

Open Access Australasia submission to the House of Representatives Standing Committee on Employment, Education and Training.

Concerning the Inquiry into the use of generative artificial intelligence in the Australian education system

July 12, 2023

Contact: Dr Virginia Barbour, Director,

Submitted on behalf of the [Open Access Australasia Executive Committee](#)

Open Access Australasia is a membership organisation of 23 Australian university libraries, eight Aotearoa New Zealand university libraries through the Council of New Zealand University Librarians, Creative Commons Australia, Tohatoha Aotearoa Commons, Australian Library and Information Association, Australian Digital Alliance, Wikimedia Australia, the Australian Citizen Science Association and National and State Libraries Australasia. Its mission is to attain open access to research in Australia and Aotearoa New Zealand through advocacy, collaboration, awareness, and capacity building across the Australian and New Zealand research sectors.

We acknowledge and support the Joint Submission from library and information related organisations to this inquiry made by Australian Library and Information Association including ALIA VET Libraries Australia, the Council of Australian University Librarians, National and State Libraries Australasia, CAVAL, AI4LAM, Open Access Australasia and the Australian Libraries and Archives Copyright Coalition.

Our interest in this Inquiry into the use of generative artificial intelligence (GAI) in the Australian education system is in relation to how these technologies could support, or hinder, the transition to openly available scholarly research that is currently underway in our region and internationally.

We note the following:

- Over the last few years, we have seen a major shift in research moving from closed to open access. Open access is defined as “free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.”¹ This openness must not be undermined by the risk of re-enclosure of information on private platforms.
- We acknowledge differences in concerns between those within the education system, including researchers, whose issues include the integrity of research practices and the commercial concerns of creators and artists.

Addressing Terms of Reference (1)

The strengths and benefits of generative AI tools for children, students, educators and systems and the ways in which they can be used to improve education outcomes.

We submit that AI/GAI can increase open access to scholarly research and data, that AI/GAI systems benefit from the existence of high-quality open information, and that many AI systems are built with open access in mind²:

- AI-assisted technologies can improve the discoverability and accessibility of resources - especially open access resources - through AI assisted search engines and automated metadata creation.
- AI-assisted technologies can improve the quality of open educational resources and support the synthesising of information for a non-technical audience.

Addressing Terms of Reference (3)

The risks and challenges presented by generative AI tools, including in ensuring their safe and ethical use and in promoting ongoing academic and research integrity.

- Copyright and licensing
 - GAI relies on copyright and text and data mining (TDM) provisions that are not fully incorporated into Australian copyright law at the present time.³
 - Many openly accessible publications make use of Creative Commons (CC) licences to set conditions for sharing. CC-BY (where creator attribution is required) licences are prevalent. While it is generally legal to use CC-BY licensed content for large language model (LLM) training purposes, there are ethical and integrity issues to consider, particularly around acknowledgement and recognition. There should always be flexible exceptions for genuine research purposes, but researchers, creators, and artists are rightly concerned about how and whether they are correctly acknowledged or compensated for commercial usage.⁴ Other CC licences with more restrictive terms, such as non-commercial (NC) or no derivatives (ND) must also be recognised as applying to open content.
 - Calls to amend copyright laws must balance the rights of public access against the interests of large online platforms.⁵
 - The use of personal data in LLMs presents unique individual privacy concerns. Data included in a massive automatically harvested training corpus cannot be separated and removed.
- Open source vs commercial AI/GAI systems
 - The importance of supporting open source systems which enable the code and model to be more easily interrogated, allowing for greater transparency.
 - Open Access Australasia shares concerns raised in the US and elsewhere about the risks to competition if power and usage of AI/GAI platforms and tools is concentrated among a handful of commercial players.⁶ This also has implications for information seeking if search engines keep users within their own platform and make it difficult to view the underlying source material.
 - The concern that commercial scholarly publishers will align with for-profit AI/GAI companies and produce systems and tools that effectively re-enclose publicly funded research.

