EDITORIAL

Check for
updates

Exorcising the ghost in the computer: ChatGPT, science publishing, and GenAI policy for *Behavioral Ecology and Sociobiology*

Theo C. M. Bakker[1] · James F. A. Traniello[2]

In a previous editorial, we discussed how recognizing the influences of word-processing technology, neglectful writing practices, and the consequences of representing the ideas of others as one's own can reduce plagiarism and maintain the quality of the scientific record (Traniello and Bakker 2016). This policy is explicit in the Instructions for Authors for *Behavioral Ecology and Sociobiology*: "No data, text, or theories by others are presented as if they were the author's own." We cautioned "When you prepare a manuscript, ask yourself introspectively *are these ideas and words my own?* before you submit your work for critical review."

Now a significant new challenge to upholding the highest standards in scientific publishing has emerged: the use of text generators to create narratives that authors can falsely represent as their own work. This violation of ethical principles has been called *aigiarism* (Brainard 2023). ChatGPT, short for Chat Generative Pre-training Transformer, is a large language model (LLM) developed by the non-profit organization OpenAI and launched in November, 2022. This generative artificial intelligence (GenAI) tool has the ability to "write" fluent human-like texts as well as tables, codes, and other "intellectual" products in response to prompts. ChatGPT 3.5 is freely available online, facilitating its widespread use. The introduction of ChatGPT has created an ethical emergency, forcing the scientific community to evaluate the lack of regulation for using GenAI and determine its consequences (Conroy 2023), risks (Clarke 2023; Thorp 2023), and potential benefits (e.g., Noy and

Zhang 2023). Consensus on the legitimacy and inclusion of AI-generated text in scientific papers has yet to be reached (Stokel-Walker and Van Noorden 2023).

Understanding the operation of ChatGPT can provide insight into its application and regulation. ChatGPT is an AI algorithm, a neural network trained by massive amounts of online text, including books, news items, Wikipedia entries, and software code (Hutson 2022) to learn statistical patterns of language (Stokel-Walker and Van Noorden 2023). Training determines the reliability of ChatGPT as an accurate source of information because texts are biased, inconsistently available, and contain errors and prejudices. Narratives must be evaluated based on when information used in training was written to validate that the information accessed is not outdated. The training data of the free version ChatGPT 3.5 dates to September 2021; the pay version, GPT-4, has improved reference accuracy (Bom 2023) and often greater performance than previous versions (Bubeck et al. 2023). ChatGPT had been trained on over 40 terabytes of text—almost 40 million books in kindle format (Khalil and Er 2023). The lack of transparency of the training data and algorithms for ChatGPT and other GenAI models have been criticized as it "makes it hard to uncover the origin of, or gaps in, chatbots' knowledge" (van Dis et al. 2023).

The use of ChatGPT to summarize, translate, and create text is multifarious: it can improve language quality, explore ideas, generate drafts, gather information quickly, assist in searching literature (Jarrah et al. 2023), and improve productivity (Noy and Zhang 2023). ChatGPT can also be used as a tool to assist data analysis by generating codes and tabulating and plotting data, applications that will be refined in the future. ChatGPT, with appropriate caution and controls, offers opportunities to advance scientific information processing and make science more inclusive and equitable (Berdejo-Espinola and Amano 2023).

All texts generated by LLMs, however, require critical evaluation. ChatGPT may produce "hallucinations" such as "erroneous references, content, and statements ….

✉ Theo C. M. Bakker
  tbakker@evolution.uni-bonn.de
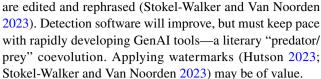
✉ James F. A. Traniello
  jft@bu.edu

1 1Bonn Institute for Organismic Biology (BIOB), Section Animal Biodiversity, University of Bonn, An der Immenburg 1, 53121 Bonn, Germany

2 Department of Biology, Boston University, 5 Cummington Mall, Boston, MA 02215, USA

intertwined with correct information, and presented in a persuasive and confident manner, making their identification difficult without close inspection and effortful fact-checking" (Bubeck et al. 2023). These inaccuracies — if left unchecked—might influence the content of manuscripts prepared for *Behavioral Ecology and Sociobiology*. For example, one of JFAT's graduate students recently attempted to use ChatGPT to locate papers on ant diets and brain evolution. Output stated that E.O. Wilson had published a relevant paper, giving title and journal, but we knew the work did not exist. When queried, ChatGPT cited *The Insect Societies* as the source and provided a page number. This was also incorrect. Since *Insect Societies* was published in 1971, the error did not appear to be due to the lack of an updated data base. Other results retrieved in the search included erroneous authorships and journal citations, even if paper titles were accurate. "ChatGPT and other LLMs have a tendency to spit out false references, which could be a signal for peer reviewers looking to spot use of these tools in manuscripts" (M. Hodgkinson in Conroy (2023)).

