

THE CENTER FOR AI AND DIGITAL POLICY (CAIDP)
Comments to the
AUSTRALIAN GOVERNMENT
DEPARTMENT OF INFRASTRUCTURE, TRANSPORT, REGIONAL
DEVELOPMENT, COMMUNICATION and the ARTS
for the
STATUTORY REVIEW of the ONLINE SAFETY ACT 2021

The Center for AI and Digital Policy (CAIDP) welcomes the opportunity to respond to the request for comments from the Australian Department of Infrastructure, Transport, Regional Development, Communication and the Arts for the Statutory Review of the Online Safety Act 2021.¹ As the consultation’s *Issues Paper* noted, “Recent advancements in technologies such as generative artificial intelligence [AI] are changing our online experiences, and can also be used to generate or amplify illegal and harmful content.”² Indeed, the online risks which accompany the advent of generative AI are expansive, and include threats to personal private data, to intellectual property, and include the potential for life-altering outcomes based on online interactions with AI-enabled decision-making models. For these reasons, CAIDP is eager to contribute to this consultation regarding our expertise.

About CAIDP

CAIDP is an independent, non-profit organization that advises national governments and international organizations on artificial intelligence and digital policy³. We advise the Global Partnership on AI, the Council of Europe, the European Union, UNESCO, the UN High-Level Advisory Body on Artificial Intelligence, the OECD and other international and national organizations, working with more than 800 AI policy experts in over 80 countries. CAIDP supports AI policies that advance democratic values and promote broad social inclusion based on fundamental human rights, democratic institutions, and the rule of law.⁴

In April, 2024, we released the third edition of our *Artificial Intelligence and Democratic Values Index (AIDV)*, providing a comprehensive review of AI strategies in 80 countries and their adherence to human rights.⁵ We noted therein that while Australia has established an admirable record developing guardrails to address a range of human rights concerns, the voluntary nature of Australia’s *AI Ethics Principles*⁶ is insufficient to match the challenges of guarding against civil and human rights violations presented by advancing and increasingly intrusive AI technologies.⁷ We observe as well that the Parliament of Australia recently released a report on the *Influence of international digital platforms*, concluding that,

¹ Australian Govt Dept of Infrastructure et al, *Statutory Review of the Online Safety Act 2021: Issues Paper* (April, 2024) <https://www.infrastructure.gov.au/departments/media/news/consultation-open-online-safety-act-review>

² Ibid.

³ CAIDP, <https://www.caidp.org/>

⁴ CAIDP, *Statements*, <https://www.caidp.org/statements/>

⁵ CAIDP, *Artificial Intelligence and Democratic Values* (April, 2024), <https://www.caidp.org/reports/aidv-2023/>

⁶ Australian Department of Industry, Science, Energy and Resources, *Australia’s AI Ethics Principles* (Nov, 2019), <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>

⁷ Ibid.

“the current regulatory system is not working effectively. Regulation of digital platforms is split across various agencies, in some cases with competing priorities.”⁸ As was recently noted in the Government’s interim response to its *Safe and responsible AI in Australia consultation*, current “laws do not sufficiently prevent harms from the deployment of AI systems in legitimate but high-risk contexts.”⁹ We would note, moreover, that it is often difficult to predict the technologies and contexts that ultimately present high-risk.¹⁰

Summary of CAIDP Recommendations

Our recommendations aim to mitigate safety, civil and human rights concerns in an era in which AI companies contravene established global privacy laws, and pertaining to online interactions with opaque AI/algorithmic models that have the potential to drastically affect users’ quality of life. In particular, we are responding to two questions within the reference document, Part 5, question 26, “*Are additional safeguards needed to ensure the Act upholds fundamental human rights and supporting principles?*” and Part 6, question 28, “*What considerations are important in balancing innovation, privacy, security and safety?*”

We recommend that the Act to be updated to place **affirmative duties** on companies developing and deploying AI online to :

