Source Transitions Pty Ltd submission to the Department of Industry, Science and Resources Discussion Paper

Safe and responsible AI in Australia

25 July 2023

Source Transitions welcomes the opportunity to provide a submission to the Department of Industry, Science and Resources discussion paper ***Safe and Responsible AI in Australia***.

We also note that we support Responsible AI in Australia and globally and are an active participant in the Australian AI Ecosystem Discovery Platform.

The discussion paper was published in the context of other reports and programmes initiated by the Australian Government which are all relevant [1].

- The Australian Government has consulted on automated decision making and AI regulation in 2022 under the previous government.
- A National Science and Technology Council's Rapid Research Report on generative AI was published on the 1st June 2023
- Reviews of the Privacy Act, Roundtable on Copyright, Ip Australia working group, Digital Platform Regulators forum, Digital Platform Services Inquiry, Select Committee on social media and Online Safety are all under way.
- The AI Ethics Principles were amongst the first ones to be published globally in 2019
- Safety by Design approach adopted.
- AI use at Department of Finance, DTA, commonwealth Ombudsman
- Sector specific initiatives and programs

This document consolidates Source Transitions views in response to Federal Government public consultation [2] and discussion paper [3], seeking feedback from all organisations and individuals about how to mitigate any potential risks of AI and support safe and responsible AI practices.

Executive summary

We understand that the fundamental underlying question being asked is:

> How do we keep Australia and Australians safe while taking advantage of AI technologies?

Source Transitions shares the Australian Government's view that legislating AI technologies development and applications is warranted given the risks they introduced into society.

However, we feel that the discussion paper could be stronger and clearly specify what the Australian Government want AI to do for Australia and for Australians; a clear intent about what we want to achieve, as a people, with these technologies will help better define the legislative framework to ensure we achieve it. It also allows us to think more positively about the technologies and empowers people to think creatively about opportunities they can create.

Our key recommendations can be summarised as follows:

- It is critical for Australia to develop trust and transparency in AI systems so that it facilitates the adoption of the technologies and helps boost innovation. It is also part of developing **AI literacy across sectors** so that economic and social agents can take part and help guide how they want to use AI and for which purposes.
- Disclosing and establishing clear guidelines on what the **real cost of AI is in terms of environmental and social indicators** and not just costs, will help organisations and individuals better identify where the opportunities really lie in a carbon and resource constrained world.
- **A bigger story of how AI can help Australia transition** to better social, economic and environmental outcomes should be framed; one with a strong intent from government to direct innovation for purpose. Without a clear direction about what AI should be doing for Australia and Australians, little of the potential opportunities will be achieved.
- While some aspects of AI legislations should be non-negotiable, others require **a more iterative approach to governance** which will be able to adapt to fast changing technologies and learning outcomes that can only be derived from "doing". In an AI world, adaptive or hybrid governance will likely be the best approach to learn and be flexible in the face of constant change.
- Finally, **providing infrastructure such as sandboxes** to test regulatory compliance of AI systems will help close the gap between industry and government knowledge and make government less susceptible to industry-centric views.

Source Transitions already works with businesses and the not-for-profit sector to help them develop pathways to sustainable futures using AI as an enabler. As a result of these initiatives, AI literacy increases significantly; to the point where participants can imagine what they would do to solve the problems they have, to engage with complexity and to dare imagine new business models or new products and services. AI technologies can serve their sustainability efforts (financial, social, environmental) in collaboration with their ecosystem.

Source Transitions takes this opportunity to show its interest to collaborate with the Department of Industry, Science and Resources and other public and private organisations to collaboratively design Responsible AI governance frameworks cross-sector and an AI for Sustainability approach for Australia.

Contents

The discussion paper contains 20 specific questions, we understood these to be prompts for consideration rather than requiring an individual answer for each. However, we have kept the questions for ease of data collection and indicated n/a when the question was not specifically answered.