Publishers have developed policies to restrict the use of ChatGPT (Brainard 2023). Large Language Models (LLMs), such as ChatGPT, do not currently satisfy Springer Nature's authorship (https://www.springer.com/us/editorial-policies/authorship-principles) criteria. Notably an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs. Use of an LLM should be properly documented in the Methods section (and if a Methods section is not available, in a suitable alternative part) of the manuscript. Springer Nature is monitoring ongoing developments in Artificial Intelligence closely and will review (and update) these policies (https://www.springer.com/gp/editorial-policies/artificial-intelligence--ai-/25428500) as appropriate. In Information for Authors of *Science* journals, AI policy is made clear: "Text generated from AI, machine learning, or similar algorithmic tools cannot be used in papers published …. nor can the accompanying figures, images, or graphics be the products of such tools, without explicit permission from the editors. In addition, an AI program cannot be an author of a *Science* journal paper. A violation of this policy constitutes scientific misconduct" (see also Thorp 2023). It is agreed that ChatGPT and similar tools should be excluded as "authors" of scientific papers (Stokel-Walker and Van Noorden 2023) and in reviews submitted to funding agencies (Kaiser 2023).

Manuscript submission procedures of many journals routinely screen for plagiarism by using detection software like iThenticate and Turnitin. Similar procedures should be applied to discriminate GenAI text. Detection software for plagiarism, however, cannot reliably detect GenAI texts and can be outcompeted by ChatGPT itself (Khalil and Er 2023). Detection will become more difficult if AI generated texts are edited and rephrased (Stokel-Walker and Van Noorden 2023). Detection software will improve, but must keep pace with rapidly developing GenAI tools—a literary "predator/prey" coevolution. Applying watermarks (Hutson 2023; Stokel-Walker and Van Noorden 2023) may be of value.

The Instructions for Authors of *Behavioral Ecology and Sociobiology*, a Springer Nature journal, are explicit about the use of GenAI as authorship and in preparing manuscripts (see above). Our role as Editors-in-Chief is to assist the Publisher in developing policies that effectively adapt to changing and challenging technological and ethical landscapes to maintain excellence in the quality of papers published in our journal. Peer reviewers play a vital role in scientific publishing. Their expert evaluations and recommendations guide editors in their decisions and ensure that published research is valid, rigorous, ethical, and replicable. Editors select peer reviewers primarily because of their in-depth knowledge of the subject matter and/or methods of the work they are asked to evaluate. This expertise is invaluable. Peer reviewers are accountable for the accuracy and views expressed in their reports, and the peer-review process operates on a principle of mutual trust between authors, reviewers, and editors. Despite rapid progress, generative AI tools have considerable limitations: they can lack up-to-date knowledge and may produce nonsensical, biased or false information. Manuscripts may also include sensitive or proprietary information that should not be shared outside the peer review process. For these reasons we ask that, while Springer Nature explores providing our peer reviewers with access to safe AI tools, peer reviewers do not upload manuscripts they are reviewing for *Behavioral Ecology and Sociobiology* into generative AI platforms (https://www.springer.com/gp/editorial-policies). If any part of the evaluation of the claims made in the manuscript is in any way supported by an AI tool, we ask peer reviewers to declare the use of such tools transparently in their peer-review report. Communications with our associate editors will ensure the policy is successfully implemented and invitation letters sent for manuscript review will be accordingly revised.

