

Exposing Fake Science in Biomedical Science

How to detect mass-produced fake
research — and why it matters for the
future of science.



for Diversity and Responsible
Research and Innovation

Fake Publications in Biomedical Science:

red-flagging method indicates mass production.

Introduction:
René Bernard



01.06.2026



12 pm (CET)



Online

The Hidden Crisis

Paper Mills

Commercial operations mass-producing fake biomedical papers using AI — then selling them to researchers under pressure to publish.

The Buyers

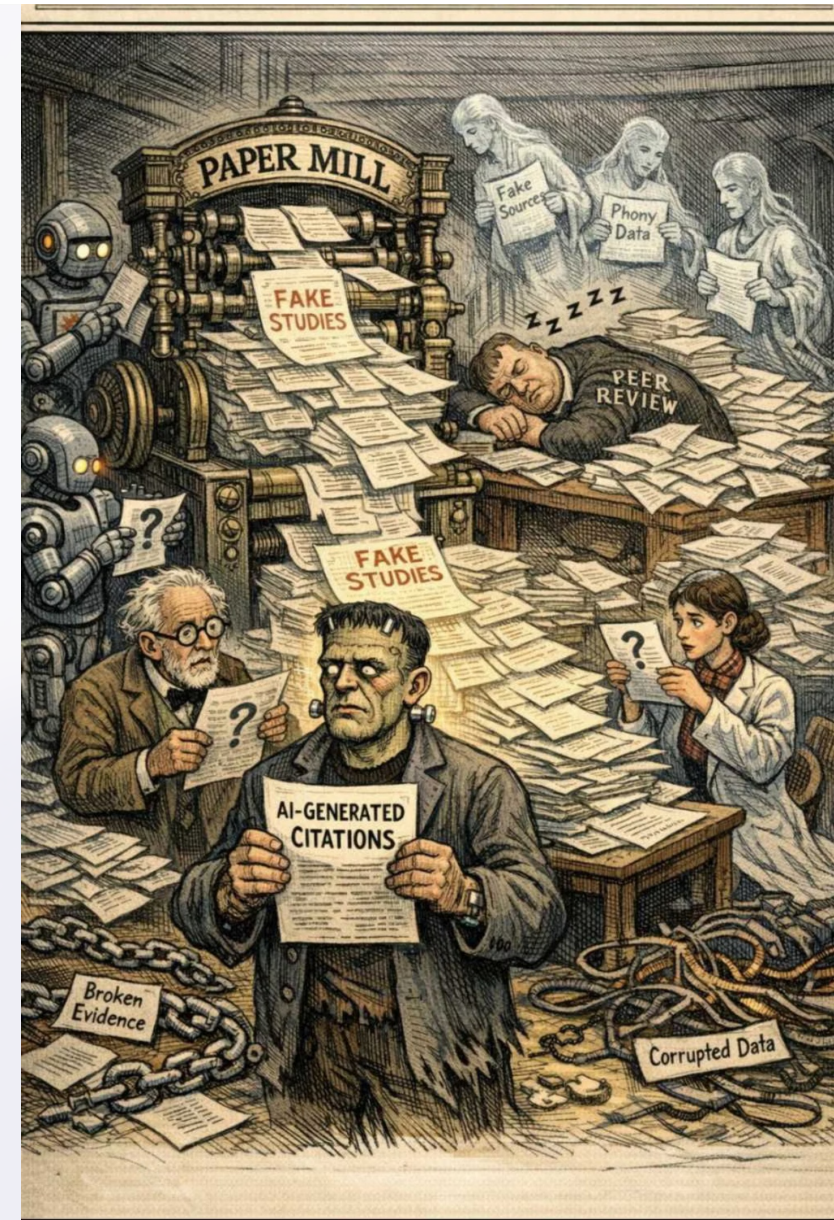
Students and academics who need publications to advance careers, often unaware of the consequence fake papers have

The Damage

Fraudulent, non-existent science enters the permanent record — polluting the evidence base that medicine and policy rely on.

