https://doi.org/10.1038/s41562-023-01740-4

Generative AI poses ethical challenges for open science

Check for updates

Open science (OS) is a move towards making research methods and outputs available to all and is founded on the principle of equity. It is seeing popularity at a time when artificial intelligence (AI) tools are also growing in use and potentially reinforcing existing inequities. We must ensure that OS does not unintentionally feed into harmful AI tools and for this we need deep debates and, possibly, new forms of governance over knowledge generation.

Globally, OS has impetus and is linked to greater equity in science. In 2021, the 193 UNESCO member states approved a 'Recommendation on Open Science', which sparked several other initiatives. For example, the US White House declared 2023 as a 'Year of Open Science' and described OS as "research … processes available to all, while respecting diverse cultures, maintaining security and privacy … and equity". One can argue that OS is enjoying unprecedented political support to become the new research paradigm.

Somewhat in parallel, generative AI — a form of data-based AI or machine learning that produces data (for example, text or images) — has been widely popularized since the release of ChatGPT in November 2022. Public releases of generative AI products have created massive global hype. As with previous technological disruptive events[1], it would seem that people must include generative AI tools everywhere, immediately and inevitably.

Nonetheless, we know that data-based AI tools (including these new generative tools) can and do cause harm[2–4]. For example, the Coding Rights project has documented over 20 cases in Latin America in which AI systems contribute to discrimination based on gender and its intersectionalities and the AI Incident Database has more than 500 examples of harms or near harms that are due to the use of AI systems, worldwide. These issues have only increased with the popularity of generative AI: several legal trials are underway in which companies that deploy AI systems have been accused of violating copyright and consumer protection laws. Meanwhile, efforts are underway to develop new policies to regulate these tools (for example, ref. 5 or the AI Act of the EU). We know that AI tools often discriminate, disrespect different cultures, violate privacy and security, and automate inequality. This is antithetical to OS definitions and values.

Generative AI often reuses open science outputs to train harmful tools. Modified by Julián Buede from The Turing Way Community & Scriberia (2023). Illustrations from The Turing Way, CC-BY 4.0.

Yet OS practices may well be contributing to the capacity of AI tools to do harm. Often, generative AI models are developed chiefly by big, for-profit technology companies. For their models to achieve acceptable performance, they must train them with large amounts of high-quality data in various formats. Such data are often costly, but OS practices seek to make them freely available to all. This means that one of the indispensable inputs of generative AI is the very output that OS works hard to generate and perfect: open data, source code, scientific articles and educational resources, all of which are provided for free and are often funded by tax-payer monies. Although intersections between AI and OS will not always be problematic (for example, the AlphaFold Protein Structure Database), it is clear that OS researchers face the dilemma of how to be a part of the OS movement by making data openly available and also demand that data are not used in a way that causes harm[6].

The issues around AI are severe enough that the OS movement and its policymakers should work on this topic. However, because generative AI has been explosive and organizations often need help to establish basic OS policies, these concerns have not yet reached the most visible parts of the OS ecosystem. For instance, the discussions centring on OS policy during the recent CERN-NASA OS summit did not include discussions about these frictions. Similarly, a recent collaborative opinion piece that outlines policy recommendations for open research software omits generative AI[7]. We are concerned that if OS policymakers continue to not explicitly consider AI in their debates, AI harm will continue to propagate and policy creation to prevent it will become increasingly difficult.

Notably, the challenges that generative AI poses to OS have not gone entirely unrecognized. The Open Source Initiative recently started discussions aimed at "reducing confusion for [open source] policymakers, helping developers understand data sharing and transparency … fighting open washing". Furthermore, we are seeing work in creating licenses similar to those developed for open

BOX 1

Resumen en Español

La IA generativa plantea desafíos éticos a la ciencia abierta

La Ciencia Abierta (CA) se basa en la equidad para compartir la investigación científica con todas las personas. En 2021, la UNESCO aprobó una 'Recomendación sobre Ciencia Abierta' y la CA tiene actualmente un respaldo político global sin precedentes. Al mismo tiempo, en paralelo, proliferó la inteligencia artificial (IA) generativa.

