

NEWS | 23 September 2022

'Papermill alarm' software flags potentially fake papers

The text-analysis tool could be used to help detect bogus research submitted to journals.



Holly Else



The Papermill Alarm looks for similarities to text found in bogus papers. Credit: Raimund Koch/Getty

A software tool that analyses the titles and abstracts of scientific papers and detects text similar to that found in bogus articles is gaining interest from publishers.

The tool, called the Papermill Alarm, was developed by Adam Day, who is director of scholarly data-services company Clear Skies in London. Day says he ran all the titles listed in citation database PubMed through the system, and found that 1% of currently listed papers contain text very similar to that of articles produced by paper mills – companies or individuals that fabricate scientific manuscripts to order. The Papermill Alarm does not say definitively whether an article is fabricated, but flags those that are worthy of further investigation.



The fight against fake-paper factories that churn out sham science

Day says his analysis is not intended to estimate the scale of paper-milling among PubMed entries, because it can recognize only papers that are similar to those from known paper mills. Many more paper mills might exist, and legitimate papers could also get flagged for having similar wording, he says. “It’s like a fishing net. It’s not a fishing rod.”

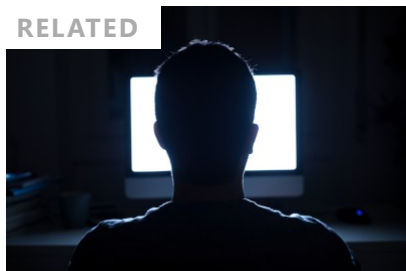
Anna Abalkina, an economist at the Free University of Berlin who studies paper mills, says that the scientific community will benefit from automated checks that can detect potentially bogus papers.

Suspicious submissions

Many publishers already use software and other methods [to help detect fraudulent activity](#) and spot junk papers. Some manuscript-processing systems can detect and flag if many submissions come from the same computer, for example – a sign that one person or organization could be churning out a large number of studies. But Day says his approach of analysing text is new. Six publishers, including SAGE in Thousand Oaks, California, where Day works as a data scientist, have expressed interest in using the Papermill Alarm to screen submitted manuscripts.

The tool uses a deep-learning algorithm to compare the language used in the titles and abstracts of manuscripts with that used in articles known to have come from paper mills. The comparison is based on lists of paper-mill articles compiled by research-integrity sleuths including [Elisabeth Bik](#) and David Bimler (also [known by the pseudonym Smut Clyde](#)). The tool uses a traffic-light system, assigning red flags to papers with many similarities to known paper-mill articles, orange flags to those with some similarities and green flags to those with none.

RELATED



What makes an undercover science sleuth tick? Fake-paper detective speaks out

There have so far been few estimates of the prevalence of articles from paper mills. A June [report by the Committee on Publication Ethics](#) in Eastleigh, UK, suggested that 2% of papers submitted to journals come from paper mills, and said that the problem “threatens to overwhelm the editorial processes of a significant number of journals”.



Even Day’s finding of 1% of published PubMed papers coming from paper mills is “too high for comfort”, says Bimler. “These junk papers do get cited. People seize on them to prop up their own bad ideas and sustain dead-end research programmes,” he adds.

Bik says that the real number of paper-mill papers listed in PubMed might be even higher, but points out that their impact on science overall is probably low, because most of these articles are not highly cited or influential. “But it damages the reputation of science and the trust that we put into research papers,” she says.

doi: <https://doi.org/10.1038/d41586-022-02997-x>

Latest on:

[Publishing](#)