The Magnitude

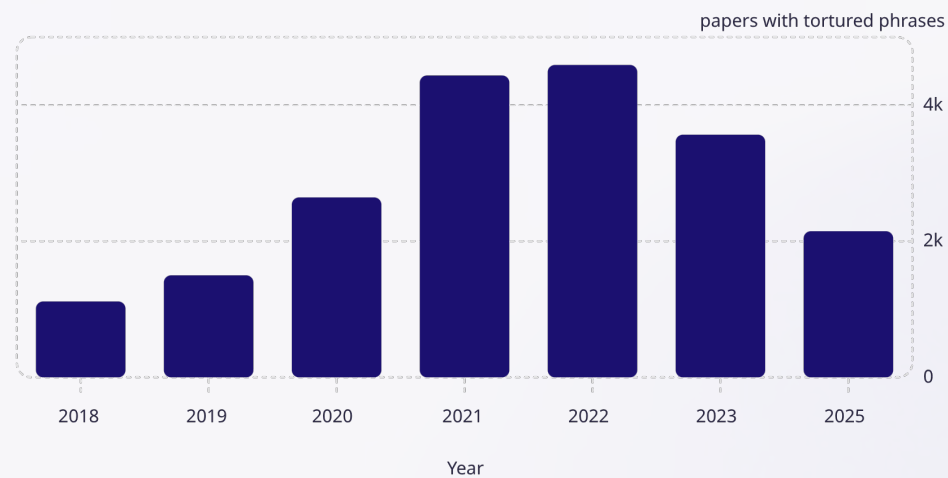
Sabel and colleagues explored how large the problem is and by how much scientific literature is already "infiltrated" by fake papers.



Paper Mills are not new

Computer scientist **Guillaume Cabanac** and his colleagues coined the term "tortured phrases" (2021) and launched the [Problematic Paper Screener \(PPS\)](#) ; 550 tortured phrases among the 16,000 new papers

Original Scientific Term	Tortured Paper Mill Phrase
Artificial intelligence	Counterfeit consciousness
Breast cancer	Bosom peril
Signal to noise	Flag to clamor



The Detective's Toolkit: Methodology



Study 1: The authors needed a **simple, scalable method** to identify potentially fake papers — without reading every article by hand.

- List institutional (!) (professional) email addresses of all co-authors (in the order of authorship)
- Did you engage a professional agency to help write your paper? [Yes/No] If yes, please state the agency name, website, and email
- If so requested by the editor, are you able and willing to provide the original data and images of your publication? [Yes/No]
- Provide name and institutional email of your academic leadership: (e.g. President/Dean); Head of HR Department

Exploration phase (studies 1-6)

Questionnaires submitted to corresponding authors of published "questionable" papers — assumption: honest researchers will answer in detail; authors of fake papers will not answer or vaguely reply.

Confirmation phase (study 8)

400 known fake papers were compared against 400 random, legitimate articles to identify distinguishing signals. Which indicators actually work reliably?

The detection switch (study 7)

Bibliometric estimate (study 9)

Hypotheses and Tests

S2-S6: Detection indicator for flagging suspicious publications: *"private email"* AND *"no international co-author"*

Hypotheses	Associated Studies	flagged papers
H1 Authors of potentially fake publications are less willing than authors of presumed genuine publications to respond transparently to editorial quality-control questions or provide verifiable underlying research information and original data.	S1 Do authors of suspicious papers avoid transparency and quality-control checks?	20.5%
H2 If the red-flag indicators identified in suspicious papers are truly associated with paper-mill activity, their <u>frequency should increase over time</u> in parallel with the reported global rise of industrialized fake publishing.	S2 Do red-flag indicators increased over time in neuroscience journals?	13.4% → 33.7% (neuroscience journals, 2010–2020 trend)
H2 extended; comparing two disciplines	S3 Performed a time-trend analysis across neuroscience and medical journals from 2010–2020.	19.4% → 24.0% (medical journals, 2010–2020 trend)
H3: If paper mills preferentially target journals with lower barriers to publication, then red-flagged papers should occur more frequently <u>in lower-impact-factor and certain high-volume open-access journals</u> .	S4 Validation of the red-flagging method in a larger neuroscience journal dataset (alliance of 70 Neuroscience journals)	10.5%
	S5 Are red-flagged papers were more common in lower-impact-factor journals?	23.8%
	S6 What is red-flag prevalence in high-volume open-access Frontiers journals?	40.3%

Study 7 - the detection switch

Authors (manually) reviewed results and noticed that many previously red-flagged papers from Studies 1–6 had hospital affiliations. Why?

