The Australia Institute
Research that matters.
Centre for Responsible Technology

No 'Responsible AI' without transparency and accountability

Submission to the Consultation on Safe and responsible AI in Australia

Jordan Guiao
Research Fellow, Centre for Responsible Technology

July 2023

About The Australia Institute

The Australia Institute is an independent public policy think tank based in Canberra. It is funded by donations from philanthropic trusts and individuals and commissioned research. We barrack for ideas, not political parties or candidates. Since its launch in 1994, the Institute has carried out highly influential research on a broad range of economic, social and environmental issues.

About the Centre for Responsible Technology

The Australia Institute established the Centre for Responsible Technology to give people greater influence over the way technology is rapidly changing our world. The Centre will collaborate with academics, activists, civil society and businesses to shape policy and practice around network technology by raising public awareness about the broader impacts and implications of data-driven change and advocating policies that promote the common good.

Our philosophy

As we begin the 21st century, new dilemmas confront our society and our planet. Unprecedented levels of consumption co-exist with extreme poverty. Through new technology we are more connected than we have ever been, yet civic engagement is declining. Environmental neglect continues despite heightened ecological awareness. A better balance is urgently needed.

The Australia Institute's directors, staff and supporters represent a broad range of views and priorities. What unites us is a belief that through a combination of research and creativity we can promote new solutions and ways of thinking.

Our purpose – 'Research that matters'

The Institute publishes research that contributes to a more just, sustainable and peaceful society. Our goal is to gather, interpret and communicate evidence in order to both diagnose the problems we face and propose new solutions to tackle them.

The Institute is wholly independent and not affiliated with any other organisation. Donations to its Research Fund are tax deductible for the donor. Anyone wishing to donate can do so via the website at https://www.tai.org.au or by calling the Institute on 02 6130 0530. Our secure and user-friendly website allows donors to make either one-off or regular monthly donations and we encourage everyone who can to donate in this way as it assists our research in the most significant manner.

Level 1, Endeavour House, 1 Franklin St
Canberra, ACT 2601
Tel: (02) 61300530
Email: mail@australiainstitute.org.au
Website: www.australiainstitute.org.au
ISSN: 1836-9014

Summary

The Australia Institute's Centre for Responsible Technology welcomes the opportunity to submit to the Government consultation on Safe and responsible AI (artificial intelligence) in Australia.

With the advent of generative AI, the public has been made aware of the potentially transformative capabilities of this generation of AI technologies, which could impact many aspects of Australian society.

This has resulted in a race where the largest technology companies try to outpace each other to dominate AI technologies, as governments attempt to develop governance frameworks to maximise opportunities while minimising risks.

Numerous ethical frameworks, safety manifestos and pledges have been developed, but these need to be supported by regulation, effective accountability measures and meaningful oversight.

To make AI safer and more responsible, the Australia Institute recommends:

- The Australian Government require transparency from AI product owners and model owners (namely technology companies like OpenAI, Google, Meta and Amazon) on what datasets are used to train their AI, and where they source these data sets from.
- The Australian Government adopt strong data privacy protections, and explicitly include AI technologies in the *Australian Privacy Act*, ensuring effective consent, data minimisation and purpose limitations.
- Compensating copyright holders and owners of any data used to train AI technologies – including authors, artists, musicians, writers, journalists, programmers, and any other original copyright holders.
- A system of accountability for privately funded AI initiatives, including any research bodies, practices, training hubs, and networks which have been funded by technology companies with vested interests in a positive return for their own organisations.
- After developing a risk register for AI, the Australian Government should impose a moratorium on what are identified as the most harmful applications of AI.

Introduction

While AI technologies have been in development for many years, the creation of mass market applications for generative AI like ChatGPT, Stable Diffusion, and DALL-E has brought AI into the public eye. Generative AI allows users to create new content and images based on simple inputs, a much more accessible version of general AI technologies.

Now a race has developed between the largest technology companies to try and outmanoeuvre each other to dominate the burgeoning consumer AI market. At the same time, governments are also attempting to ensure a safe environment for AI technologies, having learnt in recent years of the many harms that unregulated digital technologies can produce – not least in the areas of privacy, hate speech, disinformation, and predatory/monopolistic behaviours from the largest companies.

