Article:

Johnson, T.F. orcid.org/0000-0002-6363-1825, Simmons, B.I., Millard, J. orcid.org/0000-0002-3025-3565 et al. (4 more authors) (2024) Pressure to publish introduces large-language model risks. Methods in Ecology and Evolution, 15 (10). pp. 1771-1773. ISSN 2041-210X

https://doi.org/10.1111/2041-210x.14397

Received: 3 April 2024 | Accepted: 3 July 2024

DOI: 10.1111/2041-210X.14397

FORUM

Methods in Ecology and Evolution

Pressure to publish introduces large-language model risks

Thomas F. Johnson[1] | Benno I. Simmons[2] | Joseph Millard[3] | Tanya Strydom[1] | Alain Danet[1] | Amy R. Sweeny[1] | Luke C. Evans[4]

[1]Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, UK

[2]Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn, UK

[3]Biodiversity Futures Lab, Natural History Museum, London, UK

[4]Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK

Correspondence
Thomas F. Johnson
Email: t.f.johnson@sheffield.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: NE/R016801/1, NE/T003502/1, NE/V006800/1 and NE/V006916/1

Handling Editor: Robert B. O'Hara

Abstract

1. Large-language models (LLMs) have the potential to accelerate research in ecology and evolution, cultivating new insights and innovation. However, whilst revelling in the plethora of opportunities, researchers need to consider that LLM use could also introduce risks.
2. An important piece of context underpinning this perspective is the pressure to publish, where research careers are defined, at least partly, by publication metrics like number of papers, impact factor, citations etc. Coupled with academic employment insecurity, especially during early career, researchers may reason that LLMs are a low-risk and high-reward tool for publication.
3. However, this pressure to publish can introduce risks if LLMs are used as a shortcut to game publication metrics instead of a tool to support true innovation. These risks may ultimately reduce research quality, stifle researcher development and incur reputational damage for researchers and the entire scientific record.
4. We conclude with a series of recommendations to mitigate the magnitude of these risks and encourage researchers to apply caution whilst maximising LLM potential.

KEYWORDS
ecology, evolution, large-language models, paper hacking, publish or perish

Innovation invites excitement over novel uses, concern over misuses and fears about detrimental impacts on individuals and society. Large-language models (LLMs) represent a significant innovation that could impact how science is conducted, for better and for worse. Cooper et al. (2024) provide a timely overview of LLM use for research and teaching in ecology and evolution and suggest approaches to maximise LLM utility, especially in coding exercises. We agree with the points made by Cooper et al. (2024), but in this complementary extension, we highlight that the potential of LLMs extends beyond coding and could transform the entire research process from writing to reviewing and introduces new risks to scientific progress if applied incautiously. We term these risks: *paper hacking*, *stunted researcher development* and *reputational risk*.

To frame our perspective, an important piece of context is the pressure to publish and the use of publication metrics as markers of researcher accomplishment. Scientists are typically judged through academic publishing and are incentivised to publish to progress in their career, that is 'publish or perish' (van Dalen & Henkens, 2012). Indeed, over a 10-year period, researchers beginning their careers in 2000 published 2.6 times more papers than researchers beginning their careers in 1950 (Fire & Guestrin, 2019), with the number of publications rising exponentially across an expanding number of journals (McGill, 2024). Combined with the current global socio-economic climate and academic job rarity, pressure on researchers (especially early career), is high. Against this backdrop of incentivised output and employment insecurity, researchers may reason that LLMs are a valuable tool for increasing publication rates.

1 | PAPER HACKING

The advent of statistical software disrupted the field of ecology and evolution, with the scientific process shifting towards computational approaches (Petrovskii & Petrovskaya, 2012). LLMs have the capacity to rival and even surpass this disruption, as they not only have the ability to accelerate code development, but can also automate much of the research process. This could result in unparalleled innovation, but may exacerbate quality issues already creeping into our science. For instance, analytical shortcuts like improper model selection, 'causal salads' (McElreath, 2020) and p-hacking have introduced reliability issues into scientific fields (Fraser et al., 2018). Presently, these issues arise (at least partly) because researchers can rapidly try many analyses without needing a rich understanding of the methods or a deep exploration of the research topic. These issues could be supercharged with LLM use, as LLMs provide opportunities to not only shortcut analyses, but convincingly automate much of the research process, essentially 'paper hacking'.