Definitions

1. *Response to the definitions*

Definition of Artificial intelligence

The definition proposed in the paper is adequate but incomplete. AI is mainly used in the public as an umbrella term. The definition should therefore be accompanied by a list of specific techniques and approaches used for AI development or its application. The definition should be kept up to date considering market and technological developments as indicated in the preparation for the regulation of AI in the EU[1]. An important part of the role of government in the field is to ensure clarity and transparency so that trust can be built around the set of technologies AI can power. Initiatives such as the EU AI definition and set of examples to illustrate what AI does in practice will help guide the public, businesses and government agencies understand how they could be affected by AI[2].

The definition of AI has evolved over time and allowing for that definition to continue evolving is key to keeping abreast of how the field is developing.[3] The EU parliament proposes to establish a technology-neutral, uniform definition for AI that could be applied to future AI systems[4].

Definition of AI governance

The definition of AI governance in the paper is limited to regulatory and voluntary mechanisms. This is not sufficient an explanation in that it does not establish the context for governance of AI systems.

For the purpose of transparency and clarity, AI governance should be specified from the onset. The definition should indicate how legislation and principles constitute the environmental governance layer, but they are also integrated as part of practical processes and mechanisms at an organisational level (figure 1)[5]. Clearly specifying how AI governance is established in organisations and where the

[1] EU Parliament (2021) *Proposal for a Regulation of the European parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts COM/2021/206 final*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206, accessed on the 26 July 2023

[2] European Parliament News (2023) *What is artificial intelligence and how is it used?* https://www.europarl.europa.eu/news/en/headlines/society/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used, accessed on the 26 July 2023

[3] Grosz, B. J., Altman, R., Horvitz, E., Mackworth, A., Mitchell, T., Mulligan, D., & Shoham, Y. (2016). *Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence* (AI100) 2015 study panel report. https://apo.org.au/node/210721

[4] European Parliament News (2023) *EU AI Act: first regulation on artificial intelligence*, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence, accessed on 26 July 2023

[5] Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). *Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance*. 2022. http://arxiv.org/abs/2206.00335

regulatory boundaries start, and end would provide companies and organisations developing or using AI an overview of how the regulatory frameworks are only part of the issue of AI governance.

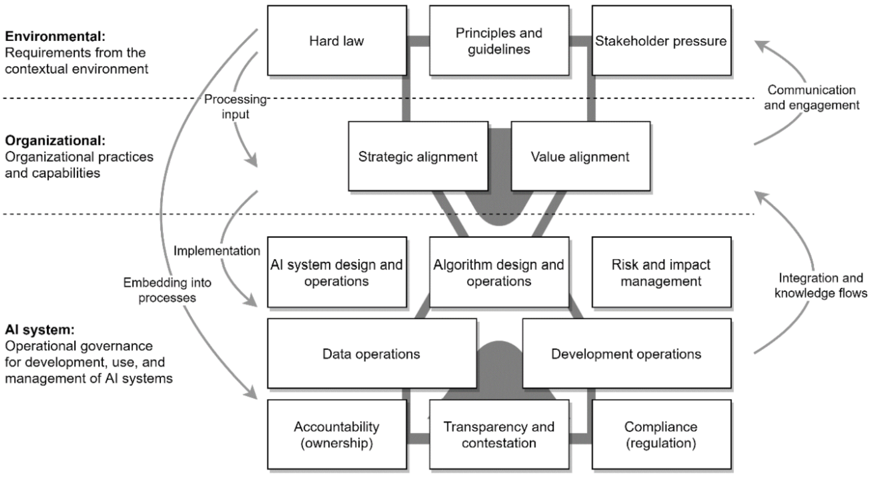
Environmental:
Requirements from the contextual environment
Organizational:
Organizational practices and capabilities
AI system:
Operational governance for development, use, and management of AI systems
Processing input
Implementation
Embedding into processes
Hard law
Principles and guidelines
Stakeholder pressure
Strategic alignment
Value alignment
AI system design and operations
Algorithm design and operations
Risk and impact management
Data operations
Development operations
Accountability (ownership)
Transparency and contestation
Compliance (regulation)
Communication and engagement
Integration and knowledge flows

Figure 1: The hourglass model of organizational AI governance (Mäntymäki, M., et al, 2022, Fig2, p.5)

Recommendations

- Increase the scope of the definitions to establish the overall context of the fields.
- Provide examples of AI in practice for the public to be able to relate to.
- Establish AI governance as a field and clarify the place of regulations in that field.
- Ensure a process for defining AI is in place so that it can cope with new domains of AI that are not yet defined and will be evolving over time.