The future of AI is uncertain. Predictions range from AI causing the end of life on Earth to forecasts of tremendous economic growth or even ending scarcity altogether. These extremes are little more than hype, but given these conflicting narratives, it is not surprising that governments have so far struggled to arrive at useful AI policy.

A consultation on safe and responsible AI is a welcome mechanism that can assess the current landscape and weigh up risks and benefits.

For technology that has the potential to be so consequential for Australian society, there needs to be more regulation and oversight which enforces transparency, data privacy, fair use and accountability, and a hardline on the most dangerous applications. Strong regulation is essential for the most prominent AI products, and oversight for the technology companies developing the most prominent AI products that have been and will be released in future.

Response to the Discussion Paper

TRANSPARENCY

The largest AI products are privately-owned black boxes – neither regulators nor the public have a clear understanding of how they work and what data they are based on.

Current AI products are often:

- Trained using public posts online, with the public not being aware that anything they post can be used to train proprietary AI.
- Trained using a corpus (a large body of writing) which has an original author, or group of authors, with those authors having not given consent to their works being used in this way.
- Trained using data that was intended for a different purpose, and not for training AI. For example, Google quietly updated their terms and conditions to determine that any data captured by Google products can now be used to train Google's proprietary AI.[1] This is of particular concern given Google's large and often monopolistic market share in search, mail, advertising and more.
- Trained using data that is assumed to be private, like financial data, health data, household data, potentially sensitive data around personal identity, such as sexuality, religion, socio-economic status, etc.

Mandating transparency in data sets used to train AI goes to the heart of safe and responsible use. Without these large data sets, AI products cannot be developed. Transparency around AI dataset training would clarify whether creators' consent was sought and whether the data included was private or permissible, and ultimately whether the data sets were sources and used responsibly.

DATA PRIVACY

Beyond data transparency, there is growing scrutiny around responsible use and sourcing of data. Technology companies like Google have faced lawsuits and breaches of data privacy laws,[2] and legislation around data privacy are being updated to account for the latest digital technologies. This includes the *Australian Privacy Act*, which is currently being reviewed.[3]

[1] Weatherbed (2023), *Google confirms its training Bard on scraped web data, too,* https://www.theverge.com/2023/7/5/23784257/google-ai-bard-privacy-policy-train-web-scraping

[2] Allyn (2022), *Google pays nearly $392 million to settle sweeping location-tracking case, https*://www.npr.org/2022/11/14/1136521305/google-settlement-location-tracking-data-privacy

[3] Australian Government Attorney General's Department (2023), *Privacy Act Review Report,* https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report

Data privacy regulations could assist in supporting members of the public in ensuring their personal information/personal data is or isn't used for training AI systems in a responsible manner.

AI systems and technologies should be included as part of the scope of the *Australian Privacy Act*. While 'personal information' is covered extensively in the review of the Australian Privacy Act, AI applications should be made explicit as part of the scope of the Act.[4]

Privacy principles that protect user personal data/personal information should be applied, including, but not limited to:

- Data minimisation – ensuring only the minimum amount of data is used for the effective operation of AI products.
- Purpose limitations – ensuring data collected is only used for the purposes of the original request – for example, when creating a Gmail account, most people would not foresee that their emails would be used to train Google's AI products.
- Consent – many data sets have been used in a way not intended by the data owner/originator.

Applying privacy principles to current AI products will likely expose how current AI practices have not considered user protection and user privacy.

COMPENSATING ORIGINAL AUTHORS

AI products are often trained using large data sets scraped from public sources. These data sets can include original works with known authors, either individuals or organisations.

For example, OpenAI, the developer of Chat GPT, sourced data from BookCorpus (a repository of unpublished book manuscripts), many news websites, online forums and more.[5] The company is facing a class action that alleges it violated the rights of millions of users by scraping their data from the public internet,[6] as well as a separate lawsuit from authors that claim the company breached copyright on their works.[7]

[4] Australian Government Attorney General's Department (2023), *Privacy Act Review Report,* https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report

[5] Hughes (2023), *ChatGPT: Everything you need to know about OpenAI's GPT-4 tool,* https://www.sciencefocus.com/future-technology/gpt-3/

[6] De Vynck (2023), *ChatGPT maker OpenAI faces a lawsuit over how it used people's data,* https://www.washingtonpost.com/technology/2023/06/28/openai-chatgpt-lawsuit-class-action/

[7] Creamer (2023), *Authors file a lawsuit over OpenAI for unlawfully 'ingesting' their books,* https://www.theguardian.com/books/2023/jul/05/authors-file-a-lawsuit-against-openai-for-unlawfully-ingesting-their-books

Given that AI companies are generating revenue and profit from their products, there should be an investigation about what data they have scraped to develop their products, the terms under which it was used, and the effect that this may have on creators.