Generative AI may extend the capabilities of traditional search engines in preparing scientific manuscripts but its misuse must be prevented to maintain the honesty and accuracy of science. With proper disclosure and critical analysis, the use of ChatGTP-generated text in scientific writing may be made ethically legitimate (Eke 2023; Jarrah et al. 2023) and accurate. Eke (2023) suggests what transparent use could look like: "….referencing ChatGPT could involve documenting date of generation, prompts used for generation, and limiting the use of direct quotation to one paragraph." Concerning manuscripts submitted to *Behavioral Ecology and Sociobiology*, we will restrict the use of ChatGPT and similar LLMs:

1. ChatGPT will be excluded from the authorship of manuscripts;
2. ChatGPT will be allowed to assist in language correction, literature search, and the preparation of figures (except images) with legends and tables with headings if output is properly corrected and annotated;
3. The use of ChatGPT must be disclosed in the Methods section by stating the LLM used, date of application, specification of prompts, and whether output was edited. It must also be stated if ChatGPT was not used.
4. Manuscripts suspected of (mis)using ChatGPT will be handled according to guidelines of the Committee on Publication Ethics on Authorship and AI Tools (https://publicationethics.org/cope-position-statements/ai-author).

We recognize that implementing these policies and practices and ensuring compliance will require effort from journal editors and assistants, editorial board members, and reviewers, but will be necessary to uphold high standards.

Broader impacts of GenAI should also be considered, together with its applications and consequences in science publishing. GenAI can be viewed as intellectual theft on an extraordinary scale: copyrighted work is used to train GenAI without consent, credit, or remuneration. Millions of poorly paid "clickworkers" and "ghostworkers" are exploited to develop the training process of AI applications by cleaning, coding, and categorizing texts and images (Viana Braz et al. 2023). This energy-demanding process has a significant ecological footprint (Stokel-Walker and Van Noorden 2023). Additionally, public discourse, which may be infiltrated by GenAI, must be free from abuses that could foster denialism and the distribution of misinformation (Sinatra and Hofer 2023).

Declarations

Conflict of interest The authors declare they have no conflict of interest.

References

Berdejo-Espinola V, Amano T (2023) AI tools can improve equity in science. Science 379:991

Bom H-SH (2023) Exploring the opportunities and challenges of ChatGPT in academic writing: a roundtable discussion. Nucl Med Mol Imaging 57:165–167

Brainard J (2023) Journals take up arms against AI-written text. Science 379:740–741

Bubeck S, Chandrasekaran V, Eldan R et al (2023) Sparks of artificial general intelligence: early experiments with GPT-4. arXiv:2303.12712

Clarke L (2023) Call for AI pause highlights potential dangers. Science 380:120–121

Conroy G (2023) Scientific sleuths spot dishonest ChatGPT use in papers. Nature Published Online. https://doi.org/10.1038/d41586-023-02477-w

Eke DO (2023) ChatGPT and the rise of generative AI: threat to academic integrity? J Responsible Technol 13:100060

Hutson M (2022) Could AI help you to write your next paper? Nature 611:192–193

Hutson M (2023) Conflicting visions for AI regulation. Nature 620:261–263

Jarrah AM, Wardat Y, Fidalgo P (2023) Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say? Online J Commun Media Technol 13:e202346

Kaiser J (2023) Funding agencies say no to AI peer review. Science 381:261

Khalil M, Er E (2023) Will ChatGPT get you caught? Rethinking of plagiarism detection. arXiv 2302:04335

Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. Science 381:187–192

Sinatra G, Hofer BK (2023) ChatGPT and other generative AI could foster science denial and misunderstanding – here's how you can be on alert. The Conversation, https://theconversation.com/chatgpt-and-other-generative-ai-could-foster-science-denial-and-misunderstanding-heres-how-you-can-be-on-alert-204897

Stokel-Walker C, Van Noorden R (2023) What ChatGPT and generative AI mean for science. Nature 614:214–216

Thorp HH (2023) ChatGPT is fun, but not an author. Science 379:313

Traniello JFA, Bakker TCM (2016) Editorial: Intellectual theft: pitfalls and consequences of plagiarism. Behav Ecol Sociobiol 70:1789–1791

van Dis EAM, Bollen J, van Rooij R, Zuidema W, Bockting CL (2023) ChatGPT: five priorities for research. Nature 614:224–226

Viana Braz M, Tubaro P, Casilli AA (2023) Microwork in Brazil: who are the workers behind artificial intelligence? Research Report DiPLab & LATRAPS, https://diplab.eu/?p=2833