Potential gaps in approaches

2. *Potential risks from AI not covered by Australia's existing regulatory approaches – suggestions for action*

While Australia has a set of AI ethics principles in place, the proposed legislation does not encompass some of the elements found in the principles. Critically, considerations for the environmental dimension (Principle 1: human, societal and environmental wellbeing) are not included.

The environmental costs of developing AI solutions or training large datasets is considerable. For example, the environmental cost of training a large natural language processing model could be as much as the total emissions produced by five cars over the cars' lifetime[6] or 150 return flights between

[6] Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP*. 57th Annual Meeting of the Association for Computational Linguistics. https://arxiv.org/abs/1906.02243v1

New York and London[7]. Furthermore, the demand for energy to power the development, training and usage for AI-powered tools is growing exponentially and is not required to be come from renewable energy sources. Finally, the demand for graphic cards and other critical materials required for the infrastructure (i.e. computers, servers etc) onto which AI solutions are operating is pushing the world towards further depletion of resources. T

The environmental and social costs of mining and manufacturing the equipment, let alone disposing of it is not considered currently. It should however be mandated in the legislation that AI systems be environmentally friendly. Similarly, to discovering bias and asking AI developers to reconsider the appropriateness of deploying the system (p.8), they should also explain how environmentally-friendly the AI system is and be asked to reconsider if the predictions for environmental frugality are not met.

Legislation should provide a framework that will support Australia towards more sustainable ways of living and doing. For example, the EU legislation draft specifies that AI systems be safe, transparent, traceable, non-discriminatory and environmentally friendly[8]. More than productivity alone, Europe has initiated the Green Digital programme[9] to tackle the green transition and the digital transition together. This approach is critical if Australia wants to meet the challenges facing ours and future generations. Furthermore, it also creates opportunities for innovation developed in Australia which can then be exported globally.

In their book on Responsible AI, Kirshner, Vidgen and Wallace[10] propose a table of AI harms on page 58 which lists different categories of harm: well-being harms, economic harms, individual harms, institutional harms, environmental harms. The table is useful in identifying the realms that the regulation should include and therefore the regulatory boundaries around these realms. For example, under institutional harms, while the paper lists surveillance and weaponry as a risk, it does not clearly account for the democratic harms (AI influencing choice in elections) or the informational harms (AI usage in government reducing accountability and trust.

Immersive tech and the metaverse will generate new types of challenges because they enable more control of human behaviour than social media[10]. They could create serious harm and organisations behind it do not seem to make any progress on public safety[11].

In the paper, not enough emphasis is put on the need to provide users of AI-powered tools to have agency. Developers of AI systems should provide the means for users (individuals, governments and companies) to:

- be able to detect an AI is used (i.e. to influence their behaviour or provide advice, make assessments and decisions), especially recommendation systems.
- be informed of what AI systems do and how they influence decisions in plain language,

[7] Guardian (2019) *How your flight emits as much CO2 as many people do in a year*. Available at: www.theguar dian.com/environment/ng-interactive/2019/jul/19/carbon- calculator-how-taking-one-flight-emits-as-much-as-many- people-do-in-a-year, accessed 20 July 2023.
[8] European Parliament News (2023) *AI Act: a step closer to the first rules on artificial intelligence* https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence, accessed on 26 July 2023
[9] European Commission (2023) *Green Digital Sector*, https://digital-strategy.ec.europa.eu/en/policies/green-digital, accessed on the 26 July 2023
[10] Kirshner, S., Vidgen, R., Wallace, C. (2022) *Checkmate Humanity: The how and why of responsible AI*, Global Stories
[11] Milmo, D (2021) *Enter the metaverse: the digital future Mark Zuckerberg is steering us toward*, The Guardian, https://digital-strategy.ec.europa.eu/en/policies/green-digital, accessed on the 24 July 2023

- be able to opt-in or opt out of using these systems
- disclose or modify the criteria used to assess them
- have options to revert decisions that were made automatically
- have controls over how and who decisions made automatically are shared with
- have the possibility to be forgotten and trust that the information and decisions made about them in such systems can be deleted.