There are also early discussions between AI companies and some publishers for licensing of their news content for training AI models.[8] Some form of licensing or compensation to news publishers should be considered as part of AI regulation given AI products also scrape a large number of news websites. These discussions are currently ad hoc in nature. While these are encouraging developments, a stronger mandate should apply to all owners of large AI systems, and not just those who voluntarily enter into agreements.

Many creative professionals are concerned about how AI might be used to exploit their creative rights, as well as used to undermine the value of their labour and output. There are already several AI products which use creative works to generate their own versions of images and music. Currently use of AI for film and TV productions are in the spotlight as the Screen Actors Guild-American Federation of Television and Radio Artists (more commonly known as SAG-AFTRA) and the Writers Guild of America go on high-profile strikes to demand fair use, compensation and protections against AI.[9]

ACCOUNTABILITY

Given the resources involved in developing the most cutting edge AI technologies, it is no surprise that only the largest technology companies – like Google, Amazon, Meta and OpenAI – have been able to develop successful products.

Governments may not necessarily be able to match the funding and investments that come from the largest technology companies like Google. Therefore large funding initiatives from technology companies may be welcome. However, at a minimum there should be transparency and accountability on any related government or non-profit initiatives that receive funding from these technologies, given they have vested interests in promoting their own businesses.

For example, one of the most visible AI initiatives in Australia at the moment is the National AI Centre, run through the CSIRO. The Centre is actively building AI networks, promoting governance models and training workshops across the country. The National AI Centre is

[8] O'Brien (2023), *ChatGPT-maker OpenAI signs deal with AP to license news stories,* https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a

[9] Chmielewski (2023), *Actors decry 'existential crisis' over AI-generated 'synthetic' actors,* https://www.reuters.com/technology/actors-decry-existential-crisis-over-ai-generated-synthetic-actors-2023-07-21/

part of the Digital Future Initiative funded by Google, a $AU1billion investment in Australian research, infrastructure and partnerships over 5 years.[10]

There should be a system of accountability which reviews the Centre's output and independent oversight over their research and projects. Google has historically punished research that calls into question its existing AI models. For example, Google fired prominent AI researcher Timnit Gebru, whose research highlighted the inequality built into large AI models. Gebru previously lead Google's AI ethics team.[11]

Given the largest technology companies are in a race to develop proprietary AI products, their funding initiatives should be scrutinised and seen as an extension of their business objectives.

MORATORIUM ON THE MOST HARMFUL APPLICATIONS

The Australian Government should develop a risk register that outlines and clarifies the most harmful applications that AI can be used for, and place a moratorium on those applications.

While the doomsday scenarios promoted by some can appear far-fetched, there are some very plausible highly dangerous scenarios that AI could be used for.

These could include:

- Widespread political manipulation using AI features like deepfakes
- Social scoring systems to discriminate and target against specific groups such as those used in repressive regimes
- Automation on military systems and weaponry

While it is important to develop level-headed regulatory initiatives with practical applications, it is also important not to be naïve to the potential harms of this transformative technology.

[10] Australian Government Australian Trade and Investment Commission (2021), *Google invests A$1 billion in Australian infrastructure, research and partnerships,* https://www.austrade.gov.au/international/invest/investor-updates/google-invests-a-1-billion-in-australian-infrastructure-research-and-partnerships#:~:text=26%20Nov%202021,investment%20in%20Australia%20to%20date.

[11] Harris (2023), '*There was all sorts of toxic behaviour': Timnit Gebru on her sacking by Google, AI's dangers and big tech's biases,* https://www.theguardian.com/lifeandstyle/2023/may/22/there-was-all-sorts-of-toxic-behaviour-timnit-gebru-on-her-sacking-by-google-ais-dangers-and-big-techs-biases

Conclusion

The role of government is to ensure the safety of its citizens. This cannot happen while the largest AI products remain secretive black boxes, run by Big Tech companies that are already too powerful.

The Australian Government should pursue regulation that enforces transparency and accountability for these AI technologies and the companies that are developing them.