it is not possible to predict all the future risks that are likely to emerge through the use of AI, and some degree of flexibility will need to be embedded within the proposed governance approach to account for these unforeseen risks.
    - The proposed risk management approach is far too broad and, in line with other jurisdictions' approaches, we advise that at least four risk categories should be implemented (low, medium, high, and very high).

3. **Interoperability**
    - Given the range and variety of technologies that fall under the umbrella term "AI", it is important to ensure that definitions used in policy are consistent across different governments departments and organisations.
    - Australia's overall approach to AI governance should be mindful of what other government departments, organisations, and bodies are doing, both within Australia and internationally.

4. **Responsibility**
    - It is important that Australia's approach to governing AI protects and ensures culturally appropriate development, applications, and uses of AI.
    - When considering the issue of responsibility, care must be taken to ensure that the wording of policies do not place responsibility inherently or inadvertently with the AI technologies. Instead, there must be clear recognition that responsibilities are held by the people involved in designing, deploying, applying, and managing AI, in particular where the AI is being made available to others.

5. **Opportunity**
    - Australia's approach to governing AI should indeed be cautious of not stifling the opportunities offered by AI. Attention should be focused on looking at opportunities AI technologies can offer for social welfare and the inclusion and betterment of marginalised groups, alongside how AI can contribute to industry and economic growth.

3. Responses to questions

Below is a more detailed response to specific questions posed in the Discussion Paper.

Q2 What potential risks from AI are not covered by Australia's existing regulatory approaches? Do you have any suggestions for possible regulatory action to mitigate these risks?

The broad risk management approach reflected in this discussion paper assumes that risks related to AI are identifiable and foreseeable. Given that AI is rapidly evolving with new use cases and applications being discovered daily, allowance should be made for some degree of flexibility that adapts to unexpected uses and risks.

We also believe that the three risk levels proposed in the Discussion Paper are inadequate for capturing the range of current and potential risks posed by AI. We recommend that, in

line with the approach taken by the EU and Canada, Australia employ a four-risk level approach to managing AI (low, medium, high, and very high).

Q3 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.

We recommend greater public education focusing on both AI and digital literacy. Digital literacy refers to "the ability to search and navigate, create, communicate and collaborate, think critically, analyse information, and address safety and wellbeing using a variety of digital technologies".[8] As AI is a rapidly evolving technology it is difficult to provide clear benchmarks for what can be considered "AI literate". However, we are in alignment with UNESCO in viewing AI as "the basic grammar of our century" and believe AI literacy will encompass the ability to understand what AI can and cannot do, when and how it can be useful, and when and how it should be questioned.[9]

- AI Literacy
  Currently, there is a great deal of uncertainty and confusion circulating about what AI technologies are realistically capable of.[10] The term "AI" has been used as an umbrella to encompass a complex range of technologies from robotics to machine learning and natural language processing. The type of AI that the general public engages with on an everyday basis actually comprises what is known in AI circles as "Narrow AI" or "Weak AI". "Strong AI" by contrast includes the categories of "Artificial General Intelligence" and "Artificial Super Intelligence" and is, as yet, not considered to exist outside of theory and fiction. Even more advanced versions of AI that are available to us, such as self-driving cars and generative AI tools like ChatGPT, do not constitute "Strong AI". However, media reports and public discussions that use the umbrella term "AI" do not differentiate between these different types of AI.

  This is leading to confusion that is fostering and overexaggerating appraisals of AI's positive and negative impacts on society. As a result, we are seeing the circulation of common AI imaginaries such as: the idea that AI will free us from the tedium of human labour, drastically improve healthcare and extend our lifespans, take our jobs and create a crisis of unemployment, or destroy us by becoming sentient and taking over the world. The circulation of such imaginaries detracts from critical conversations about AI's current and actual impact on society, including discussions about who benefits, who is harmed, and how our potential capabilities are being both afforded and stifled. Greater public education focusing on AI literacy can help clarify understandings of what AI is and how it is currently being used. This can not only help to build greater public trust and confidence in AI, but also help to mitigate any potential and as yet unforeseen risks.