One of the big risks of AI is that it is adopted by large and rich entities who will further use it to remove competition from smaller organisations or individuals. Clear and powerful laws on unfair competition should be updated to reflect that risk and ensure that prosecutors have the right levels of awareness and literacy of the issues at hand to be able to prevent and remedy them.

Australia does not have the equivalent of the GDPR[12] legislation in place (including the right to be forgotten and other regulations) that could protect AI systems users and their data. Therefore, the policy framework needs to ensure these rights are covered across the AI regulatory system. For example, data governance is a critical element to include in the legislation. It should ensure data used to train models has been obtained legally without breaching copyright laws, that the data is representative of diversity and contexts.

The wildest claims are made about AI systems and their abilities to "place a claim here". To avoid false advertisement, clear guidelines on the environmental and social costs of AI throughout its lifecycle need to be available. For example, claims are made that AI will solve the water efficiency problem or the energy problem, or the food problem. All without context or data, or even an acknowledgement that we are talking here about complex systems therefore, one solution will probably not be the answer. Claims from solution provider need to be substantiated with life cycle assessments with clearly defined boundaries and include other aspects of problems such as social, cultural, societal issues or other environmental issues.

Major technology companies exert significant influence over the legislative process through the lobbying process[13] . This could exacerbate the power imbalances and social inequalities[14].

Recommendations

- Mandating that AI systems are environmentally-friendly
- Regulating digital and green transitions together
- Adopting a taxonomy of harms to help better understand and tailor the legislation to each type of risk.
- Include the metaverse and other immersive tech in the scope of the legislation
- Strong and updated competition laws to ensure big and rich companies do not use the technologies to unfairly eliminate smaller and diverse competitors
- Ensure GDPR and EU Data Act equivalent are adopted in Australia either as part of the Responsible AI legislation across other legislations
- Strengthen legislation around AI-washing
- Push for open-source content and traceable content

[12] European Commission (2016) *General Data Protection Regulation*, gdpr-info.eu, accessed 26 July 2023
[13] Cath, C. (2018). *Governing artificial intelligence: Ethical, legal and technical opportunities and challenges.*
[14] Jobin, A., Ienca, M., & Vayena, E. (2019). *The global landscape of ethics guidelines*, Nature Machine Intelligence.1(9), 389–399.

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

Governing AI systems through their design where social, legal, ethical rules can be enforced through code to regulate the behaviour of AI systems[15]. Where government considers privacy by design, data protection by design and safety by design, it should also consider sustainable by design and ethical by design.

Human in the loop approaches to governing AI extended to society in the loop approach where society is responsible for finding consensus on the values that should shape AI[16].

Providing companion practices, tools and resources for anyone interested in developing or implementing AI responsibly should be undertaken. Resources should be available to all levels of AI literacy and AI maturity, and available freely.

Leaving AI literacy to the private sector would be to the detriment of the public, small business and not for profit who need to be and feel involved in the direction AI can take. Presenting clear case studies, including mapping of AI responsible practices, could serve as examples of sector-specific innovation. Practices should guide the design, use and deployment of AI-powered and automated systems in line with AI ethics principles and with the legislation.

AI governance is only partially based on legislation. Frameworks and models should be developed for Australia and communicated effectively. Business due diligence, approaches to de-risking AI decision-making and audits could all be part of an overall approach.

