LETTERS

Edited by **Jennifer Sills**

Undue publicity for flawed fraud detector

In the News In Depth story "Fake scientific papers are alarmingly common" (9 May, https://scim.ag/2KD), J. Brainard describes a medRxiv preprint by Sabel *et al.* (*1*) alleging that nearly 30% of all scientific articles may be the fake products of paper mills. Fraudulent papers are indeed a concern, but this story sensationalizes the study's findings while downplaying its substantial flaws. In doing so, Brainard unjustly tarnishes the scientific enterprise and causes undeserved harm to the reputation of scientists from the countries singled out in the preprint.

Sabel *et al.* propose a classifier for identifying fake papers. Rather than training this classifier using machine learning, the researchers preselected just three features to flag suspected fakes: whether the authors use private email addresses, whether they are affiliated with a hospital, and—unmentioned by the News story—whether the team lacks international collaborators. Brainard eventually cites details from the preprint showing that the algorithm misclassified more than a third of real papers as fake. This unacceptable false-positive rate indicates that the method has failed.

The structure of the Sabel *et al.* study demonstrates how such classifiers can be used to reinforce discrimination and inequity. Because papers without international coauthors, as well as authors without institutional email addresses, are

deemed untrustworthy, the classifier is more likely to identify papers from Asia and the Global South, where international collaborations are less common (*2*) and personal email addresses are more frequently used (*3*). Yet Sabel and colleagues use the output of this classifier to argue that fraud is more common among authors from these countries, without controlling for these well-documented international differences. Such circular reasoning unjustly disparages authors from these countries.

The algorithm that *Science* has heralded is fundamentally biased in its implementation, with potentially racist consequences. Using a highly biased algorithm even as a preliminary screen is unacceptable without carefully considering issues of fairness and algorithmic harm that arise from its use. No such consideration is given, and the News story gives only passing mention to the risks. *Science* should be on the vanguard of drawing attention to how machine intelligence can encode racism while perpetuating and exacerbating traditional inequities, not an accessory to such wrongs.

Carl T. Bergstrom[1]* and C. Brandon Ogbunugafor[2]
[1]Department of Biology, University of Washington, Seattle, WA, USA. [2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA.
*Corresponding author.
Email: cbergst@u.washington.edu

REFERENCES AND NOTES
1. B. A. Sabel, E. Knaack, G. Gigerenzer, M. Bilc, medRxiv 10.1101/2023.05.06.23289563 (2023).
2. A. Gazni, C. R. Sugimoto, F. Didegah, *J. Am. Societ. Info. Sci. Technol.* **63**, 323 (2012).
3. S. Shen, R. Rousseau, D. Wang, *Scientometrics* **115**, 1039 (2018).

10.1126/science.adi7104

Advancing equity and integrity in research

In the News In Depth story "Fake scientific papers are alarmingly common" (9 May, https://scim.ag/2KD), J. Brainard draws attention to a number of new tools and technologies that researchers and publishers are developing in an effort to increase the integrity of research. These tools primarily work by automating the identification of fake authors and fake papers with the use of a set of variables common in publications, such as author email address and affiliation. The goals of these technologies are valid and important, but the scientific community needs to act cautiously when determining which variables are used to identify and flag individuals and manuscripts.

A private email address could be considered a flag for further scrutinizing a publication, but private email addresses are often used by legitimate researchers, many of whom are nonnative English speakers or use non-Latin writing scripts. In countries such as India and China, more researchers use personal email addresses than institutional ones (*1*). Judging the legitimacy of research based on this flag coupled with a hospital affiliation, as one medRxiv preprint Brainard cites does (*2*), without any follow-up practice that accounts for language or cultural differences, would be a disservice to scientific advancement globally.

More rigorous author and manuscript evaluation needs to verify individuals and institutions by leveraging not

just email domain and affiliations but also author disambiguation tools and persistent identifier registries, such as Open Researcher and Contributor Identifiers (ORCIDs) and the Research Organization Registry. An additional analysis of publishing history—including the number of publications, coauthorship, and a network analysis—should be applied to create signals for additional checks on the author of a manuscript. These results should then be balanced with knowledge and sensitivity. Graduate students and new faculty with a light publication history are not fake authors, just new to the field.

Ensuring research integrity requires reducing inequities and gatekeeping in the publishing ecosystem. Some of the tools cited by Brainard use limited checks that prematurely categorize and ultimately miscategorize researchers as nefarious actors. Both the proposed tool and the choice by *Science* to promote it contribute to furthering inequities rather than improving trust in science.

The Committee on Publication Ethics (COPE) has developed sensible guidance (*3*) for editors and publishers on how to navigate any potential paper-mill case once it has been identified. Better evaluation mechanisms will allow the scientific community to simultaneously advance equity and integrity goals.

Leslie McIntosh[1]* and **Cynthia Hudson Vitale[2]**
[1]Digital Science, London, UK. [2]Association of Research Libraries, Washington, DC, USA.
*Corresponding author.
Email: l.mcintosh@digital-science.com

REFERENCES AND NOTES
1. S. Shen, R. Rousseau, D. Wang, *Scientometrics* 115, 1039 (2018).
2. B. A. Sabel, E. Knaack, G. Gigerenzer, M. Bilc, medRxiv 10.1101/2023.05.06.23289563 (2023).
3. COPE, "Addressing concerns about systematic manipulation of the publication process" (2023).

10.1126/science.adi7867

Editor's note

We thank Bergstrom and Ogbunugafor, as well as McIntosh and Hudson Vitale, for pointing out the limitations of tools for detecting fake papers such as the one developed by Sabel *et al*. Our story states prominently, in the fourth paragraph, that Sabel *et al*.'s method "isn't a perfect solution, because of a high false-positive rate." Later we specify the high false-positive numbers. Far from heralding or sensationalizing the tool, we presented it as a rough indicator of a real problem.

Tim Appenzeller
News Editor, *Science*

10.1126/science.adj3681

Undue publicity for flawed fraud detector

Carl T. Bergstrom and C. Brandon Ogbunugafor

View the article online
https://www.science.org/doi/10.1126/science.adi7104
Permissions
https://www.science.org/help/reprints-and-permissions