

Large language models challenge the future of higher education



ChatGPT is a chatbot based on a large language model (LLM) that generates text in dialogue format. It was publicly released by OpenAI in December 2022 and has sent shockwaves through the higher education sector for its ability to create polished, confident-sounding text, which could be used to write essays and assignments. While for now it can produce answers¹ that are only competent enough to achieve a passing mark, it is capable of correctly answering multiple-choice questions across several subject areas, including passing sample questions from high-profile licensing examinations. The rate of progress of such applications has been such that it is not difficult to imagine that a much-improved successor of ChatGPT will be released soon.

One question that arises is whether and how higher education should react. Should universities ban its use? Or should academics instead accept that language models will become integral to their professional toolkit, and incorporate them in our teaching and assessment practices?

On a practical level, allowing the use of LLM-based tools would impact the structure of assessment. And on the level of professional conduct, many share the sentiment that using text that is produced by a LLM is on a par with committing plagiarism. As universities already have harsh penalties in place to sanction plagiarism by other means, it seems natural to extend them to LLMs. A problem with this approach, however, is that it will be challenging to enforce. Unlike copy-and-pasting or paraphrasing, LLMs produce new text that is not traceable to a single source, and although software to check the likelihood of LLM-aided cheating has been released (ref. 2), their reliability appears to be low for now. Moreover, any attempt to upgrade detection software is likely to fail³ in the face of fast-evolving LLMs.

Another reaction by some universities has been to (at least temporarily) revert to old-fashioned pen-and-paper, invigilated examinations as their primary mode of assessment. While this solution will dramatically reduce LLM-related cheating in the short-term,

it is unlikely to be a sustainable or widely applicable one. The approach can only be used in traditional institutions where students are physically present, and it is a regressive move with respect to the digital transformations in higher education⁴ delivery and assessment that were instigated by the global COVID-19 pandemic. Transforming written assessment into oral exams may be better suited to digital environments, yet this brings concerns of reliability, validity and scalability.

A third type of reaction to LLMs, and perhaps the only sustainable one, is to adapt and embrace them, as envisaged in a recent editorial⁵ in this journal and consistent with the International Baccalaureate's recent announcement regarding their qualifications⁶. There are many possibilities to experiment and be creative with ChatGPT when teaching and assessing students. However, the adoption of ChatGPT (or similar privately owned applications) as part of standard practice raises serious risks of negative operational, financial, pedagogical and ethical consequences for universities. In particular, OpenAI is under no obligation to cater to the needs of educational institutions when it comes to maintenance and access to its model, thus creating basic operational issues if this forms part of the assessment.

The long-term pedagogical implications of accepting LLMs as learning tools also need consideration. Practicing academic writing is a common way to teach and assess logical argumentation and critical thinking⁷ (which ironically are necessary skills to evaluate a LLM's output). Foreign-language students or students who are educationally disadvantaged are likely to be the most affected, with educators placing less emphasis on learning how to craft well-written and argued texts. This could end up strengthening social divides and diminishing social mobility once students graduate and are thrown into working environments where LLMs may not be available or useful.

Another challenge concerns the trust that educators can put in the model, how it was trained and on what data. Text produced by LLMs is a reflection of patterns⁸ in the training data. Its use in education could further

entrench representational harms in ways that are insidiously difficult to document and redress⁹. OpenAI made some progress in improving the accuracy of ChatGPT on factual prompts and also in moderating toxic content. However, the limits of this engineering are impossible to test, and they have come at the cost of exploiting the labour of data workers who, it has emerged¹⁰, were contracted to view and label toxic content. Educators adopting ChatGPT in their teaching would implicitly validate these harmful and extractive practices.

Finally, there should be concern about the resources that are required for running LLMs, particularly in light of hundreds of universities' net-zero and low-carbon commitments. A recent article estimates ChatGPT's daily carbon footprint to be around 23 kg CO₂e, about the same as a single return trip from London to Paris on the Eurostar, but this does not include the cost of training the model. While this may appear relatively small, it will rapidly increase as the technology becomes ubiquitous. Educational institutions should, therefore, be mindful of asking students to use a model whose operation is actively contributing to the climate crisis, unless the value that can be derived from its use demonstrably exceeds the environmental cost.

Given these challenges, what can academics do? One step could be the creation of publicly funded LLMs in collaboration with open, stakeholder-led initiatives like the BigScience project. Such models could be specifically developed for educational settings, ensuring that they are auditable and transparent with regards their human and environmental costs. This will require a forward-looking vision, substantial investments and the active involvement and lobbying of educational institutions and their funders. Excitement about ChatGPT and other LLM tools foreshadows the huge political issue of who owns and sets the standards for education in the age of AI.

Silvia Milano¹✉, Joshua A. McGrane²
& Sabina Leonelli¹

¹Exeter Centre for the Study of the Life Sciences (Egenis), University of Exeter, Exeter, UK. ²Melbourne Graduate School

of Education, University of Melbourne,
Melbourne, Victoria, Australia.

✉ e-mail: s.milano@exeter.ac.uk

Published online: 31 March 2023

References

1. Choi, J. H., Hickman, K. E., Monahan, A. & Schwarcz, D. SSRN <https://doi.org/10.2139/ssrn.4335905> (2023).
2. Heikkilä, M. How to spot AI-generated text. *MIT Technology Review* <https://www.technologyreview.com/2022/12/19/1065596/how-to-spot-ai-generated-text/> (2022).

3. *Nature* **613**, 612–612 (2023).
4. Garcia-Morales, V. J., Garrido-Moreno, A. & Martin-Rojas, R. *Front. Psychol.* **12**, (2021).
5. *Nat. Mach. Intell.* **5**, 1 (2023).
6. Hill, C. & Lawton, W. J. *High. Educ. Policy Manag.* **40**, 598–610 (2018).
7. Liu, F. & Stapleton, P. *Assess. Writ.* **38**, 10–20 (2018).
8. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–623 (ACM, 2021).

9. Weidinger, L. et al. Preprint at arXiv <http://arxiv.org/abs/2112.04359> (2021).
10. Time. Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer. *Time* <https://time.com/6247678/openai-chatgpt-kenya-workers/> (18 January 2023).

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Machine Intelligence* thanks the anonymous reviewers for their contribution to the peer review of this work.