COMMENTARY

“Research paper mills”: A factory outlet for dubious research

SHUBHADA NAGARKAR

Abstract
The primary objective of any research, regardless of its domain such as health, technology, psychology, or any other subject, is to enhance the overall well-being of individuals. Rigorous processes are involved in conducting research ethically and in communicating its outcomes to society. However, as publishing research has become a mandatory requirement for career advancement and appointments, academics are resorting to several unethical practices to get standard work published quickly. Consequently, predatory publishing markets have emerged, which publish data that is falsified and fabricated, along with plagiarised textual matter. The emergence of “paper mills” is a further step in the corruption of research, where a group of persons or automated systems generate papers for publication. Anyone desirous of publishing a paper can purchase one, akin to any desired fast-moving consumer product, with the added guarantee of publication in indexed journals. Therefore, paper mills and their unethical modus operandi are discussed in this paper in detail, with relevant examples. The article unfolds the consequences of publishing such fraudulent research papers and concludes with the challenges in combating paper mills.

Keywords: paper mills, dubious research papers, publication misconduct, machine generated papers

Introduction
Honestly and ethically conducting research, and transparently communicating its results to society, are rigorous and demanding processes. To carry out any original research, a researcher is required to formulate and conduct a study, undertake an exhaustive literature search, and uphold the integrity of experimentation and data collection processes. These endeavours demand significant time and concerted efforts. Following meticulous editorial scrutiny, the outcomes of such research are published in the form of a “research paper.” Thus, quality is the soul of the research.

Over the last few decades, the publication of research has been made compulsory for the recruitment and promotion of researchers and teachers in higher educational institutions. This has led to the emergence of the predatory journal market and exacerbated the production of bogus research papers [1]. Furthermore, during annual appraisals, the practice of assigning marks to papers published in high-impact-factor journals has led to the emergence of over 20 counterfeit journal impact factors [2]. Software programmes and ghostwriters are used to produce bogus papers that are either machine-generated or are written by groups of subject experts and are often coordinated by companies motivated only by monetary returns. The conceptual term for the mass production of research papers, either by people or machines, is “paper mills.” Analogous to grain mills that focus on large-scale flour production, research paper mills have emerged as a response to the “publish or perish” culture plaguing academia.

The existence of “term paper mills” has been documented in the published literature since the 1960s in the United States (US) and a few other countries. Term paper mills primarily cater to undergraduate students who wish to avoid spending hours in the library accessing and collating references. Consequently, purchasing readymade term papers became an easy way out, which engendered a profitable business for the term paper mills or companies manufacturing term papers [3–6]. At the post-graduation level, PhD theses are produced and made available without conducting research. According to Bik, paper mills produce scientific papers on demand [7]. She cites the case of medical practitioners in China who buy research papers, as they are plagued by a dearth of time, material, and hypotheses for actual research [7]. Gaby, a reviewer of biomedical literature with over 40 years of experience, flagged the publication of fabricated data in natural medicine papers over the past 10–15 years. He particularly underscored the hurried reporting of clinical trial results and related issues with examples [8].

The contentious issue of research paper mills came into the limelight again following the retraction of 68 papers published in RSC Advances by the Royal Society of Chemistry in the United Kingdom [9]. After a detailed and careful investigation, the executive editor of this journal not only
found similarities in texts but in charts and titles as well [9]. Pérez-Neri et al, as reported in the Retraction Watch database, traced 306 retracted papers that were associated with the term “paper mill” [10]. Analysis indicates that roughly one-half of the retracted papers were related to the field of health sciences, including medicine. It also shows that several such papers were published in China [10].

The continued increase in the number of paper mills is also attributed to the high rejection rates of credible journals, in addition to the pressure to publish to secure promotions, pad one's curriculum vitae, and inflate one's citation count.

**Modes of operation**

Paper mills primarily profit from “writing” fraudulent research papers for bogus researchers; it is very unfortunate that there are thousands of such establishments [11]. Paper mills typically utilise two methods to generate fake research papers: a) employ persons who may or may not be experts in the field; or b) use software programmes.