- Hospitals » Research institutions
- Research institutions are more likely to track their publications (for incentives, internal & external reporting)
- Researchers at hospitals have large fluctuation rate
- Clinicians often need publications for promotion

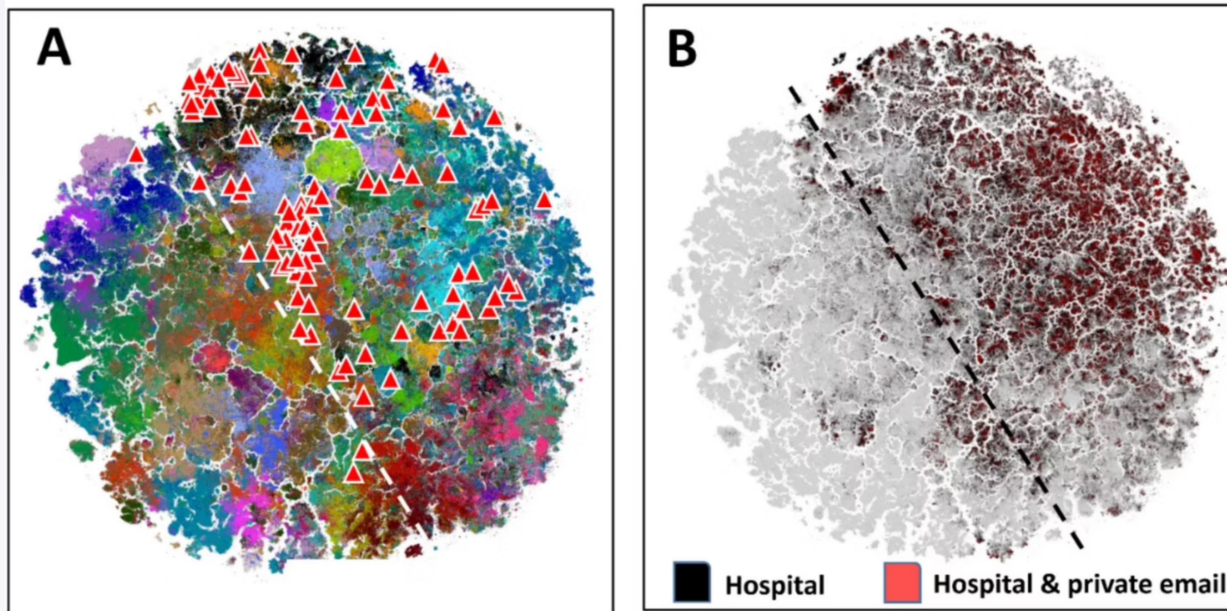
Study 7 detection rule: Private email address AND Hospital affiliation = Red-flagged publication (RFP) | screening population: 4000 pubs from 45 journals published in 2020. | Pure metadata analysis

Study 7 issues

- many legitimate clinicians use private gmail addresses, especially in in lower-resource systems
- overflagging of local/regional clinical environments (China 42%, India 33%, Rest: Iran, Turkey, Russia)
- classifier does not test any of the paper content (are data realy, conclusions valid)
- Study 7 does not detect fraud, it flaggs an anomaly in publication data compared to Western standards

Figure 2

2D Topic Landscape of Paper Mill Publications



A Color patches represent different disciplines (topics) in basic science to the left of the white line (e.g., immunology, physiology, and chemistry) and in clinical science to the right (e.g., pharmacology, neurology, and radiology) (for details, see González-Márquez et al. 2023). The position of actual paper mill publications obtained by one of us (B.S.) (98 red triangles in A) shows that the large majority are from clinical and not basic science disciplines. The paper mill publications in the 12 o'clock cluster are from cancer research and the cluster in the center from microbiology/virology.