Several professional associations have strong code of ethics and code of conducts. While this is being explored by some parts of the engineering profession, it has not been adopted. With the implication and potential harm that could be done by AI systems, the government should enter discussions with the AI development profession to at least consider implementing a “do no harm” approach.

AI ethics, responsible AI and similar subjects should be taught in schools and universities. Failing to tackle the subject before professionals enter the workforce means that organisations will not have trained staff capable of understanding and adapting to what a Responsible AI world should be. They will not have a clear understanding of the risk mitigation strategies they should put in place until they are caught in it.

Recommendations

- Provide responsible AI tools, practices and resources freely to the public
- Sponsor human-in-the-loop by design approach
- Sponsor AI literacy programmes across sectors
- Develop databases of responsible AI-powered projects across sectors to inform about risks and to present available and innovative opportunities.
- Provide clear guidelines on AI governance in practice

[15] Leenes, R., & Lucivero, F. (2014). *Laws on robots, laws by robots, laws in robots: Regulating robot behaviour by design*. Law, Innovation and Technology, 6(2), 193–220.

[16] Rahwan, I. (2018). *Society-in-the-loop: Programming the algorithmic social contract. Ethics and Information Technology*, 20(1), 5–14.

- Facilitate conversations on the adoption of a code of ethics and code of conducts with engineering and computer science professions, and other AI development or design professions.
- Sponsor the teaching of Responsible AI and AI ethics in schools and universities

4. *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.*

n/a

Responses suitable for Australia

5. *Governance measures being taken or considered by other countries that are relevant, adaptable and desirable for Australia*

The following measures should be in place in Australia's context:

- In the UK example, principles are legislated to create a statutory duty to have due regard to the principles. Guidelines could be adhered to and can evolve more rapidly than legislation. Regulation to implement the guidelines and enforcement on it will be most likely able to cope with the pace of change[17]
- In the UK, the Algorithmic Transparency Standard publishes reports on the use of algorithmic tools in government decision making.
- In China, a mandatory registration system is in place for recommendations algorithm. Whilst the goal for the registration will differ with Australia, China has rightly identified that recommendation engines have a particularly powerful capacity to influence behaviours without the consent or awareness of the systems users.
- Italy forces generative AI platforms to allow users data not to be used in training, to disclose how they process user data, to allow users to opt-out.
- The city of Amsterdam host a searchable public AI register on the algorithmic systems that it uses
- The EU has other legislation in place to scaffold the EU AI act that we should ensure is in place in Australia or that the key intent for the legislation is covered in other parts of the legislation. This includes the GDPR, the Digital Services Act, the Digital Markets Act, the Digital Governance Act and the Unfair Commercial Practices Act.

Additionally, the need for a different model of governance is arising in a context where flexibility and adaptiveness are so critical. Adaptive governance or hybrid governance models move away from a command-and-control model of governance and instead looks to emphasize the role of non-state actors and the need for ongoing assessment of the balance of power.

To enable more flexible approaches and iterative adjustments and improvement of regulations and policies as new information is gathered from the public, a combination of industry standards and public

[17] Taeihagh, A. (2021) *Governance of artificial intelligence*, https://academic.oup.com/policyandsociety/article/40/2/137/6509315 , Policy and Society, accessed on the 26/06/2023

regulatory oversight is required. It will also be able to better suited to adapt to varied approaches to ethics based on different contexts and cultures. For example, IEEE published an ethically-aligned design guide for autonomous and intelligent systems[18] in 2020 under the Creative Commons which could be part of a hybrid public-private governance model.

Examples of adaptive governance include:

- laws that require regular risk assessments of the regulated activity,
- soft law approaches that involve collaboration with the affected stakeholders to develop guidelines, and legal experimentation and
- regulatory sandboxes to test innovative frameworks for liability and accountability for AI that will be adapted in iterative phases

Recommendations

- Have a statutory duty to have due regard to the AI ethics principles
- Develop algorithms databases for models used in government settings and provide public access to them
- Recognise recommendations engines for the potential damage they could cause to individuals and society
- Embed user agency by design
- Ensure principles such as the right to be forgotten and other user agency principles are embedded in the legislation
- Principles legislated to create a statutory duty to have due regard to them.