Hvistendahl analysed and explained the working methods of China’s Publication Bazaar [12]. Ter Smut Clyde and TigerBBB (both pseudonyms) have indicated another mode of operation, where a few people cut and paste pieces of information from published papers, submit the document as a new paper, and subsequently, employ/provide “experts” to facilitate its publication in journals that have a good impact factor. Ter Smut Clyde and TigerBBB have also documented highly sophisticated paper mills that conduct actual animal experiments, collect and analyse data, and write manuscripts [13].

**Customised paper providers**

Several companies provide customised research papers, viz http://papersowl.com/ and https://www.sharkpapers.com/, to name two of many. Russian Paper Mills by International Publisher LLC is one of the largest companies that offers scholars the option of purchasing co-authorship in hundreds of articles [14]. http://123mi.ru/ and http://123mi.ru/1/ are Russian websites where one can sell or purchase papers.

**Software programmes**

Software programmes (both online and offline) designed to generate research papers are offered by several companies through easily accessible platforms. These programmes provide sample essays based on keywords to entice customers to place an order for a paper. Two such companies are Essaybot and Drassignment [15,16]. Google lists several companies that produce papers based on keywords. They provide personal assistance to improve the essays and even assign writers or students with masters or doctoral degrees to work on client requirements. In addition, these companies provide free services to check and correct plagiarism, spelling, and grammar, and they often offer discounts on the first paper. The charges are based on the number of words, tables, figures, and images used in the production of the paper. Scigen and Mathgen are machine-generated programmes that generate professional computer science and mathematics papers. Machine-generated fake papers are difficult to distinguish, but their out-of-context language often aids in the identification of papers accepted at conferences without proper screening, leading to retractions by several journals [17].

**ChatGPT and paper mills**

Recently developed artificial intelligence (AI)-based tools such as ChatGPT, Bard, and Human, which have been discussed in the literature, are used not only for generating papers but also in the peer review process [9]. Van Dis et al discussed five priorities in research and the use of ChatGPT or AI tools. They listed their relative advantages and disadvantages and identified several errors in text generated by ChatGPT. After conducting experiments with ChatGPT to create text in the field of medicine, concerns were raised regarding the validity of the content, the references generated, and the authors’ responsibility for the results [18]. Stokel-Walker raised concerns about ChatGPT being listed as an author and highlighted examples of a few medical journals in nursing and oncology in which ChatGPT has been listed as one of the authors [19]. He carried out conversations with the editors of a few journals, who clearly indicated that the use of such artificial authors would call into question the integrity of research and raise concerns about the use of plagiarised text and false data [19]. Paper mills will exploit this AI text-generation tool to create fraudulent papers.

**Detection**

Some of the tactics used by experts to identify research papers generated by paper mills are the duplication of graphics, the use of meaningless text, questionable peer-review processes, and similarities in grammatical structure [20,21,22].

Bik has compiled a list of over 400 papers from paper mills in China that she has dubbed “tadpole paper mills” [7]. Machine-generated “Western blot” images are another reason for paper retraction. The Western blot is a laboratory method used to detect specific protein molecules in a mixture of proteins associated with a specific tissue or cell type [23]. Bik has visually screened Western blot images from 20,621 papers published in 40 scientific journals from 1995 to 2014. Using in-house and open-source software, she located 3.8% duplicate and altered images [20]. Despite this, she still felt that human eyes remained superior to any other tool. Recently, Bik identified “Iranian plant paper mills” which contains plant sampling locations that are shared between 70 papers². Papers shows same coordinates but names of cities, provinces and regions are different. Additionally, she has presented citation rings (groups agreeing to cite each other’s papers) in these papers as evidence, where common papers have been cited to increase the citation index of authors [24].
Byrne and Christopher emphasise that many milled papers use readymade templates from credible journals to maintain consistency, making it difficult to identify data fabrication. In addition to counterfeiting layout and design, they misuse micrographs, photographs, scatter plots, bar graphs, stock images, and numerical datasets to prepare papers [25].

