Paperpal Preflight

Safeguarding Research Integrity

Using AI Tools and Human Insights to Overcome Fraud in Research

APRIL
2024

Nishchay Shah
Shilpi Mehra
Kriti Gupta
Minhaj Rais

Table of contents

Executive Summary

At CACTUS Communications, we believe that maintaining research integrity is crucial for progress in science. This white paper addresses the growing challenges with maintaining robust research integrity. In light of the widespread adoption of generative AI within the scientific community, this paper proposes a hybrid model that merges human insight with AI efficiency to tackle the issue of questionable research practices.

Nishchay Shah
Chief Technology Officer & Head, Emerging Products
Cactus Communications

Shilpi Mehra
Head, Research Integrity,
Cactus Communications

Kriti Gupta
Vice President, Institutional Marketing
Cactus Communications

Minhaj Rais
Associate Director, Strategy and Corporate Development,
Cactus Communications

Our argument is grounded in CACTUS's extensive experience in science and technology communication spanning over two decades. The proliferation of "fake science," fueled by advancements in AI, poses a significant threat to the credibility of scientific publishing. While AI offers efficiency, it also opens avenues for fraud and manipulation, making human oversight indispensable.

We analyze various integrity-checking models and conclude that the hybrid human-AI approach is the most effective and feasible solution. Our stance is not against technological progress but rather acknowledges the distinct strengths of both human and AI domains. Human expertise remains crucial for ethical judgment, deep understanding of complex issues, and maintaining scholarly authenticity.

Ensuring research integrity is not solely the responsibility of individual researchers but of the entire scientific community, including publishers, institutions, funders, and stakeholders. Beyond ethical considerations, there are significant business implications, as credibility directly impacts trust and reputation. Moreover, the consequences of fraudulent science extend beyond academia, affecting society at large.

The proposed hybrid model represents a harmonious integration of human and AI capabilities, envisioning a future where technology advances while anchored in human integrity. It promotes a science that is progressive yet grounded, innovative yet integral, and ultimately, reliable like the human spirit propelling it forward.

Science is Under Siege

Research integrity, the backbone of scientific progress, faces a rising adversary – fraud. Plagiarism, fabrication, and data manipulation are becoming increasingly common in scholarly work, posing a threat to the reliability of research, funding allocation, and public trust.

Traditional Peer Review processes, a cornerstone of research, struggle to detect sophisticated misconduct.

The problem is worsened by the explosive growth of papermills that mass-produce fake research for profit. The rise of generative AI creates a complex problem – it refines the sophistication of fake research, while churning out vast amounts of fraudulent material.

An article published in Nature Magazine in November 2023 titled "How big is science's fake-paper problem?"[1] indicated that of all scientific papers published in 2022, 1.5% closely resemble papermill outputs. This is reflected in the rapidly growing number of retractions recorded on the Retraction Watch[2] database that has over 46,000 retractions as of January 2024 (over 10,000 added just in the last year!).

A Bumper Year for Retractions

Retraction notices in 2023 have passed 10,000, largely because of more than 8,000 retractions by Hindawi.