Target areas

6. Applying different approaches to public and private sector use of AI technologies

The legislation will need to be tailored towards not just products (a critique of the proposed EU AI Act[19]) but also to the open-source community (researchers, developers, not-for-profits).

Inclusivity and representation are particularly important for governments who need to represent all their constituents. A higher level of accountability is therefore expected of government departments. Updated and up-to-date approaches to data governance and to interoperability[20] are expected for governments which might not be required in the private sector (outside of regulated industries).

Data governance, a critical area: Part of AI governance, as multiple organisational and technological challenges exist that impede effective control over data and attribution of responsibility for data-driven decisions made by AI systems[21].

[18] IEEE. (2020). *A call to action for businesses using AI – ethically aligned design for business*, https://standards.ieee.org/wp-content/uploads/import/documents/other/ead/ead-for-business.pdf?

[19] McKinley, S. (2023) *The EU Act can get democratic control of artificial intelligence -but only if open-source developers get a seat at the table*, https://fortune.com/2023/07/17/eu-ai-act-democratic-control-artificial-intelligence-open-source-developers-tech-politics-shelley-mckinley/, accessed on the 26 July 2023

[20] Taeihagh, A. (2021) *Governance of artificial intelligence*, https://academic.oup.com/policyandsociety/article/40/2/137/6509315, Policy and Society, accessed on the 26/06/2023

[21] Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). *Data governance: Organising data for trustworthy artificial intelligence*, Government Information Quarterly, 37(3), 101493.

Interoperability, a critical area: Data fragmentation and lack of interoperability between systems limits an organisation's control over data flows throughout its entire life cycle and shared roles between different parties in data sharing clouds the chain of accountability and causation between AI-driven decisions/events and the parties involved in facilitating that decision[22].

Government should use transparency by design approaches and open-source models as defaults.

Recommendations

- Legislation not just for products but also to the open-source community
- Higher expectations of inclusivity and representation in government
- Data governance legislation required
- Interoperability legislation required
- Open-source as default for government

7. *Australian Government further support to responsible AI practices in its own agencies*

Codesign and participatory design of solutions ensure that the complexity and prioritisation of problems is handled by the people who know best and are directly involved. Only then should a range of solutions be considered. AI initiatives are often driven by solution providers rather than by the problems and the users themselves. This leads to solutions being implemented before anyone knows what the problem is. It also leads to a tendency to fit the organisation to a solution rather than the other way around. Human-centered design approaches should be the default for solution design.

There is an urgent need to develop and operationalise clear guidelines on what the potential trade-offs of a responsible AI solution are. These frameworks will help government departments and decision-makers to understand what they potentially impact with their proposed solution and realise if it is worth implementing (e.g. social versus financial, short term versus long term, etc). Ethical principles can come into conflict with each other; therefore policymakers will need to establish relative importance between them for the public to be compliant[22].

Providing examples of opportunities of Responsible AI applicable to different sectors would illustrate what is possible and help engage sectors on responsible AI innovation. Working with AI in collaboration with others create opportunities to learn what works, what does not. It creates a forum for conversations about what ethics are, and how they et stretched in designing, implementing and using AI-powered solutions. Learning what issues were discussed and how they were resolved will support others who can relate to similar problems.

Recommendations

- Codesign and participatory design of government solutions
- Guidelines on ethical trade-offs and prioritisation between principles to facilitate operationalisation
- Cross-sector and transdisciplinary approaches to real-life projects

[22] Piano, S. L. (2020). *Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward*, Humanities and Social Sciences Communications, 7(1), 1–7.

8. *In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better? Please provide some examples.*

n/a

9. *Transparency across the AI lifecycle*

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

Part of improving public trust and confidence in AI would be to ensure providers of AI products document their data sources, explain how they ensured bias minimisation, or how they are going to monitor and correct bias over time.

Further, the Australian government should ensure that location-specific and inclusive datasets are used for models used by government agencies.