In his preprint, Day revealed that authors create fake email addresses for reviewers prior to journal submission. So, when the journal sends the paper for review to these addresses, it returns to the author, who then reviews their own paper. In some cases, a group of people peer review each other’s work to make the process easier. Many paper editors are unaware of the fraud perpetrated by paper mills through the creation of fake papers, authors, and fake reviewers. Day has questioned the efficacy of the peer review system of these journals and equated the paper mill fraud with cancer [26].

Language usage is another indicator of fraud and can be identified via the use of identical phrases, paraphrasing, and context-free grammar. Filion reported a very interesting example of this when he discovered a collection of disturbingly similar scientific papers. He found 25 papers published by Chinese doctors in which they used similar structures and figures in the same order. The last figure in each paper is the “Beggar funnel plot.” This term does not exist but is a derivative of the names of two statisticians — Colin Begg and Matthias Egger. Both gave the name to a test for publication bias [27].

Teixeira discusses three paraphrased fake papers in credible journals related to cancer, citing machine-generated papers by Pan et al and Lue et al. The former is on prostate cancer in females and the latter on ovarian cancer in males, which is self-contradictory as prostate glands are not found in females and ovarian structures are absent in males. Teixeira was deeply concerned by the fact that credible journals had published such fake articles [28].

Cabanac et al, in a preprint on arXiv, exposed the context-free grammar in machine-generated papers using the GPT-2 Output Detector. They found several meaningless, out-of-context, and irrelevant sentences [29].

It has been observed that many editors seem unable to detect such malpractices, allowing fake or fraudulent articles to pass their stringent vigilance. Even reputable publishers such as Springer, Taylor & Francis, Wiley, and the Public Library of Science are vulnerable to the machinations of paper mills [7].

The scenario in India

The scenario in India is rendered highly complex by prevalent socioeconomic factors. Predatory journals, the quality of doctoral studies, and research in India have been debated since 2006, particularly after the Sixth Pay Commission in India [30,31]. Predatory journals as well as paper mills have proliferated in India after the publication of research was made mandatory for the appointment and career advancement of teachers in universities as well as in medical colleges. As per the gazette notification by the National Medical Council (NMC), the publication of research papers is a prerequisite for the direct appointment of professors and associate professors, as well as for career advancement in medical colleges [32]. Furthermore, once candidates secure faculty or equivalent positions in institutions, publications are a vital driver for their further progress, once again affording golden opportunities to paper mills. Several online paper mills exist in India, which help doctors select thesis topics as well as publish research papers.

The mentioned Russian website (http://123mi.ru/) lists 344 papers accepted for publication by journals from different countries where authorship is available for a fee, among which, 73 papers (the highest) belong to journals based in India. Many of these papers are from biomedical, engineering, and other fields of science [33].

Consequences

The numerous effects of paper mills on research have been observed and documented in the published literature. Fraudulent papers with out-of-context text, particularly in science and technology, pose a veritable threat to the integrity of academic research and are detrimental to society. Editors and reviewers work under the assumption that authors have conducted research honestly and have reported it correctly; therefore, any abuse can cause widespread disruption. Sabel noted that the loss of trust will result in confusion and pose a great risk to health and humanity at large, in addition to incurring a complete waste of time, money, and energy [34].

Additionally, it has been noted that paper mills target journals indexed in highly credible databases. These facilities generate counterfeit research papers and strategically submit them to journals they anticipate or believe will approve them. Once accepted, such fabricated papers are extensively referenced in other deceptive publications. Sometimes, unscrupulous editors assume control of special issues of journals to flood them with fake papers that cite one another, thereby artificially increasing the citations and status of these fraudulent studies. Bricker-Anthony and Herzog highlighted that this malpractice in citing papers helps increase the journal’s impact factor. Moreover, open-access journal publishers collect publication fees, helping authors publish “high impact” papers and get promoted, while institutions celebrate high productivity and professors may receive high-ranking positions [35].

