Science in Australia Gender Equity (SAGE)

Submission to TEQSA consultation:
Assessment reform for the age of artificial intelligence

20 October 2023

Overview

There are significant inequities in the global workforce producing and maintaining artificial intelligence (AI). Alongside other inequalities in the sector, AI is overwhelmingly male-dominated (only 22% of AI professionals are women). This workforce's susceptibility to bias poses a risk for education and the future of learning.

An adequate response to the challenge of assessment in the age of AI must therefore be holistic. We recommend considering not only the implications for student learning but also equipping university staff with the skills to identify, navigate and teach about these biases.

Recommendations

1. Assess students on their ability to recognise and minimise social biases and discrimination in AI.
2. Build staff capacity to identify and teach about equity, diversity and inclusion issues in AI.
3. Increase the representation of women and marginalised groups amongst AI researchers and developers.

Q1: What feedback do you have on the two principles and five propositions?

> **RECOMMENDATION 1**
>
> To ensure that students do not inadvertently reproduce or reinforce social biases when they use AI in their professional and civic lives, **TEQSA should expand Guiding Principle 1 and Proposition 1 to stipulate that an EDI lens should be applied to any assessment framework for the responsible and ethical use of AI.**

As SAGE's work focuses on workplace equity, diversity and inclusion (EDI) in higher education and research, we will only comment on **Guiding Principle 1** and **Proposition 1**.

Broadly speaking, we agree with treating AI like any tool or instrument by teaching students how it works, how and when to use it, and what its strengths and limitations are.

We note that Guiding Principle 1 and Proposition 1 encourage educators to teach and assess students on the "responsible and ethical use of AI". Generative AI poses many ethical challenges, such as misinformation, violent and sexualised content, and privacy and copyright infringement, to name a few.[1] Many of these stem from AI's susceptibility to bias.

There are many points at which bias could creep into AI models, from the data and algorithms with which they are trained, to the developers' design and policy decisions.[2] These biases are then reproduced in the outputs of these models. When given general prompts, AI text-to-image generators tend to produce images that reinforce whiteness, heteronormativity and American ideals.[3] AI generators can also amplify harmful stereotypes – for example, they have been shown to associate Muslims with violence,[4] portray people with disabilities as needing to be "fixed" or saved,[5] and suggest that women are less suited to leadership roles[6] or certain occupations.[7,8,9]

All of these EDI challenges have serious implications for the likely impacts of AI on learning and assessment processes. Although students may benefit from generative AI tools that can explain, summarise and critique content for them,[10] occasionally, these tools will confidently present users with false or misleading information as fact.[11] These errors can be difficult to verify without expertise in the relevant subject area, and as such they pose great potential risks for students who are not only developing their knowledge of course material, but are simultaneously forming social views. In an EDI context, this may confirm or even generate students' own biases instead of

[1] Bird C, Ungless E and Kasirzadeh A (2023) 'Typology of risks of generative text-to-image models', *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 396–410.

[2] Ferrara E (2023) 'Should ChatGPT be biased? Challenges and risks of bias in large language models', *ArXiv*, 2304.03738, accessed 13 October 2023.

[3] Bianchi F, Kalluri P, Durmus E, Ladhak F, Cheng M, Nozza D, Hashimoto T, Jurafsky D, Zou J and Caliskan A (2023) 'Easily accessible text-to-image generation amplifies demographic stereotypes at large scale', *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1493–1504.

[4] Abid A, Farooqi M and Zou J (2021) 'Large language models associate Muslims with violence', *Nature Machine Intelligence*, 3:461–463.

[5] Gadiraju V, Kane S, Dev S, Taylor A, Wang D, Denton E and Brewer R (2023) '"I wouldn't say offensive but...": disability-centred perspectives on large language models', *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 205–216.

[6] Newstead T, Eager B and Wilson S (2023) 'How AI can perpetuate – or help mitigate – gender bias in leadership', *Organizational Dynamics*, in press.

[7] Zhou M, Abhishek V and Srinivasan K (2023) 'Bias in generative AI', manuscript in preparation, accessed 13 October 2023.

[8] Thomas RJ and Thomson TJ (2023) 'What does a journalist look like? Visualising journalistic roles through AI', *Digital Journalism*.

[9] Sun L, Wei M, Sun Y, Suh YJ, Shen L and Yang S (2023) 'Smiling women pitching down: auditing representational and presentational gender bias in image generative AI', *ArXiv*, 2305.10566, accessed 13 October 2023.