Aspects of LLM automation are already entering the literature, with papers containing made-up (hallucinated) citations (Joelving, 2023), and authors forgetting to remove LLM prompts from writing (Zhang et al., 2024). Given LLMs are known to struggle with several tasks—see Cooper et al. (2024)—there is a risk that even with sound intentions, LLM use could reduce work quality. With skewed intentions the risks would be far more severe and the antithesis of the slow science movement (Frith, 2020). We anticipate a litany of convincing LLM errors and hallucinations entering and compromising the scientific record over the next decade. One could argue these risks will be reduced by the peer-review process, where human assessors will catch and correct these errors. However, the burden on reviewers and editors is already high, and LLMs are convincing, if not always correct. Risks could be further inflated if publishers and journals use LLMs as part of the review process (Liu & Shah, 2023), with LLMs marking their own homework. As a community, we must apply caution and due diligence when using LLMs to reduce these risks, without stifling their tremendous potential.

2 | STUNTED RESEARCHER DEVELOPMENT

There are multiple components of the job of a scientific researcher: writing papers and grants, designing experiments and teaching students. Through doing these things, a researcher learns them. Senior researchers, in theory, are experienced enough in these tasks to judge the accuracy of outputs from an LLM. For early-career researchers, there is a risk that individuals learn to equate writing with prompting and that researchers learn the habits of a tool that is not trained to teach them. Ultimately, LLMs may mature and improve to the extent that the value of conventional scientific skills, like writing, may depreciate. However, the risks of use, in the short-term, are not fully apparent. For instance, there are concerns that AI-based tools like LLMs inflate confidence in our understanding, but not necessarily improve understanding to the same extent, resulting in overconfidence (Messeri & Crockett, 2024).

3 | REPUTATIONAL RISK

Given the importance of proper attribution and reliability of findings in science, authors may risk losing credibility if it is discovered that their work is primarily an LLM output, or of low quality (see Section 1). This is especially concerning as the guidelines of LLM use are still being defined, meaning LLM practices that are acceptable now may be deemed unacceptable in the future. This could be particularly problematic when it comes to *who* is most likely to make use of LLMs. LLMs are marketed as bridging tools for non-native speakers, and this group of authors are the most at risk of further scrutiny as rules and opinions about the use of LLMs are altered, further alienating authors who already face challenges within research and publishing spheres. Damages to the credibility of science as a whole also risk further reducing an already low public trust in science (Tyson, 2023).

Cooper et al. (2024) provide a series of guidelines for LLM use within the Methods in Ecology and Evolution journal. These guidelines, whilst helpful, may not mitigate the above risks and we need to be on our guard against potential misuses, whilst still embracing the opportunities this technology presents. It is important to note, too, that the risks we identify are very much a function of LLM technology, and wider society, in its current state. There is a huge research interest and investment in minimising phenomena like hallucinations; this technology is still young, and thus the technological concerns raised here are likely to reduce as LLMs mature. Moreover, as AI becomes more dominant, cultural norms may change—it is not impossible to imagine a future where fully automated paper writing is accepted and 'manual writing' is seen as an antiquated skill. Whether this is desirable is a different question. Thus, our concerns about deskilling could be a product of the time in which they are written.

Our concerns are not solely attributable to LLMs; they are a product of the global socio-economic climate and the rarity of academic jobs and funding. Solutions to mitigate or at least dampen the risks of LLMs may be structural as well as technological: First, to maintain credibility and improve trust within science, authors must be candid regarding the contribution of LLMs and consider the ethics of applications. Given the novelty of LLMs, a sensible rule of application could be to only use LLMs when the user or someone in the team has the expertise to review, verify, validate and take responsibility for the outputs, a value echoed in Cooper et al. (2024). However, it is worth noting that cognitive biases can impede our ability to self-assess expertise (Kruger & Dunning, 1999; Rahmani, 2020). Second, to ensure early-career researchers develop into highly competent and well-rounded scientists, universities and mentors need to rapidly develop a strong grasp of LLM pedagogy, and probe students to ensure they gain a rich understanding of their work, and the importance of quality. Third, we should continue the shift away from entirely metric-based judgement, favouring alternatives like narrative CVs and the adoption of DORA declarations, which allow peers to see achievements within context and appreciate the broader quality and impact of one's work.

We should also not allow the risks associated with LLM use from stifling their adoption, instead we need to find the instances where the benefits of LLMs outweigh the risks, with real promise in areas from evidence synthesis (Berger-Tal et al., 2024) to computer vision (Berrios et al., 2023). More broadly, as a field, we need to continue discussions over appropriate LLM use, and be prepared to adapt guidelines. As scientists, we strive for innovation, but not at the cost of the quality of science.