- Lack of transparency in AI/GAI processes and systems
 - Many widely used LLM training datasets are not fully detailed or transparent.⁷ This obscurity is compounded by the use of dynamic datasets that are constantly changing. We do not know what data is being used or whether it is openly available or not. OpenAI has just disabled functionality that appeared to bypass some paywalls.⁸
- Implications for the integrity of research
 - Research assessment bodies, such as funders, are adopting policies prohibiting the use of GAI tools in applications in an effort to ensure academic integrity.⁹ There are fears that the proliferation of AI tools for academic research may result in decline in the quality of academic output.¹⁰
- Indigenous research and data
 - Indigenous peoples' control of their data, culture and language is threatened when data is scraped en masse from the web without knowledge and consent.¹¹
 - Indigenous data sovereignty is compromised if Indigenous research data are harvested and stored at locations remote to the owners and beyond their control.
 - The importance of full inclusion of Indigenous scholars in any consultation about the inclusion of Indigenous research and data in the training data of LLMs.

Addressing Terms of Reference (6)

Recommendations to manage the risks, seize the opportunities, and guide the potential development of generative AI tools including in the area of standards.

- We recommend that Australia embrace a considered and leading role in the region with regards to the responsible use of AI in higher education.
- We support a robust, inclusive approach to developing further recommendations and guidelines in this area - libraries and open access advocates can play a key role in this due to our provision of both content and information skills, and our experience in the complex landscape of open access.
- We believe that AI can make responsible reuse of open access scholarly research, but it is critical that licensing and acknowledgement issues around AI are clarified.
- We emphasise that the success of LLMs' contributions to higher education depends on the quality of their training data.
- Any national approach to open access will need to consider the implications for AI/GAI.

References

1. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Max Planck Society. October 22, 2003. Accessed July 13, 2023. <https://openaccess.mpg.de/Berlin-Declaration>
2. The Relationship Between Open Access Science and AI. March 20, 2023. Azorobotics. Accessed July 10, 2023 <https://www.azorobotics.com/Article.aspx?ArticleID=588>
3. Bradley, F. Representation of libraries in artificial intelligence regulations and implications for ethics and practice. J Aus Libr Inf Assoc. 2022;71(3) doi:10.1080/24750158.2022.2101911
4. Milmo D. Sarah Silverman sues OpenAI and Meta claiming AI training infringed copyright. The Guardian. July 10, 2023. Accessed July 12, 2023. <https://www.theguardian.com/technology/2023/jul/10/sarah-silverman-sues-openai-meta-copyright-infringement>
5. Google calls for relaxing of Australia's copyright laws so AI can mine websites for information. The Guardian. April 19, 2023. Accessed July 10, 2023 <https://www.theguardian.com/technology/2023/apr/19/google-calls-for-relaxing-of-australias-copyright-laws-so-ai-can-mine-websites-for-information>
6. Generative AI Raises Competition Concerns. Federal Trade Commission. June 29, 2023. Accessed July 10, 2023. <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns>
7. Venturebeat. Generative AI's secret sauce — data scraping— comes under attack. Cisco. July 6, 2023. Accessed July 10, 2023. <https://venturebeat.com/ai/generative-ai-secret-sauce-data-scraping-under-attack/>
8. @OpenAI. [tweet] July 4, 2023. Accessed July 10, 2023. <https://twitter.com/OpenAI/status/1676072388436594688>
9. Australian Research Council. Policy on Use of Generative Artificial Intelligence in the ARC's grants programs. July 7, 2023. Accessed July 12, 2023. <https://www.arc.gov.au/sites/default/files/2023-07/Policy%20on%20Use%20of%20Generative%20Artificial%20Intelligence%20in%20the%20ARCs%20grants%20programs%202023.pdf>
10. Humane Ingenuity 47: AI Is Coming for Scholarship Next. July 10, 2023. Accessed July 12, 2023. <https://newsletter.dancohen.org/archive/humane-ingenuity-47-ai-is-coming-for-scholarship/>
11. Chandran R. FEATURE-Indigenous groups in NZ, US fear colonisation as AI learns their languages. April 3, 2023. Accessed July 12, 2023. <https://www.reuters.com/article/newzealand-tech-lawmaking-idUSL8N2UQ0EC>