Finally, government data from employees' use of the tools could end up in foreign databases and be used for training of decision-making tools. Therefore, data governance framework should be in place and records management strictly enforced to protect Australian government and private data.

Recommendations

- Enforce documentation of data sources bias minimisation strategies
- Ensure location-specific datasets are used for decision-making in Australia
- Enforce records management policy on data storage and capture

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

For transparency along the value chain, the environmental costs of AI[23] and the human cost of training models, often akin to modern slavery[24][25] should be taken into consideration.

Consider sandbox development environments where governance and transparency processes are evaluated. This would allow innovation to take place while supporting the public trust in AI technologies because they have been evaluated and accredited.

Publish transparency reports on content moderations decisions and algorithms used (p.17 of paper), adopt a similar approach to the UK Algorithmic Transparency Standard.

[23] Gordon, A., Jafari, A., Higgs, C. (023) The hidden cost of the AI boom: social and environmental exploitation, https://theconversation.com/the-hidden-cost-of-the-ai-boom-social-and-environmental-exploitation-208669, The Conversation, 29 July 2023 , accessed on the 26 July 2023

[24] Perrigo, B. (2023) OpenAI used Kenyan workers on less than $2 per hour to make ChatGPT less toxic, https://theconversation.com/the-hidden-cost-of-the-ai-boom-social-and-environmental-exploitation-208669, Time, accessed on the 26 July 2023

[25] Perrigo, B. (2022) Inside Facebook's African sweatshop, https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/, Time, 14 February 2022, accessed on the 26 July 2023

Recommendations

- Value chain transparency include environmental and social cost of AI systems
- Provide sandbox environments for governance and transparency processes
- Publish transparency reports on content moderation decisions

10. Suggestions for:

Whether any high-risk AI applications or technologies should be banned completely?

Recommendations

Research and applications should potentially be regulated differently so that we have:

- a knowledge-based on how to counter high-risk AI
- clear guidelines for is allowed to be deployed in society
- open innovation which is discussed with the public (people who affect and affected by it)

In terms of high-risk AI applications or technologies, the following should be banned:

- Social scoring
- Facial recognition
- Deepfake

Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?

The main problem is that something might be accepted today that should be banned tomorrow and we will not know until it is too late. As we learn about it, the conversation should continue to unban old technologies that we understand or can regulate better or ban new technologies where new risks have appeared.

Power structure is critical in this. Transparency on who decides the criteria, assesses the requirements and what are their interests and biases is key to keeping the trust. A public-led discussion is necessary, and should not be conducted behind closed doors with a handful of big tech or consultancy companies.

Recommendations

- Iterative process to ban and unban AI systems and applications as we learn more about them
- Open the process to the public

11. Initiatives or government action that can increase public trust in AI deployment to encourage more people to use AI

Seeing that government is actively engaged with Responsible AI will provide some certainty to the public that it is and remains in touch with the evolving market. More than just regulations, the government should be engaged in a set of efforts and initiatives to continuously better its understanding, adjust the legislation and provide updated guidance as new systems and risks unfold.

AI literacy is a society-wide area that will need to be driven both by governments and by the private sector. To ensure the latest science is also commercially viable and that learning together about what Responsible AI will require government, academia and the private sector to work together on real-life problems. Public-private partnerships and academia-private research is difficult, often cumbersome, and expensive. Infrastructure, initiatives and funding could make this cross-sector learning framework easier to access and would help share knowledge and practical outcomes.

The Australian Government should incentivise innovation for social good and for transformative impact[26] to address complex societal needs and advance Australia sustainable future.

Systems thinking, transdisciplinarity and codesign are key set of skills and approaches which are required of ethical and responsible AI. These skills are necessary to ensure bias is recognised, solutions serve the people who affect and are affected by AI. The design of solutions which align to the AI ethics principles already laid out in Australia[27][28] will depend on these skills being available and taught across the curriculum. The government needs to recognise these as a priority and ensure education and communication programmes support these approaches.