- 1) Establish **redlines** for developers, providers, and deployers of AI systems regarding training data, prohibiting practices which contravene the **Australian Privacy Principles**¹¹ including the **non-consensual web-scraping** of personal data and intellectual property.¹²
- 2) Require **transparent** and **contestable data provenance** regarding AI models trained on web-scraped data so that data subjects may be made aware when their personal, private data and intellectual property has been used to train AI models, providing an opportunity for compensation and extrication of data.¹³

⁸ Parliament of Australia, *Influence of international digital platforms: The way forward* (Nov. 2023), https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_1_0_-_The_way_forward

⁹ Australian Dept of Industry, Science and Resources, *Safe and responsible AI in Australia consultation* (Jan. 17, 2024), <https://www.industry.gov.au/news/australian-governments-interim-response-safe-and-responsible-ai-consultation>

¹⁰ CAIDP, *Comment of The Center for AI and Digital Policy (CAIDP) to the National Institute of Standards and Technology in response to Request for Information Related to NIST’s Assignments Under Sections 4.1, 4.5 and 11 of Executive Order Concerning Artificial Intelligence* (Feb. 2, 2024), <https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence/comments#C>

¹¹ Office of the Australian Information Commissioner, *The Australian Privacy Principles* (Dec. 12, 2012), <https://www.oaic.gov.au/privacy/australian-privacy-principles>

¹² Parliament of Australia, *Influence of international digital platforms: Data* (Nov. 2023), https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_5_-_Data

¹³ Parliament of Australia, *Influence of international digital platforms: Data* (Nov. 2023), https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_5_-_Data

- 3) Require rigorous, independent **impact assessments prior to deployment** to identify and mitigate potential online harms, including **biases** and **rights violations**, with ongoing re-assessments across the AI lifecycle.¹⁴
- 4) Require **algorithmic transparency** from models operating online such that users are aware when they are interacting with an AI/algorithmic system and are provided with **clear and valid reasons for outcomes** affecting their lives.¹⁵
- 5) Require human oversight and control over AI systems operating online and an **affirmative obligation to terminate** if human control of the system is no longer possible and/or if the system fails to uphold human/civil rights in keeping with the *Universal Guidelines for AI*,¹⁶ a precursor to the Australia-endorsed *UNESCO Recommendation on the Ethics of Artificial Intelligence*.¹⁷

Explanation of CAIDP Recommendations

Advancing AI technologies pose new risks to online safety and individual rights that Australia's *Online Safety Act 2021* in its current form does not adequately address. To uphold the rights established within *The Australian Privacy Principles* as well as those enshrined in international human rights law, the Act must be adapted to secure Australians' rights and agency in the face of AI systems that can track, profile, and profoundly affect their lives in opaque and unaccountable ways.

*Recommendation #1: Establish **redlines** for developers, providers, and deployers of AI systems regarding training data, prohibiting practices which contravene **Australian Privacy Principles**¹⁸ including the **non-consensual web-scraping** of personal data and intellectual property.¹⁹*

We commend the robust *Australian Privacy Principles* (APP), which establish mandatory requirements for the processing of personal data, requiring that individuals be informed about any collection of personal information and given the opportunity to consent or opt-out.²⁰ The web-scraping

¹⁴ CAIDP, *Comments to the National Institute of Science and Technology (NIST) Risk Management Framework Following White House Executive Order on AI*, (Feb. 2, 2024), https://www.caidp.org/app/download/8500506063/CAIDP%20Comments_NIST%20RFI_88%20FR%2088368_02022024.pdf?t=1717731063

¹⁵ Parliament of Australia, *Influence of international digital platforms: Algorithmic transparency* (Nov 2023), https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_6_-_Algorithmic_transparency

¹⁶ The Public Voice, *AI Universal Guidelines Memo*, (2018), <https://thepublicvoice.org/ai-universal-guidelines/memo/>