B Distribution of publications with hospital affiliation only (black) or hospital affiliation and private email address (red). The distribution shows that private email addresses are widespread across fields, forming clusters in some topics but not in others. The private/institutional email address ratio is $894,124/5,800,617 = 15.4\%$

Flaws in Figure 2

Visualization is misleading us

A: Color patches represent disciplines (but a strict border between clinical and basic science is totally arbitrary; no metadata)

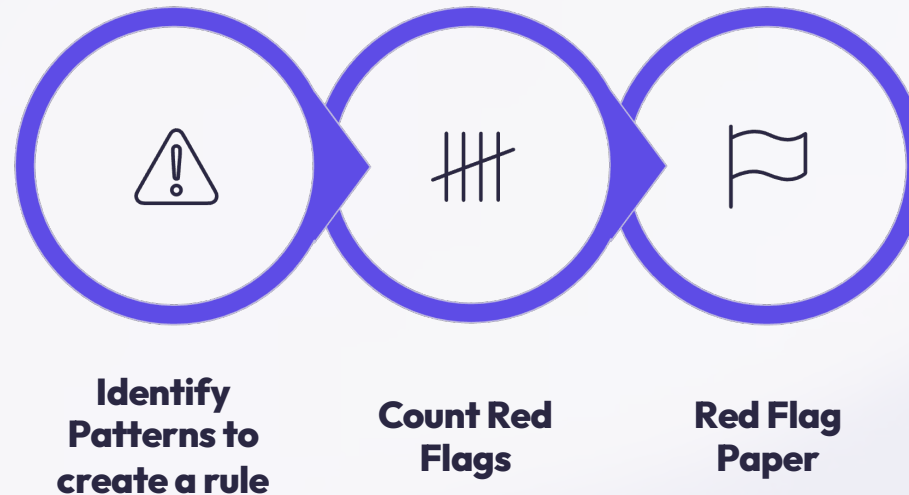
- Colorized map consists of 21,000,000 data points (papers) and the 98 large triangles represent position of paper mill papers. How were the actual positions were obtained? Bernhard Sabel – no explanation - sounds fishy;
- Clustering is arbitrary and is related to the triangle size that out of proportion. Moreover, the clusters are described but not shown (no arrows; guess work)
- Line between basic and clinical science is arbitrary (interdisciplinary papers, pharmacology is not per se clinical)

B: Color choice is terrible (every red dot must by default overlay with a black because there are many hospital + institutional combos).

- Only 15% of hospital dots are red. 85% of publications use a pattern that is not flagged. A pattern is not visible in the figure.
- In the end figure tells us that clinical disciplines are associated with hospitals.

The Red-Flagging (Tallying) Rule

If a paper triggers enough suspicious patterns, it gets flagged for review.



- Tallying rule: positive points (+1) private email, hospital affiliation, suspicious country, institutional publication output < 200/year
- Tallying rule: negative points (-1) <1 or >15 coauthors, institution output >1000
- If total score is > 2—> red flagged publication
- The goal was to test how well the red-flagging rules could distinguish the two groups of fake and genuine papers
- Assumption: Both groups are 100% accurate and representative (genuine vs fake) due their selection pattern
- This "multifactorial tallying rule" catches **94% of fake papers** with very few false alarms — making it a practical tool for large-scale screening. But there is also **11,5% false positive rate**.