Recommendations

- Engage Government with responsible AI efforts and initiatives to also "be seen" using, thinking, validating AI systems and learning about Responsible AI approaches
- Drive AI literacy programmes across sectors
- Sponsor cross-sector real-life projects with academia, government and private companies on Responsible AI projects
- Support systems thinking, transdisciplinary and codesign agendas across academia

Implications and infrastructure

12. *How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?*

n/a

13. *What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?*

As mentioned earlier, AI sandboxes would help provide an environment where innovation can happen within the regulatory framework. It also can provide some assurances to the public that AI systems were tested and compliant.

[26] World Economic Forum (2023) *The Presidio Recommendations on Responsible Generative AI*, June 2023, https://www3.weforum.org/docs/WEF_Presidio_Recommendations_on_Responsible_Generative_AI_2023.pdf, accessed on the 26 July 2023

[27] Camaréna, S. (2022). *Artificial intelligence in food system redesign – Designing for the benefit of the whole*, RMIT University.

[28] Camaréna, S. (2021). *Engaging with artificial intelligence (AI) with a bottom-up approach for the purpose of sustainability: Victorian Farmers' Markets Association , Melbourne Australia*, Sustainability, 13(16). https://doi.org/https://doi.org/10.3390/su13169314

Risk-based approaches

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

A risk-based approach is the most urgent piece of legislation that should be put in place to get some type of controls on the technologies and its developers and deployers.

However, this needs to be done with a view to implement an iterative way of reviewing and adapting the legislation as the technologies evolve. It should also not deter government from guiding the type of society we want to have and the way in which these technologies will support the vision for a fair, sustainable and healthy society.

Iteration and adaptiveness, along with intent will call on an innovation-led approach to legislation.

Recommendations

- Urgently implement risk-based approaches to AI in Australia
- Implement adaptive and hybrid governance models in tandem to ensure the legislation can keep pace with the changes in technologies and applications

15. Main limitations of a risk-based approach

Risk management does nothing about power imbalances. It manages the harms without addressing power imbalance, implying that even if people do not want it, the activity can happen as long as risk are mitigated. It could all be a tick box exercise[29].

Developers and users of AI have no obligation to be beneficial to or promote the interests of people affected by AI or the broader national and global challenges we face therefore missing out on what AI could do for us.

Recommendations

- Promote sponsor and legislate an agenda which has a clear purpose for AI to be beneficial and in the interests of people

16. Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?

n/a

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

Yes

[29] Kirshner, S., Vidgen, R., Wallace, C. (2022) *Checkmate Humanity: the how and why of responsible AI*

18. *How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?*

n/a

19. *How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?*

n/a

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:*

public or private organisations or both?

n/a

developers or deployers or both?

n/a

About the Author

Throughout her career, Stéphanie Camaréna PhD has focused on ways of designing with people to ensure technology was an enabler rather than a barrier to sustainable futures. This relies on making the societal, environmental and financial trade-offs visible for better decision-making.

Stéphanie has developed the "Bottom-Up Engagement Framework" (BUE) which allows and empowers people in organisations to leverage existing iterative practices (i.e. Agile, Lean, others) as a way to engage with AI for sustainability and ethics for the benefit of the whole. Participatory design and participatory futures play an important role to how Source Transition engages with people at work.

A systems thinker, Stéphanie embraces an approach which focuses on deep leverage points, the places where "a small shift in one thing can produce big changes in everything".

About Source Transitions

Source Transitions is a for-purpose management consulting group dedicated to supporting companies, organisations and government entities in the detection, design, shaping and delivery of differentiated perspectives for radical transitions to sustainable systems using AI.

We enable this shift in perspective through the delivery of our services and pro-bono work dedicated to supporting communities and Not for Profit (NFP) organisations.

Source Transitions takes a human-centered approach to developing strategies using design thinking and peer-reviewed research. By putting people first, we enable technological implementation to reflect the changing world and ethical standards. Our management consulting experts assist our clients transform faster to a more sustainable way of doing things.