



Correcting the record: retracting papermill articles

UKSG eNews 560

Elliott Lumb, Co-Founder, **Signals**

In 2023, the scholarly literature had the highest number of retractions ever seen in a single year, driven by over 8000 retractions at **Hindawi**. The large set of retractions were fraudulent articles created by papermills and published after **Wiley** acquired Hindawi in 2021. The consequences of this have been huge for Wiley, and the industry as a whole. Wiley estimated that the problem cost them \$35-40 million in lost revenue in 2023. They have closed several Hindawi journals and will stop using the Hindawi brand. The broader industry has become acutely aware of the papermill issue. Many publishers are taking steps to investigate their journal portfolios for issues, tightening their integrity checks, establishing new research integrity teams equipped with innovative tools, and joining industry initiatives to tackle papermills. It's unlikely that Hindawi is the only publisher with papermill issues, and with prolific research integrity sleuths combing through the published literature to identify problematic articles, 2024 will likely see even more retractions than 2023.

In recent years we have seen an increase in the number of papermill articles published in scholarly journals. Generally speaking, these papermills create fake articles and sell authorship to individuals who require a publication. Papermills have several ways of manipulating the publishing process. They use peer review rings, enlist rogue editors, take advantage of the 'guest editor' role in Special Issues, or fabricate convincing articles that pass peer review.

Fraudulent papermill articles harm research. They erode trust in the scholarly literature, which should lay the foundation of knowledge. Nonsensical articles can be used in training data for AI models, and researchers waste resources differentiating between real and fake articles, or even worse, use fake research in their own work. One necessary solution is to identify and retract these articles.

Unfortunately, article retractions are typically perceived as a bad thing. Publishers can be reluctant to retract articles due to fear of being delisted from indexes such as Web of Science and Scopus and consequently receiving fewer high-quality article submissions. We need to change this narrative around retractions. Ultimately, the most important thing is that the

scholarly record is trustworthy and accurate, enabling research to progress. Publishers should be encouraged to actively identify problematic articles in their portfolios and, when necessary, correct the scholarly record by issuing retractions or corrections.

Our industry has recognized that there is an issue with fraudulent articles. This is evident from the recent formation of several industry initiatives, including the **STM** Integrity Hub and **United2Act**. Publishers must own their past mistakes, accept that there will be mistakes in the future, and see retractions as best practice when it comes to dealing with published fraudulent articles. Wiley has been upfront about fake articles in the Hindawi portfolio, which is a step in the right direction.

It is not trivial for publishers to identify and retract problematic articles. Identifying them can be difficult. Many large publishers have hundreds of journals, each publishing hundreds of articles per year. Research integrity teams are growing within publishers, but they are still a relatively new and limited resource, and searching through published articles for issues is extremely time-consuming. Fortunately, there are new tools available to help research integrity experts in this task. At **Signals**, we can analyze a publisher's whole portfolio, and show them which journals have the most issues, which articles should be investigated, and why. This streamlines the investigation process and helps publishers to correct the scholarly record.

Retracting an article can be a slow, challenging process. New retraction workflows may be required to help publishers deal with new forms of publishing fraud. As an industry, we can also improve retraction notices. Not all retractions are equal; researchers, institutions, and publishers should be able to differentiate between papermill retractions and retractions due to honest mistakes.

While retractions help correct the scholarly record, they do not solve the underlying problem of publishing fraud. The papermill problem is likely to continue. The pressure for researchers to publish and the financial incentives for open-access publishers to publish increasing volumes of content have not gone away, and there are no indications that this will change in the near term. Publishers will need to expand their research integrity teams, install additional checks at article submission, and equip editors with tools to help them identify problematic articles.

A new challenge is that papermills will use generative AI to create convincing but entirely fake articles at scale. The rapidity of this process also means that papermills can adapt to research integrity tools that assess content, such as plagiarism and image manipulation checks, resulting in an evolutionary arms race. A new approach, used by **Signals**, to identify fraudulent articles avoids the arms race by focusing on metadata, such as citation networks, which papermills can't avoid and become more robust over time as researchers engage with articles.

Retracted papermill articles are often cited by unidentified papermill articles. This is an effective way of discovering fraudulent articles as it cuts to the root of the papermill business model. Approaches like this, combined with expert input will be key to identifying and preventing publication fraud.

As an industry, we should work towards establishing a practice of efficient retraction of fraudulent articles. By doing so, we can help correct the scholarly recording and restore trust in research.