It is also important that news media outlets undergo AI literacy training. UK research into media coverage of AI revealed that news media have a tendency to inflate the possibilities of these technologies, over-rely on industry figures as authoritative sources, and uncritically accept the promises offered by these industry figures and technologies.[11] These tendencies are also being reflected in Australian news media and are impacting the public's perception, understanding, and trust in these technologies. Greater AI literacy in news media will therefore have two benefits: a) more accurate reporting of AI and its impacts on society; and b) further enhance the AI literacy of the public via more accurate and realistic communication about AI.

- Digital Literacy
  It is important to recognise that many of the risks related to AI use, particularly in relation to media and communication, are deeply embedded in society and not tied to specific technologies. Similar risks have emerged in the past in relation to social media and the internet (e.g. the "black box" of algorithms) but have not been adequately managed. By focusing on public education and digital literacy, we equip ourselves better to manage the currently known and foreseeable risks, and better insulate ourselves against the unforeseeable and unexpected.

Q4 Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia.

In responding to this question, we note a number of current government reviews on this topic as well as international interest in AI governance by organisations such as the Association of Internet Researchers (AoIR). It is important that there is a focus on interoperability of definitions and regulatory approaches. Any policies or governance approach developed should also be mindful of what is happening elsewhere within Australia and internationally. The current Multicultural Framework Review, as well as the inquiry into the use of generative artificial intelligence in the Australian education system, are both examples of ongoing policy initiatives that will impact Australia's broader governance and regulatory approach to AI.

Q9 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.**

Transparency is vital at **all** stages of the AI lifecycle, from design and development through to implementation and use. There is currently a lack of public trust in both governments and industry regarding AI development, use and governance.

Public trust in AI can be improved by communicating information in plain language and readily available formats that are accessible by all members of our population including disability groups and non-English speaking communities. To this end we recommend an inclusive design approach to developing appropriate strategies for communicating information about AI. We have elaborated on this further in our response to Q11.

The other element of transparency needed, particularly for generative AI, is to be clear what data sources the AI has been developed from. This will be important for both public trust and understanding, but also business confidence in using different AI tools and the risks associated with potential litigation from owners of the data on which the AI tools draw.

Q11 What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?

To build public trust in AI it is important that the public clearly understands what AI is and how it is being used. Therefore, our recommendations for government initiatives and action are focussed around developing clear communication with the public about AI. We have three key recommendations:

1. Communication about AI must be inclusive and accessible.

   We recommend an inclusive design approach to communication strategies. Inclusive design focuses on edge users or people overlooked in the design process (such as people with disability or non-English speaking communities), to improve value for all users. Our research on health communications during the COVID-19 pandemic for example found that when communication strategies were designed for people with disability, the entire population benefited and were better able to understand and trust important government messaging.[12]

2. Focus on building AI literacy and digital literacy.

   As mentioned in our response to Q3, we recommend greater public education that focuses on AI literacy *and* digital literacy. This is in line with UNESCO's *Recommendation on the Ethics of Artificial Intelligence*, which views AI and digital literacy as important for ensuring effective public engagement and participation, and vital to helping the public avoid undue influence from the misuse of AI or AI-generated content as well as make informed decisions about their own AI use.[13] Such a public education initiative will help build public confidence and trust in AI use,

better empower the public in their engagement with AI technologies, and foster greater resilience to as yet unforeseen risks posed by future AI developments.

As an example of why AI literacy is important, we refer to the recent instance of Australian doctors and hospitals using ChatGPT to write medical notes.[14] We note that the Australian Medical Association (AMA) have called for stronger AI rules and healthcare-specific regulation in their submission response to this Discussion Paper. We would also add that greater AI literacy would have been a benefit in this instance, helping both medical staff and patients recognise the inherent privacy and security risks, as well as the possibility that the material generated by the chatbot could be incomplete, misleading, or wrong. In this instance, both medical staff and patients also needed to be aware that any data entered into ChatGPT will be used by its parent company to further develop and commercialise future AI products, thus posing a considerable ethical concern regarding the use of patients' information.

