



A chatbot called ChatGPT can help to write text for essays, scientific abstracts and more.

ChatGPT: five priorities for research

Eva A. M. van Dis, Johan Bollen, Robert van Rooij, Willem Zuidema & Claudi L. Bockting

Conversational AI is a game-changer for science. Here's how to respond.

Since a chatbot called ChatGPT was released late last year, it has become apparent that this type of artificial intelligence (AI) technology will have huge implications on the way in which researchers work.

ChatGPT is a large language model (LLM), a machine-learning system that autonomously learns from data and can produce sophisticated and seemingly intelligent writing after training on a massive data set of text. It is the latest in a series of such models released by OpenAI, an AI company in San Francisco, California, and by other firms. ChatGPT has

caused excitement and controversy because it is one of the first models that can convincingly converse with its users in English and other languages on a wide range of topics. It is free, easy to use and continues to learn.

This technology has far-reaching consequences for science and society. Researchers and others have already used ChatGPT and other large language models to write essays and talks, summarize literature, draft and improve papers, as well as identify research gaps and write computer code, including statistical analyses. Soon this technology will evolve to the point that it can design experiments, write and complete manuscripts, conduct peer review and support editorial decisions to accept or reject manuscripts.

Conversational AI is likely to revolutionize research practices and publishing, creating both opportunities and concerns. It might accelerate the innovation process, shorten time-to-publication and, by helping people to write fluently, make science more equitable and increase the diversity of scientific

perspectives. However, it could also degrade the quality and transparency of research and fundamentally alter our autonomy as human researchers. ChatGPT and other LLMs produce text that is convincing, but often wrong, so their use can distort scientific facts and spread misinformation.

We think that the use of this technology is inevitable, therefore, banning it will not work. It is imperative that the research community engage in a debate about the implications of this potentially disruptive technology. Here, we outline five key issues and suggest where to start.

Hold on to human verification

LLMs have been in development for years, but continuous increases in the quality and size of data sets, and sophisticated methods to calibrate these models with human feedback, have suddenly made them much more powerful than before. LLMs will lead to a new generation of search engines¹ that are able to produce detailed and informative answers to complex user questions.

But using conversational AI for specialized research is likely to introduce inaccuracies, bias and plagiarism. We presented ChatGPT with a series of questions and assignments that required an in-depth understanding of the literature and found that it often generated false and misleading text. For example, when we asked 'how many patients with depression experience relapse after treatment?', it generated an overly general text arguing that treatment effects are typically long-lasting. However, numerous high-quality studies show that treatment effects wane and that the risk of relapse ranges from 29% to 51% in the first year after treatment completion^{2–4}. Repeating the same query generated a more detailed and accurate answer (see Supplementary information, Figs S1 and S2).

Next, we asked ChatGPT to summarize a systematic review that two of us authored in *JAMA Psychiatry*⁵ on the effectiveness of cognitive behavioural therapy (CBT) for anxiety-related disorders. ChatGPT fabricated a convincing response that contained several factual errors, misrepresentations and wrong data (see Supplementary information, Fig. S3). For example, it said the review was based on 46 studies (it was actually based on 69) and, more worryingly, it exaggerated the effectiveness of CBT.

Such errors could be due to an absence of the relevant articles in ChatGPT's training set, a failure to distil the relevant information or being unable to distinguish between credible and less-credible sources. It seems that the same biases that often lead humans astray, such as availability, selection and confirmation biases, are reproduced and often even amplified in conversational AI⁶.

Researchers who use ChatGPT risk being misled by false or biased information, and

VITOR MIRANDA/ALAMY

incorporating it into their thinking and papers. Inattentive reviewers might be hoodwinked into accepting an AI-written paper by its beautiful, authoritative prose owing to the halo effect, a tendency to over-generalize from a few salient positive impressions⁷. And, because this technology typically reproduces text without reliably citing the original sources or authors, researchers using it are at risk of not giving credit to earlier work, unwittingly plagiarizing a multitude of unknown texts and perhaps even giving away their own ideas. Information that researchers reveal to ChatGPT and other LLMs might be incorporated into the model, which the chatbot could serve up to others with no acknowledgement of the original source.