The retraction of papers is on the rise, especially of papers produced by paper mills. Mass retraction by leading publishers indicates serious problems in the peer-review process. Flynn has listed such retraction cases by Wiley and Hindawi publishers, due to which 50+ journals were delisted by Clarivate [36]. Retraction Watch maintains a database of
retracted papers, with “paper mills” listed as one of the criteria for retraction. Candal-Pedreira et al identified 1,182 retractions of milled papers from the Retraction Watch database, and the results show that the number of such papers is increasing, especially from China, and is linked with specific hospitals. More global efforts are needed to identify such papers because they disseminate fabricated images and data, which will adversely affect future research [37].

Citations to retracted papers, even if received before the retraction, can potentially misrepresent metrics such as journal impact factors and rankings of journals, authors, and institutions, etc. Bolland and Grey investigated the citations to retracted papers and noticed that citations are from unreliable sources. When publications cite retracted articles, reliability is compromised, affecting readers, researchers, patients, and clinicians [38].

Open Researcher and Contributor ID — ORCID is a global non-profit organisation developed by a dedicated knowledgeable professional, which determines author credibility. ORCID builds transparent and trustworthy connections between researchers, their contributions, and their affiliations by issuing a unique identifier. Many leading publishers as well as research institutions and libraries are members of ORCID and support the creation of a permanent, clear, and unambiguous record of research and scholarly communication (https://orcid.org/). Unfortunately, paper mills have exploited ORCID iDs and created ghost author iDs, which render the authors unverifiable and disrupt the system [39].

Solutions

Eradicating paper mills poses a considerable challenge due to the integral role of research and publications in the career advancement of faculty members. Instead of focusing solely on the quantity of research papers, a shift towards evaluating the quality of research and publications could be considered. Given that faculty members play a pivotal role in fostering highly skilled human resources and aid in departmental progress, their overall contributions could also be included as assessment criteria.

To detect fake papers generated by paper mills, researchers should examine the original dataset. There are several data repositories available, especially where researchers can deposit their research data in the public domain — for example, the Open Science Framework (OSF) data repository by the Centre for Open Science (COS). These open datasets can be reused for further research [40]. Open data repositories enhance transparency in data, foster collaboration across institutions, and increase the visibility of work. The most significant feature of these repositories is that the data remains in the public domain for verification and is preserved for future use. Several such repositories are available for researchers.

Publishers must clean up their publication lists by retracting such milled papers. This can be done by hiring research integrity experts to check for duplicate images and text that lacks context [34]. The government and regulatory authorities of each country must take steps to curb paper mills and invest funds in developing tools to detect such papers.

Awareness programmes for researchers and graduates should be organised to shed light on unethical research practices. All educational institutions should have strict leadership at the senior and expert decision-making level.

Flynn suggested that funders, academic institutions, researchers, and research integrity officers, must work in collaboration to address the issues of paper mills including research integrity [36].

Researchers should keep themselves updated with the latest publishing practices, for which they can consider subscribing to the following blogs:

- Research Integrity Digest (https://scienceintegritydigest.com/)
- For Better Science (https://forbetterscience.com/)
- Image Integrity (https://image-integrity.com/)
- Retraction Watch (https://retractionwatch.com/)
- Committee on Publication Ethics (COPE) (https://publicationethics.org/resources/research/paper-mills-research)
- Scholarly Kitchen (https://scholarlykitchen.sspnet.org/)

Conclusion

The rise of research paper mills and machine-generated research presents a serious threat to academic integrity. These fraudulent companies sell plagiarised or fabricated papers, making it difficult to distinguish genuine research papers from fake ones. Addressing this issue requires collective action from worldwide stakeholders within the academic community. Authors must uphold research integrity, while peer reviewers and editors must be cautious in identifying and rejecting fraudulent papers. The scientific community, as a whole, must prioritise ethical research practices and support initiatives like the Committee on Publication Ethics (COPE) and Retraction Watch to combat paper mills.

In India, there is an urgent need for training programmes to develop experts in “research integrity” who can effectively detect fraud in research. Sensitisation and awareness programmes should be implemented to educate researchers at all levels about detecting research fraud. The establishment of open repositories for raw data, the development of detection tools, and changes in how research output is measured are essential steps forward. Online forums can help expose papers generated by paper mills and contribute to their dismantling.