AUTHOR CONTRIBUTIONS

Writing—Original draft: Thomas F. Johnson, Joseph Millard, Benno I. Simmons, Luke C. Evans. Writing—Review and editing: Thomas F. Johnson, Joseph Millard, Benno I. Simmons, Tanya Strydom, Alain Danet, Amy R. Sweeny, Luke C. Evans.

ACKNOWLEDGEMENTS

TFJ and AD were supported by a UKRI-NERC Grant NE/T003502/1, LCE was supported by a UKRI-NERC Grant NE/V006916/1, JM is funded by the NERC Highlights grant GLiTRS NE/V006800/1. ARS is supported by a large NERC grant NE/R016801/1. Large-language models did not contribute to this perspective.

CONFLICT OF INTEREST STATEMENT

We have no conflicts of interest to report.

DATA AVAILABILITY STATEMENT

No data or code was used in the creation of this manuscript.

ORCID

Thomas F. Johnson https://orcid.org/0000-0002-6363-1825

Joseph Millard https://orcid.org/0000-0002-3025-3565

Alain Danet https://orcid.org/0000-0002-1592-9483

REFERENCES

Berger-Tal, O., Wong, B. B. M., Adams, C. A., Blumstein, D. T., Candolin, U., Gibson, M. J., Greggor, A. L., Lagisz, M., Macura, B., Price, C. J., Putman, B. J., Snijders, L., & Nakagawa, S. (2024). Leveraging AI to improve evidence synthesis in conservation. *Trends in Ecology & Evolution*, *39*(6), 548–557. https://doi.org/10.1016/j.tree.2024.04.007

Berrios, W., Mittal, G., Thrush, T., Kiela, D., & Singh, A. (2023). Towards language models that can see: Computer vision through the LENS of natural language. *arXiv*, arXiv:2306.16410. https://doi.org/10.48550/arXiv.2306.16410

Cooper, N., Clark, A. T., Lecomte, N., Qiao, H., & Ellison, A. M. (2024). Harnessing large language models for coding, teaching and inclusion to empower research in ecology and evolution. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.14325

Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: Observing Goodhart's law in action. *GigaScience*, *8*(6), giz053. https://doi.org/10.1093/gigascience/giz053

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS One*, *13*(7), e0200303. https://doi.org/10.1371/journal.pone.0200303

Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, *24*(1), 1–2. https://doi.org/10.1016/j.tics.2019.10.007

Joelving, A. F. (2023). Withdrawn AI-written preprint on millipedes resurfaces, causing alarm. *Retraction Watch*. https://retractionwatch.com/2023/09/01/withdrawn-ai-written-preprint-on-millipedes-resurfaces-causing-alarm/

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Liu, R., & Shah, N. B. (2023). ReviewerGPT? An exploratory study on using large language models for paper reviewing. *arXiv*, arXiv:2306.00622. https://doi.org/10.48550/arXiv.2306.00622

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and STAN* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

McGill, B. (2024). The state of academic publishing in 3 graphs, 6 trends, and 4 thoughts. *Dynamic Ecology*. https://dynamicecology.wordpress.com/2024/04/29/the-state-of-academic-publishing-in-3-graphs-5-trends-and-4-thoughts/

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*(8002), 49–58. https://doi.org/10.1038/s41586-024-07146-0

Petrovskii, S., & Petrovskaya, N. (2012). Computational ecology as an emerging science. *Interface Focus*, *2*(2), 241–254. https://doi.org/10.1098/rsfs.2011.0083

Rahmani, M. (2020). Medical trainees and the Dunning–Kruger effect: When they don't know what they don't know. *Journal of Graduate Medical Education*, *12*(5), 532–534. https://doi.org/10.4300/JGME-D-20-00134.1

Tyson, B. K. (2023). *Americans' trust in scientists, positive views of science continue to decline*. https://www.pewresearch.org/science/2023/11/14/americans-trust-in-scientists-positive-views-of-science-continue-to-decline/

van Dalen, H. P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. *Journal of the American Society for Information Science and Technology*, *63*(7), 1282–1293. https://doi.org/10.1002/asi.22636

Zhang, M., Wu, L., Yang, T., Zhu, B., & Liu, Y. (2024). The three-dimensional porous mesh structure of Cu-based metal-organic-framework—Aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries. *Surfaces and Interfaces*, *46*, 104081. https://doi.org/10.1016/j.surfin.2024.104081

How to cite this article: Johnson, T. F., Simmons, B. I., Millard, J., Strydom, T., Danet, A., Sweeny, A. R., & Evans, L. C. (2024). Pressure to publish introduces large-language model risks. *Methods in Ecology and Evolution*, *15*, 1771–1773. https://doi.org/10.1111/2041-210X.14397