Study 8: Problems with the verification samples

suspicious/fake papers group (N=400) by the community

- flagged papers within the scientific publishing ecosystem (retractions, editorial concerns, prior forensic analyses, paper-mill publications)
- what is the criterion for label "papermill" (source i, ii, iv) - image or text manipulation
- some papers may only contained one "fraudulent" elements and could be non-fake (retraction ≠ fake)
- potential circularity problem: authors are training on papers already detected by current systems → may reinforce existing detection bias and miss out on novel patterns.
- authors went along with the pragmatic assumptions

Randomly selected legitimate publications (N=400)

- "presumed non-fake" - because of age (prior to paper mill popularity) sample 2000-2010 (2016 - first wave of retractions; first appearance 2004)
- Control group = Mostly real papers + Unknown number of hidden fake papers (level of uncertainty)
- What are hard and testable criteria for a real, genuine paper?
- If the control group contains undetected fake papers
 - the 11.5% false-alarm rate could be overestimated
 - the distinction between groups becomes blurred
- authors went along with the pragmatic assumptions

Study 9: Testing the System at Scale

“Does this paper statistically resemble papers previously associated with paper mills or fraud?”

- applied the **multifactorial tallying rule** from Study 8 (If total score is > 2 —> red flagged publication)
- screened a very large PubMed dataset,
- and then used **Bayes' theorem** to estimate the *true prevalence* of fake/suspicious papers (**11,5% false positive rate, 94 % true positive rate**).

17,120

Papers Screened

From PubMed, covering random samples from two years.

14.9%

Flagged in 2020

Of sampled papers triggered the red-flag threshold.

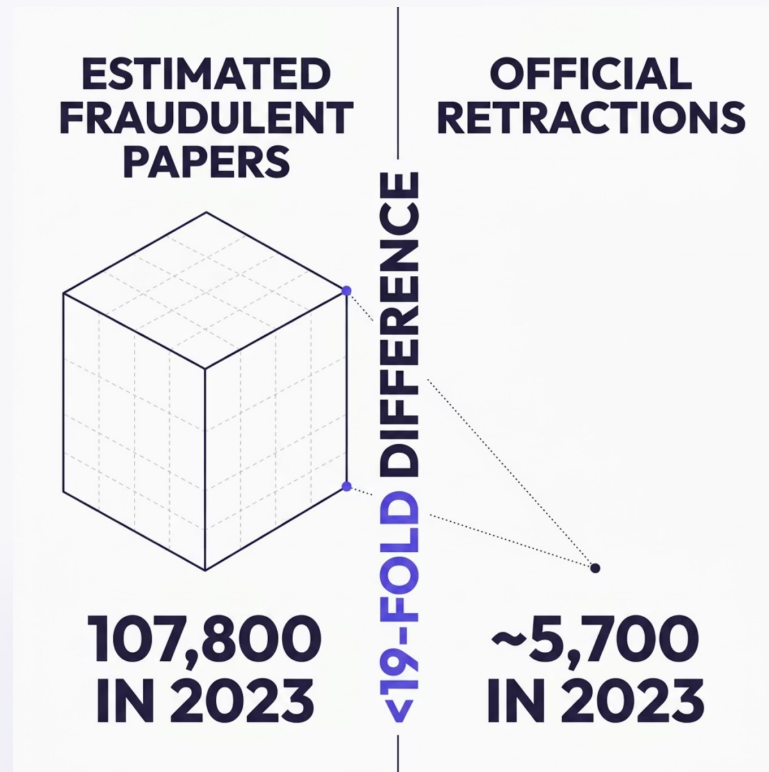
16.3%

Flagged in 2023

The proportion increased — suggesting the problem is growing.

Estimations based on Study 9

Using **Bayes' rule** — a method for estimating probability — the authors calculated how many flagged papers are *actually* fake.



5.8% of all biomedical papers

Are estimated to be fraudulent — a staggering proportion of the published record.

~107,800 fake articles in 2023 alone

This is **19 times higher** than the total number of official retractions that year.

Main goal: Awareness – Why Does This Matter?



Trust in Science

Fake papers corrupt the scientific record, eroding public and professional confidence in research findings.



Public Health Risk

False data can enter medical guidelines and drug research, potentially endangering patients.



Wasted Resources

Time, funding, and effort are spent chasing or building on fraudulent findings — a significant economic cost.