¹⁷ UNESCO, *Recommendations for the Ethics of Artificial Intelligence*, (Nov 2019), <https://en.unesco.org/about-us/legal-affairs/recommendation-ethics-artificial-intelligence>

¹⁸ Office of the Australian Information Commissioner, *The Australian Privacy Principles* (12 December, 2012), <https://www.oaic.gov.au/privacy/australian-privacy-principles>

¹⁹ Parliament of Australia, *Influence of international digital platforms: Data* (November, 2023), https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_5_-_Data

²⁰ Office of the Australian Information Commissioner, *The Australian Privacy Principles* (12 December, 2012), <https://www.oaic.gov.au/privacy/australian-privacy-principles>

practices which AI companies have attempted to normalize quite blatantly contravene these principles. As CAIDP highlighted in our complaint to the U.S. Federal Trade Commission (FTC) regarding OpenAI, gathering personal data and intellectual property without knowledge or consent represents a clear violation of established data protection norms.²¹ We noted further, in consultation to the UK Information Commissioner’s Office, that web-scraping of personal data and intellectual privacy without consent should be explicitly prohibited as these practices blatantly flout UK GDPR,²² which parallels the APP. In order for the APP to be upheld, the Online Safety Act must be amended to explicitly prohibit the non-consensual scraping of Australians’ personal data and intellectual property for AI training purposes.

*Recommendation #2: Require **transparent and contestable data provenance** regarding AI models trained on web-scraped data so that data subjects may be made aware when their personal, private data and intellectual property has been used to train AI models, providing an opportunity for compensation and extrication of data.*²³

The risks to Australians stemming from non-transparent data provenance for AI training data pertains not only to online privacy and the theft of intellectual property, but to the potential for **highly biased outcomes** which, given the nature of generative AI models, can be **self-reinforcing** at alarming and societally-damaging rates. In her testimony before the US Congress, CAIDP President Merve Hickok emphasized that “high-risk AI systems replicate existing biases in the datasets, as well as biases and choices of their developers, resulting in discriminatory decisions.”²⁴ To address this range of risks, the amended Online Safety Act should mandate that AI companies provide clear and accessible information about the data *that has been and is currently being* used to train their models, to empower data subjects to contest the inclusion of their personal data and intellectual property, and establish mechanisms for redress – including financial liability – when data has been used without consent or in a manner that produces biased or discriminatory outcomes.

*Recommendation #3: Require rigorous, independent **impact assessments prior to deployment** to identify and **mitigate potential online harms**, including **biases and rights violations**, with ongoing re-assessments across the AI lifecycle.*²⁵

²¹ CAIDP, In Re OpenAI, Supplement to Original Complaint, (Jul. 10, 2023),

<https://www.caidp.org/app/download/8466615863/CAIDP-FTC-Supplement-OpenAI-07102023.pdf>

²² CAIDP, *Comments to the 1st ICO Consultation on Lawful Basis for Web Scraping to Train Gen AI Models*, (Mar. 2, 2024),

https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caidp-ico-ai-and-webscraping-mar-1-activity-7169349419979493376-wXus?utm_source=share&utm_medium=member_desktop

²³ Parliament of Australia, *Influence of international digital platforms: Data* (Nov. 2023),

[https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_5 - Data](https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_5_-_Data)

²⁴ Merve Hickok Testimony. *Advances in AI: Are We Ready For a Tech Revolution?* House Committee on Oversight and Accountability Subcommittee on Cybersecurity, Information Technology, and Government Innovation (March 8, 2023), p. 3, https://oversight.house.gov/wp-content/uploads/2023/03/Merve-Hickok_testimony_March-8th-2023.pdf

²⁵ CAIDP, *Comments to the National Institute of Science and Technology (NIST) Risk Management Framework Following White House Executive Order on AI*, (Feb. 2, 2024),

https://www.caidp.org/app/download/8500506063/CAIDP%20Comments_NIST%20RFI_88%20FR%2088368_02022024.pdf?t=1717731063

