Editorial

Living in a brave new AI era

🔴 Check for updates

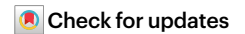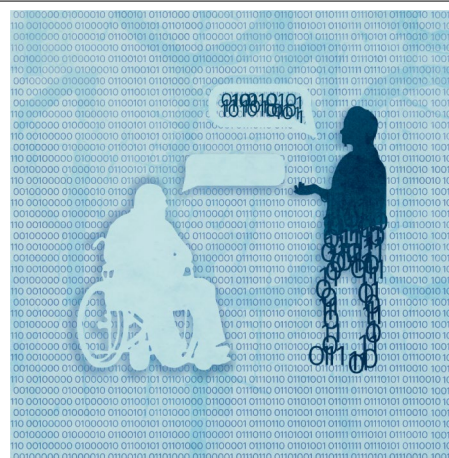**Although artificial intelligence (AI) was already ubiquitous, the recent arrival of generative AI has ushered in a new era of possibilities as well as risks. This Focus explores the wide-ranging impacts of AI tools on science and society, examining both their potential and their pitfalls.**

No AI tool was used to write this editorial. Twelve months ago, such a statement would have seemed absurd — but over the past year, the world has changed dramatically. As a journal, we are committed to learning more about how AI tools are changing our science and our society. To achieve this, we have brought together scientists from a range of disciplines to contribute their views on how we can and should navigate the AI frontier.

AI tools are created by humans and, as with any tool, the stated aim of their creators is often to improve what we do: to make our work more efficient or assist us when we are struggling. This is in many ways incredibly exciting for our communities. For example, large language models (LLMs) can translate and proofread, which could make science more accessible[1]. Clinicians can use AI tools to detect disease and predict responses to treatment[2]. But efficiency can have a less desirable side. Experts warn of job market upheaval in almost all occupations due to the rapid uptake of AI technology[3]. The use of generative AI tools may also lead to other social harms. As we highlight in out Focus, AI tools may carry substantial risks for democracy, as people can use them to create more effective disinformation that is increasingly difficult to detect.

Biases that are built in and hard to overcome can make it challenging to use AI tools for social good. LLMs perform poorly for minoritized languages and this may worsen existing inequalities in access to technology. AI tools are also often biased towards Western ways of thinking and promote Western interests. Their rapid spread across the world can be considered a form of digital colonialism, and appropriation of AI tools is one way to resist this.

The biases inherent in AI tools come mainly from the data that are used to train them. But could scientists also be unwittingly contributing to these biases? In recent years, scientists, journals (including ourselves) and governments have advocated for scientists to make all data publicly accessible as part of a movement towards open and reproducible science. Yet, this means that large volumes of scientific data can be used as training data for generative AI tools — potentially leading to harms and inequalities.

New AI technologies promise many exciting opportunities for human behavioural and cognitive scientists. Researchers can use LLMs to better understand human cognition, as long as the providers make them openly accessible for researchers. As society adopts generative AI systems, AI is likely to affect our culture. Generative AI tools now contribute, transmit and select for cultural traits, including through novel content generation, personalized recommendations and the need for new skill sets in labour markets. This marks a notable change in the way that human culture is evolving.

This brings us back to the practicalities of living and working with AI tools. We do not consider generative AI tools as qualifying for authorship, and we agree with one of the authors of our Focus who writes that current probability-based language models are no substitute for scientific review by expert peers. However, we do see promise in their use for zero-shot translation by scientists (in which the LLM is given human-authored text), and our own authors have used AI for this purpose. For all manuscripts in which authors use generative AI tools, we ask them to declare this in their Methods (for research papers) or Acknowledgements (for opinion and review pieces).

It has been less than a year since ChatGPT was made publicly available and since then generative AI has had substantial social and scientific impacts. At *Nature Human Behaviour*, we are positive about the potential of AI tools, but mindful of their pitfalls. We intend for this to be an ongoing collection and welcome future submissions at the intersection of AI and human behaviour.

References

1. Berdejo-Espinola, V. & Amano, T. *Science* **379**, 991 (2023).
2. Lenharo, M. *Nature* **622**, 686–688 (2023).
3. OECD. *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market* (OECD, 2023).