Why is the study great?

1. Paper mills pollute the scientific knowledge base with their business model but awareness and magnitude are still largely unknown. It is important to investigate the current state of fake papers infiltrating our knowledge base.
2. Paper used a series of clever approaches to estimate how big the problem of paper mill generated publication is on a larger scale
3. Despite their inherent biases, they used a validation dataset to correct for overcounting in their final estimation.
4. Model provides an approach for any field or journal (editor) to narrow down on paper mill publications and to enforce/suggest retraction and potential early detection mechanisms.

Why is this study problematic?

1. Designed as a self-fulfilling prophecy - but how else can you classify and identify "fake publications" ? There is almost no objective way - you need to operate on assumptions and test their validity.
2. The indicators used in later studies were derived from earlier suspicious-paper samples. Set of suspicious papers —> Extract patterns —> Build classifier —> Classifier finds papers with same patterns in new cohort (Confirmation bias)
3. 400 older papers that predate paper mills, that predate the rise in OA publications and Gmail account availability. Therefore, the validation dataset is not truly independent.
4. Study heavily relied on metadata proxies (suspicious publication profiles but no proven false data or false studies)
5. Confounders with global scientific inequality (many indicators strongly correlate with regional publishing culture, institutional wealth, degree of internationalization) Conflating risk: Mixing up low-resource science with suspicious science
6. Ethically controversial: Suspicious country as positive risk factor leads to geographic stigmatization, circular reasoning, and algorithmic bias amplification
7. No internal / external validation attempt, not even on a randomized subsample (in the field of authors' expertise). Could have revealed hard factors from a scientific side. They could have tried tortured phrases. Such a connection would have made a much stronger impact and given their approach more credibility. Without causal demonstration, findings are just correlations.
8. 11,5% false detections on a "selected historic sample" generates high numbers of potentially incorrect flags, making estimates very unstable.
9. Validation dataset was balanced, but real world data are not. The predictive performance changes with prevalence. Application of the same FDR and sensitivity to the entire biomedical literature is "statistically delicate issue"



The Big Picture: A Systemic Problem

The "publish or perish" culture creates powerful incentives for researchers — especially early-career ones — to cut corners.

The Root Cause

Career advancement tied to publication counts, not quality. Buy-in: publications, bribing editors

The Opportunity

We now have scalable tools to detect fakes. Education, sur

The Goal

Reform incentives and strengthen peer review to make fraud harder to commit. Sanctions for authors, editors and journals.

Danger

Paper mills are undermining the (unpaid) volunteer peer review process. Generative AI will accelerate and potentiate this

Discussion Questions

1 Peer Review Under Pressure

How does the mass production of fake papers change your trust in the peer-review system? Can peer review alone ever be enough?

2 AI & Academic Integrity

If AI can produce plausible-looking fake research, how should we change the way we evaluate student or academic work?

3 Who Is Responsible?

Is it the responsibility of journals, authors, or institutions to police "paper mills"? Where should accountability lie?

Protecting the Future of Science – Stockholm Declaration

- (i) Academia should resume control of publishing using non-profit publishing models (e.g. diamond open-access).
- (ii) Adjust incentive systems to merit quality, not quantity, in a reputation economy where the gaming of publication numbers and citation metrics distorts the perception of academic excellence.
- (iii) Implement mechanisms to prevent and detect fake publications and fraud which are independent of publishers.
- (iv) Draft and implement legislations, regulations and policies to increase publishing quality and integrity.

Fake publishing - the greatest scientific fraud

Publishing

🕒 5 minute read 📅 05 November 2025

Science is built on integrity and trustworthiness so when a flood of low-quality or fraudulent papers enter the system unchecked, trust in science is damaged. **Bernhard A Sabel** at the Otto-von-Guericke-University of Magdeburg, Germany and **Dan Larhammar** at Uppsala University, Sweden present the '**Stockholm Declaration**' for the 'Reformation of Science Publishing' published in *Royal Society Open Science*.



Made with GAMMA