3. Use of AI must be transparent and clearly stated.

   Public trust in AI is more likely to increase when the public are aware of how and when AI is being used. To this end, we recommend that future AI regulation mandate that uses of AI should be clearly stated as such, especially when the end result, product, or service will impact the public.

   We particularly advise such transparency of use in relation to AI-generated news media content. The use of AI to create and publish news articles, images, and video has been rising in recent years. Recent revelations that News Corp Australia have been publishing as many as 3,000 AI-generated articles a week have highlighted the extent of this practice,[15] which is likely to increase as more news organisations explore AI-generated content. This has significant implications for public trust in news media as AI not only circulate embedded biases but potentially spread misinformation.[16] More careful consideration of the use of AI in news media is warranted, however as a start we recommend that all news media organisations in Australia be required to clearly state when AI has been used to generate news articles and news-related content.

Q14 Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?

A risk-based approach is functional but should not be the only approach to regulating and governing AI in Australia. Combining a risk-based approach with measurable targets for increasing public literacy in AI, for example, will help build a society more aware of and resilient to the risks associated with AI, as well as being prepared to embrace its potential.

Q17 What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?

We support the elements suggested in Attachment C and particularly endorse the use of Notices and Explanations. We again highlight the importance of clear and accessible communication with the public and recommend that Notices and Explanations should be made easily available in plain language and formats that are accessible by members of the population with disabilities or for whom English is not their dominant language.

Q20 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?**

We assert that the risk-based approach for responsible AI should be mandated through regulation to ensure compliance, build public trust and confidence, and better ensure best possible practice and use of AI. It should apply in both private and public organisations, and to both developers and deployers.

4. Conclusion

Finally, in addition to recommending a four-level risk-based approach (low, medium, high, and very high), tightly integrated across all government sectors and keenly aware of international approaches, definitions and breakthroughs, we encourage the review go further and engage in an AI literacy program. It is important that we remember that rushed and hasty use and implementation of such technologies, as we saw with Robodebt, can have disastrous consequences. It is also vital that we recognise that these technologies are not only likely to have significant impacts on our people and society, but also our lived and natural environments. When considering the impacts of AI technologies, we should be mindful of the costs of the Climate Crisis we have and are currently witnessing in Australia and around the world. In order to make the most of the benefits that AI has to offer, we must ensure that our governance of AI and assessment of risk should incorporate the potential cost to lives, livelihoods, and living environment.

---

[1] Ellis, K., Kent, M., & Peaty, G. (2017). Captioned Recorded Lectures as a Mainstream Learning Tool. *M/C Journal, 20*(3). doi:10.5204/mcj.1262
McRae, L. (2022). *Ethical AI, criminal justice and sign language: A literature review*. Centre for Inclusive Design
[2] Bender, S.M. (2023). Coexistence and creativity: screen media education in the age of artificial intelligence content generators, *Media Practice and Education*, DOI: 10.1080/25741136.2023.2204203
Ellis, K., Kao, K., & Kent, M. (2020). *Automatic Closed Captions and Immersive Learning in Higher Education*. Curtin University

Bender, S.M. (2023). How Might We Co-Exist With Artificial Intelligence Content Generators in Subject English? *English Teachers Association (WA) Conference*, May 13th, University of Western Australia.

[3] Benjamin, R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity Press.