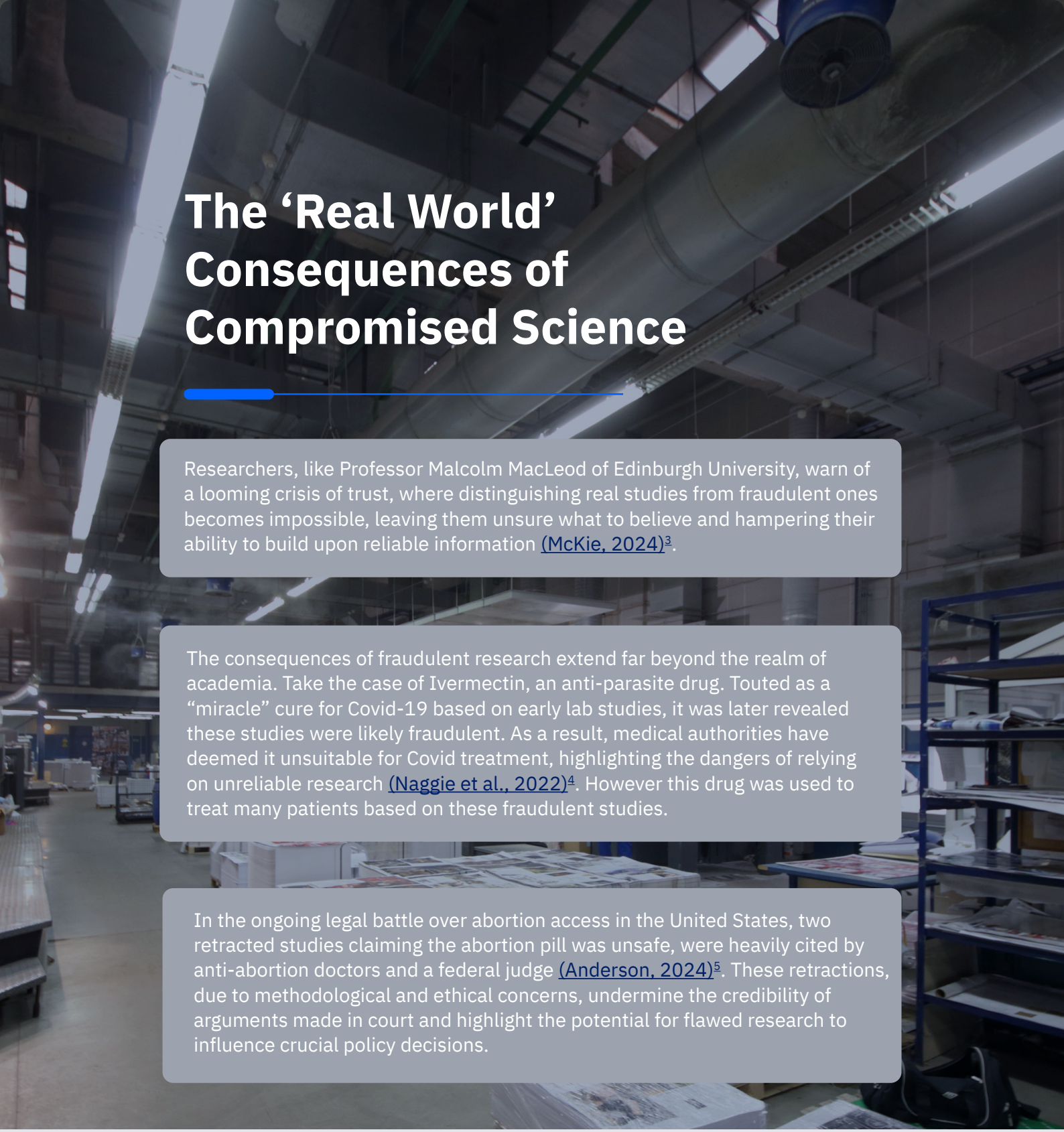
Figure 1: More than 10,000 research papers were retracted in 2023 — a new record (Van Noorden, 2023)[2]

Papermills: Churning Out Fake Research for Profit

In academic publishing, papermills represent a growing concern, engaging in the production of fraudulent research for financial gain. They operate like factories, churning out fake research papers for researchers willing to pay.

Moreover, the rise of AI technology has worsened this problem by enabling more sophisticated methods of generating fake research. AI algorithms can now produce convincingly structured papers with plausible-sounding arguments and citations, further blurring the line between authentic and fabricated research. These ghostwritten works lack real data and proper methodology, yet secure quick publications and inflate CVs.

As a result, detecting fraudulent papers becomes increasingly challenging, amplifying the impact of papermills on the integrity of academic publishing and the credibility of honest researchers.



The ‘Real World’ Consequences of Compromised Science

Researchers, like Professor Malcolm MacLeod of Edinburgh University, warn of a looming crisis of trust, where distinguishing real studies from fraudulent ones becomes impossible, leaving them unsure what to believe and hampering their ability to build upon reliable information (McKie, 2024)[3].

The consequences of fraudulent research extend far beyond the realm of academia. Take the case of Ivermectin, an anti-parasite drug. Touted as a “miracle” cure for Covid-19 based on early lab studies, it was later revealed these studies were likely fraudulent. As a result, medical authorities have deemed it unsuitable for Covid treatment, highlighting the dangers of relying on unreliable research (Naggie et al., 2022)[4]. However this drug was used to treat many patients based on these fraudulent studies.

In the ongoing legal battle over abortion access in the United States, two retracted studies claiming the abortion pill was unsafe, were heavily cited by anti-abortion doctors and a federal judge (Anderson, 2024)[5]. These retractions, due to methodological and ethical concerns, undermine the credibility of arguments made in court and highlight the potential for flawed research to influence crucial policy decisions.

Compromised research integrity can impact judicial rulings, public policy, medical practices, and societal debates. It is imperative that scholarly publications maintain rigorous standards to ensure the dissemination of reliable, accurate information. Failing to do so, can have serious ramifications, from misinforming decision-makers to eroding public trust in scientific findings. Additionally, it hinders researchers’ ability to distinguish good science from bad, impacting how new research ideas are developed.

Overburdened Editorial & Peer Review Systems

Traditional peer review systems are increasingly proving inept at detecting sophisticated fraudulent research. Subjectivity and potential bias, limited expertise in identifying markers of fraudulent activity, and above all, constraints of time and workload, collectively compromise the effectiveness of this conventional approach (Fang et al., 2012)[6].

Recognizing these limitations, research integrity teams within publisher ecosystems play a pivotal role by conducting regular audits and investigations. Leveraging human reviews, these teams work diligently to identify compromised journals and processes, serving as a crucial checkpoint against the infiltration of fraudulent research into scholarly publications (Resnik & Elmore, 2016)[7].