[10] Rowland DR (2023) 'Two frameworks to guide discussions around levels of acceptable use of generative AI in student academic research and writing', *Journal of Academic Language & Learning*, 17(1):T31–T69.

[11] Alkaissi H and McFarlane SI (2023) 'Artificial hallucinations in ChatGPT: implications in scientific writing', *Cureus*, 15(2):e35179.

educating them, or it might lead students to adopt AI recommendations for accessibility and inclusion that are not in line with best practice.[12]

In practice, this means it is necessary to assess students on their ability to use AI in a way that does not unfairly discriminate against, or cause harm to, people of any gender, sexual orientation, race, ethnicity, age, disability, religion and so on.

For example, assessment tasks should encourage students to demonstrate:

- their understanding of how biases are introduced into AI systems, and how these shape AI outputs and decisions; and

- how they recognise and mitigate potential biases or social harms in any AI-generated or AI-modified content, whether it is made by themselves or others.

Q3: What do you think needs to happen next to support the required change in the sector and/or at your institution?

> **RECOMMENDATION 2**
>
> Universities should build staff capacity to identify emerging EDI issues in AI and to teach students about these in an accurate and culturally sensitive manner.

University staff and educators may not have the EDI knowledge and competencies required to understand the harmful biases generated and perpetuated through AI. Therefore, organisations need to invest in adequate training as part of staff development programs. Educators should be able to model the responsible use of generative AI technologies and set clear expectations for students to do the same.

> **RECOMMENDATION 3**
>
> The AI research sector (of which universities are a key part) should diversify their workforce to include more women and marginalised people, such as those who are queer, racialised or have a disability.

Today's AI workforce is overwhelmingly male, limiting its ability to identify and address the gendered risks of AI technologies. Only 22% of AI professionals globally are women.[13] The figure is similar in Australia, where women made up 23% of information technology graduates in 2021.[14]

---

[12] Glazko K, Yamagami M, Desai A, Avery Mack K, Potluri V, Xu X and Mankoff J (2023) 'An autoethnographic case study of generative artificial intelligence's utility for accessibility', *ArXiv*, 2308.09924, accessed 13 October 2023.

[13] World Economic Forum (2018) *Reader: global gender gap report 2018*, World Economic Forum, accessed 16 October 2023.

[14] Department of Industry, Science and Resources (2023) *University enrolment and completion in STEM and other fields*, STEM Equity Monitor, accessed 16 October 2023.

According to OECD.AI, only half of all academic publications in AI have at least one female author.[15] This stands in stark contrast with the 90% of AI publications that have at least one male co-author. 55% of AI publications have all-male authors, whereas only 11% are authored solely by women.

To design AI technologies that are safe and inclusive for everyone, diverse perspectives should be incorporated into AI education, research and development. This is explicitly relevant when developing assessment guidelines and criteria.

---

[15] OECD.AI (2023) *Live data: AI research*, OECD.AI Policy Observatory, accessed 16 October 2023.

OECD.AI's visualisations are based on Elsevier (Scopus) data. Elsevier uses an algorithm to infer an author's gender based on their first name, last name and country of origin. Authors were only assigned a gender value if the algorithm returned a high enough probability for one gender. See the OECD.AI website for detailed methodological notes.

SAGE acknowledges that this method does not always correctly determine an author's gender, and so this data should be treated as an approximation only. We hope that more accurate data will become available in future, as more academic publishers provide options for individual authors to self-report their gender identity.

About Science in Australia Gender Equity (SAGE)

Who we are

SAGE is a not-for-profit organisation founded by the Australian Academy of Science and the Australian Academy of Technological Sciences and Engineering.

Our purpose is to promote and advocate for gender diversity and equity in Australian higher education and research, thereby fostering excellence, building trust in the sector, and strengthening its future.

What we do

We support our subscribers to create healthy and productive workplaces.

We provide evidence-driven strategies, capacity building, professional development and collaborative practitioner networks that foster sustainable and embedded change, while providing a pathway to accreditation through the internationally recognised Athena Swan Charter.

"SAGE is the only transformative gender equity program of its kind in Australia designed to achieve sustained cultural change via a national accreditation framework. Measures to enable SAGE to be adopted by all higher education and research institutions across Australia would bring unparalleled impact."

— Women in STEM Decadal Plan

Contact

Dr Janin Bredehoeft
Chief Executive Officer
Science in Australia Gender Equity (SAGE)
sage@sciencegenderequity.org.au