Australia’s AI Ethics Principles state that AI systems “should be monitored and tested to ensure they continue to meet their intended purpose, and any identified problems should be addressed with ongoing risk management as appropriate.”²⁶ It is critical that such testing is done independently of the company that will be profiting from the release, and that initial evaluations precede deployment to the public. Currently, Australia’s voluntary framework is out-of-step with the regulatory requirements suggested within the Australia-endorsed UNESCO Recommendation on the Ethics of AI, which urged member states to “develop due diligence and oversight mechanisms to identify, prevent, mitigate and account for how they address the impact of AI systems...in particular, enforceable transparency protocols should be implemented...before releasing them in the market.”²⁷ Given the prevalent risk for online AI-driven models to perpetuate and amplify biases and discriminatory outcomes, and the evolving nature of these technologies accompanying novel risks, ongoing re-assessments throughout the AI lifecycle will be necessary to ensure they remain safe and rights-respecting.

*Recommendation #4: Require **algorithmic transparency** from models operating online such that users are aware when they are interacting with an AI/algorithmic system and are provided with **clear and valid reasons for outcomes** affecting their lives.*²⁸

Australia’s voluntary *AI Ethics Principles* state, “People should be informed when an algorithm is being used that impacts them, and they should be provided with information about what information the algorithm uses to make decisions.”²⁹ This proposition reflects a fundamental expectation that human decisions should be under human control to ensure safe and non-biased outcomes, and to ensure accountability. Yet an absence of legal guardrails has allowed for a potentially dystopian reality such that AI models are contributing to – and sometimes finalizing – decisions about healthcare, justice rulings, bank loans, and other deeply meaningful outcomes for individuals and communities based on opaque factors dependent on non-transparent training data and model-training which may be highly biased. To safeguard against far-reaching accompanying risks, the amended Online Safety Act should mandate algorithmic transparency, requiring online AI models to disclose themselves and to provide clear and accurate information about the factors which contribute to decisions affecting users’ lives.

*Recommendation #5: Require human oversight and control over AI systems operating online and an **affirmative obligation to terminate** if human control of the system is no longer possible and/or if the system fails to uphold human/civil rights in keeping with the*

²⁶ Australian Department of Industry, Science, Energy and Resources, *Australia’s AI Ethics Principles* (Nov. 2019), <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>

²⁷ UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, (Nov. 23, 2021) <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

²⁸ Parliament of Australia, *Influence of international digital platforms: Algorithmic transparency* (November, 2023), https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Digitalplatforms/Report/Chapter_6_-_Algorithmic_transparency

²⁹ Australian Department of Industry, Science, Energy and Resources, *Australia’s AI Ethics Principles* (Nov. 2019), <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>

Universal Guidelines for AI,³⁰ a precursor to the Australia-endorsed UNESCO Recommendation on the Ethics of Artificial Intelligence.³¹

None of the recommended guardrails will have meaning if AI systems which evade human control, fail to adhere to fundamental laws, and/or fail to uphold human rights are permitted to keep operating. CAIDP urges the amendment of the Online Safety Act such that AI models operating online must be terminated if core expectations regarding human control and adherence to civil/human rights are not upheld.

CAIDP thanks the Australian Department of Infrastructure, Transport, Regional Development, Communication and the Arts for the opportunity to contribute to its statutory review of the Online Safety Act 2021, and to recommend additional legal safeguards to ensure the Act upholds fundamental human rights, fairness and privacy principles in an era of rapidly advancing AI technologies. We hope you will consider our views, and look forward to further engagement with the Australian government regarding AI governance.

³⁰ The Public Voice, *AI Universal Guidelines Memo*, (2018), <https://thepublicvoice.org/ai-universal-guidelines/memo/>

³¹ UNESCO, *Recommendations for the Ethics of Artificial Intelligence*, (Nov. 2019), <https://en.unesco.org/about-us/legal-affairs/recommendation-ethics-artificial-intelligence>