Cave, S., & Dihal, K. (2020). The Whiteness of AI. *Philosophy & Technology*, *33*(4), 685–703. https://doi.org/10.1007/s13347-020-00415-6

Crawford, K. (2016, June 25). Artificial Intelligence's White Guy Problem. *The New York Times*. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

[4] Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency*, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html

Collett, C., Neff, G., & Gomes, L. G. (2022). *The Effects of AI on the Working Lives of Women*. UNESCO, OECD, IDB. https://unesdoc.unesco.org/ark:/48223/pf0000380861

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating Systems: Gender, Race and Power in AI*. AI Now Institute. https://ainowinstitute.org/ discriminatingsystems.html

Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., Morris, M. R., Rankin, J., Rogers, E., Salas, M., & West, S. M. (2019). *Disability, Bias and AI*. AI Now Institute. https://ainowinstitute.org/disabilitybiasai-2019.pdf

[5] Hajkowicz, S., Karimi, S., Wark, T., Chen, C., Evans, M., Rens, N., Dawson, D., Charlton, A., Brennan, T., Moffatt, C., Srikumar, S., & Tong, K. (2019). *Artificial intelligence: Solving problems, growing the economy and improving our quality of life*. CSIRO Data61.

[6] Sandry, E. (2023). HMC and Theories of Human–Technology Relations. In Guzman, A. L., McEwen, R., & Jones, S. (Eds), *The Sage handbook of human–machine communication*. SAGE Publications Ltd, https://doi.org/10.4135/9781529782783.

[7] Amin, M., & Reid, G. (2018). Prejudice in Binary: A Case for Inclusive Artificial Intelligence. Retrieved from https://acola.org/wp-content/uploads/2019/07/acola-ai-input-paper_inclusive-design_amin-reid.pdf

[8] McLean, P., Oldfield, J., & Stephens, A. (2020). *Foundation Skills for Your Future Digital Framework*. Australian Government, Department of Education, Skills and Employment. https://www.dewr.gov.au/foundation-skills-your-future-program/resources/digital-literacy-skills-framework

[9] UNESCO. (2022, February 23). UNESCO releases report on the mapping of K-12 Artificial Intelligence curricula. UNESCO. https://www.unesco.org/en/articles/unesco-releases-report-mapping-k-12-artificial-intelligence-curricula

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380455

[10] Bridle, J. (2023, March 16). The stupidity of AI. *The Guardian*. https://www.theguardian.com/technology/2023/mar/16/the-stupidity-of-ai-artificial-intelligence-dall-e-chatgpt

Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence*, *26*(1), 3–18. https://doi.org/10.1177/1354856517715164

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology*, *24*(1), 4. https://doi.org/10.1007/s10676-022-09624-3

[11] Brennen, J. S., Howard, P. N., & Nielsen, R. K. (2018). *An Industry-Led Debate: How UK Media Cover Artificial Intelligence*. University of Oxford. https://reutersinstitute.politics.ox.ac.uk/our-research/industry-led-debate-how-uk-media-cover-artificial-intelligence

[12] Goggin, G., & Ellis, K. (2020). Disability, communication, and life itself in the COVID-19 pandemic. *Health Sociology Review, 29*(2), 168-176. doi:10.1080/14461242.2020.1784020

[13] UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000380455

[14] Taylor, J. (2023, July 27). AMA calls for stronger AI regulations after doctors use ChatGPT to write medical notes. *The Guardian*. https://www.theguardian.com/technology/2023/jul/27/chatgpt-health-industry-hospitals-ai-regulations-ama

[15] Meade, A. (2023, July 31). News Corp using AI to produce 3,000 Australian local news stories a week. *The Guardian*. https://www.theguardian.com/media/2023/aug/01/news-corp-ai-chat-gpt-stories

[16] Gal, U. (2023, June 23). *Replacing news editors with AI is a worry for misinformation, bias and accountability*. The Conversation. http://theconversation.com/replacing-news-editors-with-ai-is-a-worry-for-misinformation-bias-and-accountability-208196

Leffer, L. (2023, January 17). *CNET Is Reviewing the Accuracy of All Its AI-Written Articles After Multiple Major Corrections*. Gizmodo. https://gizmodo.com/cnet-ai-chatgpt-news-robot-1849996151