Assuming that researchers use LLMs in their work, scholars need to remain vigilant. Expert-driven fact-checking and verification processes will be indispensable. Even when LLMs are able to accurately expedite summaries, evaluations and reviews, high-quality journals might decide to include a human verification step or even ban certain applications that use this technology. To prevent human automation bias – an over-reliance on automated systems – it will become even more crucial to emphasize the importance of accountability⁸. We think that humans should always remain accountable for scientific practice.

Develop rules for accountability

Tools are already available to predict the likelihood that a text originates from machines or humans. Such tools could be useful for detecting the inevitable use of LLMs to manufacture content by paper mills and predatory journals, but such detection methods are likely to be circumvented by evolved AI technologies and clever prompts. Rather than engage in a futile arms race between AI chatbots and AI-chatbot-detectors, we think the research community and publishers should work out how to use LLMs with integrity, transparency and honesty.

Author-contribution statements and acknowledgements in research papers should state clearly and specifically whether, and to what extent, the authors used AI technologies such as ChatGPT in the preparation of their manuscript and analysis. They should also indicate which LLMs were used. This will alert editors and reviewers to scrutinize manuscripts more carefully for potential biases, inaccuracies and improper source crediting. Likewise, scientific journals should be transparent about their use of LLMs, for example when selecting submitted manuscripts.

Research institutions, publishers and funders should adopt explicit policies that raise awareness of, and demand transparency about, the use of conversational AI in the preparation of all materials that might become part of the published record.

Publishers could request author certification that such policies were followed.

For now, LLMs should not be authors of manuscripts because they cannot be held accountable for their work. But, it might be increasingly difficult for researchers to pinpoint the exact role of LLMs in their studies. In some cases, technologies such as ChatGPT might generate significant portions of a manuscript in response to an author's prompts. In others, the authors might have gone through many cycles of revisions and improvements using the AI as a grammar- or spellchecker, but not have used it to author the text. In the future, LLMs are likely to be incorporated into text processing and editing tools, search engines and programming tools. Therefore they might contribute to scientific work without authors necessarily being aware of the nature or magnitude of the contributions. This defies today's binary definitions of authorship, plagiarism and sources, in which someone is either an author, or not, and a source has either been used, or not. Policies will have to adapt, but full transparency will always be key.

Inventions devised by AI are already causing a fundamental rethink of patent law⁹, and

“One of the most immediate issues for the research community is the lack of transparency.”

lawsuits have been filed over the copyright of code and images that are used to train AI, as well as those generated by AI (see go.nature.com/3y4aery). In the case of AI-written or -assisted manuscripts, the research and legal community will also need to work out who holds the rights to the texts. Is it the individual who wrote the text that the AI system was trained with, the corporations who produced the AI or the scientists who used the system to guide their writing? Again, definitions of authorship must be considered and defined.

Invest in truly open LLMs

Currently, nearly all state-of-the-art conversational AI technologies are proprietary products of a small number of big technology companies that have the resources for AI development. OpenAI is funded largely by Microsoft, and other major tech firms are racing to release similar tools. Given the near-monopolies in search, word processing and information access of a few tech companies, this raises considerable ethical concerns.

One of the most immediate issues for the research community is the lack of transparency. The underlying training sets and LLMs for ChatGPT and its predecessors are not publicly available, and tech companies might conceal the inner workings of their

conversational AIs. This goes against the move towards transparency and open science, and makes it hard to uncover the origin of, or gaps in, chatbots' knowledge¹⁰. For example, we prompted ChatGPT to explain the work of several researchers. In some instances, it produced detailed accounts of scientists who could be considered less influential on the basis of their h-index (a way of measuring the impact of their work). Although it succeeded for a group of researchers with an h-index of around 20, it failed to generate any information at all on the work of several highly cited and renowned scientists – even those with an h-index of more than 80.