La IA generativa puede causar daño y discriminación. Esto va en contra de los valores de la CA. Sin embargo, la CA provee libremente datos abiertos de alta calidad, uno de los insumos indispensables para producir IA generative. Se nos plantea el dilema: ¿cómo compartir datos abiertos sin que sean usados para IA generativa de manera perjudicial?

Aunque están en marcha conversaciones preliminares y existen propuestas aisladas, este dilema no ha sido discutido aún en profundidad ni abordado adecuadamente en las políticas para CA. Se necesitaría elaborar un sistema de gobernanza que proteja el conocimiento como bien común y el derecho a la investigación como un derecho humano.

El movimiento de la CA es un lugar idóneo para abordar estos desafíos éticos y encontrar enfoques para hacer que los productos y procesos de investigación estén disponibles para todas las personas de forma responsable.

source[8], such as Open Responsible AI licences. This licensing proposes to open AI products but protect them from uses that, for example, could result in harmful AI tools.

Yet the fundamental question remains that imposing these limitations may go against the nature of OS openness[6]. Selective licensing is an option, but it is important to underline the nontriviality of direct use (that is, open-source code to build an autonomous weapon) and indirect use of open data and source code (that is, open-source library use within another algorithm that is in itself harmful)[6]. Indirect uses may be untraceable.

We echo previous calls[6] to advance a new governance system for knowledge generation from the perspective of the common good and the right to research as a human right. This should include the creation of mechanisms to safeguard the ethical reuse of this common good, protecting it from private appropriation through international and national regulations. We know it is no easy task. It requires deep debate and work that we hope to spark within OS with this Correspondence. We believe the OS movement — particularly in its current promising moment — is the ideal scientific ecosystem to debate these thorny issues and find concrete and ethical approaches to make research products and processes available to all, responsibly.

This Correspondence was edited in English, with a Spanish summary (Box 1) and full translation (Supplementary Information) provided by the author. The translations were not checked for correctness by Springer Nature.

Laura Acion[1,2] ✉, **Mariela Rajngewerc**[1,2,3], **Gregory Randall**[4] **& Lorena Etcheverry**[5]

[1]MetaDocencia, Ciudad Autónoma de Buenos Aires, Argentina. [2]Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina. [3]Departamento de Computación, Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Córdoba, Argentina. [4]Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay. [5]Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay.

✉e-mail: laura.acion@metadocencia.org

Published online: 20 November 2023

References

1. Christensen, C. M., McDonald, R., Altman, E. J. & Palmer, J. E. *J. Manage. Stud.* **55**, 1043–1078 (2018).
2. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. On the dangers of stochastic parrots: can language models be too big? In *Proc. 2021 ACM Conf. on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623 (ACM, 2021).
3. Ricaurte, P. *Media Cult. Soc.* **44**, 726–745 (2022).
4. Meskó, B. & Topol, E. J. *Digit. Med.* **6**, 120 (2023).
5. Belli, L., Curzi, Y. & Gaspar, W. *Comput. Law Secur. Rev.* **48**, 105767 (2023).
6. Randall, G. & Díaz Charquero, P. *Integración y Conocimiento* **12**, 51–68 (2023).
7. McKiernan, E. C. et al. *PLoS Biol.* **21**, e3002204 (2023).
8. Contractor, D. et al. Behavioral use licensing for responsible AI. In *Proc. 2022 ACM Conf. on Fairness, Accountability, and Transparency (FAccT '22)*, 778–788 (ACM, 2022).

Acknowledgements

This publication has been made possible by NASA Grants 80NSSC23K0854 and 80NSSC23K0857, CZI grant DAF2021-239366 and grant https://doi.org/10.37921/522107izqogv from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation (funder https://doi.org/10.13039/100014989), and Programa de Dedicación Total, Universidad de la República, Uruguay. We thank P. Díaz Charquero, C. Gentemann, K. Hertweck, and the Open Life Science and The Turing Way communities for their input on this piece.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01740-4.