To counter this opacity, the development and implementation of open-source AI technology should be prioritized. Non-commercial organizations such as universities typically lack the computational and financial resources needed to keep up with the rapid pace of LLM development. We therefore advocate that scientific-funding organizations, universities, non-governmental organizations (NGOs), government research facilities and organizations such as the United Nations – as well tech giants – make considerable investments in independent non-profit projects. This will help to develop advanced open-source, transparent and democratically controlled AI technologies.

Critics might say that such collaborations will be unable to rival big tech, but at least one mainly academic collaboration, BigScience, has already built an open-source language model, called BLOOM. Tech companies might benefit from such a program by open sourcing relevant parts of their models and corpora in the hope of creating greater community involvement, facilitating innovation and reliability. Academic publishers should ensure LLMs have access to their full archives so that the models produce results that are accurate and comprehensive.

Embrace the benefits of AI

As the workload and competition in academia increases, so does the pressure to use conversational AI. Chatbots provide opportunities to complete tasks quickly, from PhD students striving to finalize their dissertation to researchers needing a quick literature review for their grant proposal, or peer-reviewers under time pressure to submit their analysis.

If AI chatbots can help with these tasks, results can be published faster, freeing academics up to focus on new experimental designs. This could significantly accelerate innovation and potentially lead to breakthroughs across many disciplines. We think this technology has enormous potential, provided that the current teething problems related to bias, provenance and inaccuracies are ironed out. It is important to examine and advance the validity and reliability of LLMs so that researchers know how to use the technology judiciously for specific research practices.

Some argue that because chatbots merely learn statistical associations between words in their training set, rather than understand their meanings, LLMs will only ever be able to recall and synthesize what people have already done and not exhibit human aspects of the scientific process, such as creative and conceptual thought. We argue that this is a premature assumption, and that future AI-tools might be able to master aspects of the scientific process that seem out of reach today. In a 1991 seminal paper, researchers wrote that “intelligent partnerships” between people and intelligent technology can outperform the intellectual ability of people alone¹¹. These intelligent partnerships could exceed human abilities and accelerate innovation to previously unthinkable levels. The question is how far can and should automation go?

AI technology might rebalance the academic skill set. On the one hand, AI could optimize academic training – for example, by providing feedback to improve student writing and reasoning skills. On the other hand, it might reduce the need for certain skills, such as the ability to perform a literature search. It might also introduce new skills, such as prompt engineering (the process of designing and crafting the text that is used to prompt conversational AI models). The loss of certain skills might not necessarily be problematic (for example, most researchers do not perform statistical analyses by hand any more), but as a community we need to carefully consider which academic skills and characteristics remain essential to researchers.

If we care only about performance, people's contributions might become more limited and obscure as AI technology advances. In the future, AI chatbots might generate hypotheses, develop methodology, create experiments¹², analyse and interpret data and write manuscripts. In place of human editors and reviewers, AI chatbots could evaluate and review the articles, too. Although we are still some way from this scenario, there is no doubt that conversational AI technology will increasingly affect all stages of the scientific publishing process.

Therefore, it is imperative that scholars, including ethicists, debate the trade-off between the use of AI creating a potential acceleration in knowledge generation and the loss of human potential and autonomy in the research process. People's creativity and originality, education, training and productive interactions with other people will probably remain essential for conducting relevant and innovative research.

Widen the debate

Given the disruptive potential of LLMs, the research community needs to organize an urgent and wide-ranging debate. First, we recommend that every research group immediately has a meeting to discuss and try ChatGPT for themselves (if they haven't

Questions for debate

Issues for discussion at a forum about conversational AIs.

- Which research tasks should or should not be outsourced to large language models (LLMs)?
- Which academic skills and characteristics remain essential to researchers?
- What steps in an AI-assisted research process require human verification?
- How should research integrity and other policies be changed to address LLMs?
- How should LLMs be incorporated into the education and training of researchers?
- How can researchers and funders aid the development of independent open-source LLMs and ensure the models represent scientific knowledge accurately?
- What quality standards should be expected of LLMs (for example, transparency, accuracy, bias and source crediting) and which stakeholders are responsible for the standards as well as the LLMs?
- How can researchers ensure that LLMs promote equity in research, and avoid risks of widening inequities?
- How should LLMs be used to enhance principles of open science?
- What legal implications do LLMs have for scientific practice (for example, laws and regulations related to patents, copyright and ownership)?

already). And educators should talk about its use and ethics with undergraduate students. During this early phase, in the absence of any external rules, it is important for responsible group leaders and teachers to determine how to use it with honesty, integrity and transparency, and agree on some rules of engagement. All contributors to research should be reminded that they will be held accountable for their work, whether it was generated with ChatGPT or not. Every author should be responsible for carefully fact-checking their text, results, data, code and references.

Second, we call for an immediate, continuing international forum on development and responsible use of LLMs for research. As an initial step, we suggest a summit for relevant stakeholders, including scientists of different disciplines, technology companies, big research funders, science academies, publishers, NGOs and privacy and legal specialists. Similar summits have been organized to discuss and develop guidelines in response to other disruptive technologies, such as human gene editing. Ideally, this discussion should

result in quick, concrete recommendations and policies for all relevant parties. We present a non-exhaustive list of questions that could be discussed at this forum (see ‘Questions for debate’).

One key issue to address is the implications for diversity and inequalities in research. LLMs could be a double-edged sword. They could help to level the playing field, for example by removing language barriers and enabling more people to write high-quality text. But the likelihood is that, as with most innovations, high-income countries and privileged researchers will quickly find ways to exploit LLMs in ways that accelerate their own research and widen inequalities. Therefore, it is important that debates include people from under-represented groups in research and from communities affected by the research, to use people's lived experiences as an important resource.

Science, similar to many other domains of society, now faces a reckoning induced by AI technology infringing on its most dearly held values, practices and standards. The focus should be on embracing the opportunity and managing the risks. We are confident that science will find a way to benefit from conversational AI without losing the many important aspects that render scientific work one of the most profound and gratifying enterprises: curiosity, imagination and discovery.

The authors

Claudi L. Bockting is a professor of clinical psychology in the Department of Psychiatry at Amsterdam UMC, University of Amsterdam, the Netherlands, and co-director of the Centre for Urban Mental Health at the Institute for Advanced Study, University of Amsterdam.

Eva A. M. van Dis, Johan Bollen, Robert van Rooij, Willem Zuidema.

A full list of author affiliations and Supplementary information accompanies this Comment online (see go.nature.com/3ww5fyz).

e-mail: c.l.bockting@amsterdamumc.nl

- Grant, N. & Metz, C. *The New York Times* (21 December 2022).
- Jelovac, A., Kolshus, E. & McLoughlin, D. M. *Neuropsychopharmacol.* **38**, 2467–2474 (2013).
- Kato, M. et al. *Mol. Psychiatry* **26**, 118–133 (2021).
- Vittingl, J. R., Clark, L. A., Dunn, T. W. & Jarrett, R. B. *J. Consult. Clin. Psychol.* **75**, 475–488 (2007).
- van Dis, E. A. M. et al. *JAMA Psychiatry* **77**, 265–273 (2020).
- Rich, A. S. & Gureckis, T. M. *Nature Mach. Intell.* **1**, 174–180 (2019).
- Thorndike, E. L. *J. Appl. Psychol.* **4**, 25–29 (1920).
- Skitka, L. J., Mosier, K. & Burdick, M. D. *Int. J. Human-Comp. Stud.* **52**, 701–717 (2000).
- George, A. & Walsh, T. *Nature* **605**, 616–618 (2022).
- Rudin, C. *Nature Mach. Intell.* **1**, 206–215 (2019).
- Salomon, G., Perkins, D. N. & Globerson, T. *Edu. Res.* **20**, 2–9 (1991).
- Melnikov, A. A. et al. *Proc. Natl Acad. Sci. USA* **115**, 1221–1226 (2018).

The authors declare no competing interests.