However, it's essential to acknowledge that these teams, designed for thorough examinations, may struggle to cope with the scale at which fraud is occurring today. This emphasizes the urgency to adopt innovative and scalable solutions to fortify research integrity across the evolving landscape of scholarly publishing.

Empowering Editorial Workflows

As a response to the growing need of detecting fraudulent activity at scale, several AI solutions are attempting to identify markers of compromised research integrity early on. AI tools can highlight problematic issues to editorial desks, making it easier to decide which manuscripts to reject outright, which ones to focus on for further investigations, and which ones to take forward for peer review.

Benefits of AI Checks vs Human Checks

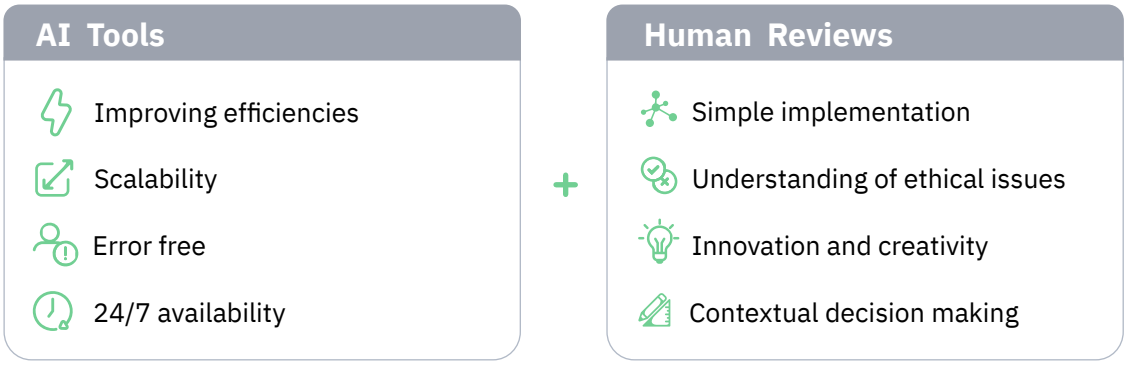


Figure 2: AI and Human Checks have complementary benefits.

An exclusive reliance on either AI tools or human review proves impractical for accuracy and scale. While AI excels at processing vast datasets swiftly, it may miss nuanced human judgment and/or introduce biases in decision making. Also, AI cannot pick out new/developing patterns of fraudulent activity. On the other hand, relying only on human reviewers can be time-consuming. Also, the process is susceptible to oversight, is resource intensive, and often impractical —especially for large publishers, since scalability remains a huge problem.

By combining the strengths of AI and human expertise, a Hybrid approach offers a more comprehensive and efficient solution. AI can rapidly scan large volumes of data, flagging potential issues, while human reviewers bring nuanced understanding and contextual insight. This collaborative strategy can help in fortifying the defenses of scientific publishing against fraudulent research, ensuring a robust and adaptable system for maintaining research integrity. This, in turn, ensures that publishers can maintain the highest levels of authentic research content, thereby safeguarding their credibility and reputation.

Enhancing Research Integrity:

Assessing Current Solutions to Combat Fraudulent Activity

To counter fraudulent activity in research submissions today, several AI solutions are being designed for detecting specific issues like AI-generated text or image duplication, showcasing proficiency in these specific domains. While these tools represent significant advancements in research integrity, they often lack comprehensive coverage across the spectrum of fraudulent activities, leaving potential gaps in detection capabilities. Additionally, the rapid evolution of deceptive strategies poses a challenge, as these tools may struggle to adapt quickly to newer types of fraudulent tactics.

The comparative analysis below aims to illustrate the diversity of checks offered by leading research integrity solutions available today (across 10 consolidated categories). It does not intend to provide an exhaustive representation of the suite of capabilities offered by each solution. Rather, the comparison serves to highlight that no single solution encompasses all aspects required to safeguard research integrity effectively. It underscores the need for more comprehensive and integrated approaches that leverage the strengths of various solutions to address the multifaceted challenges of ensuring research integrity.

Comparative Analysis of Integrity Tools

Category	AIRA	Dimensions	ImageTwin	Morressier	Preflight for Editorial Desk	Proofig	Signals	STM Integrity Hub
AI-generated/ Fabricated Content	✓			✓	✓		✓	✓
Authorship	✓	✓		✓	✓		✓	✓
Disclosures/ Transparency	✓	✓		✓	✓			
DOI Analysis				✓	✓			✓
Duplicate Submissions								✓
Human Reviews	✓				✓			
Image Duplication	✓		✓			✓		
Manuscript-level Risk Rating					✓		✓	
References/ Citations	✓			✓	✓		✓	
Retractions	✓			✓	✓		✓	✓

*Disclaimer: This information is derived from online sources accessed on March 15, 2024, and may not reflect latest updates.

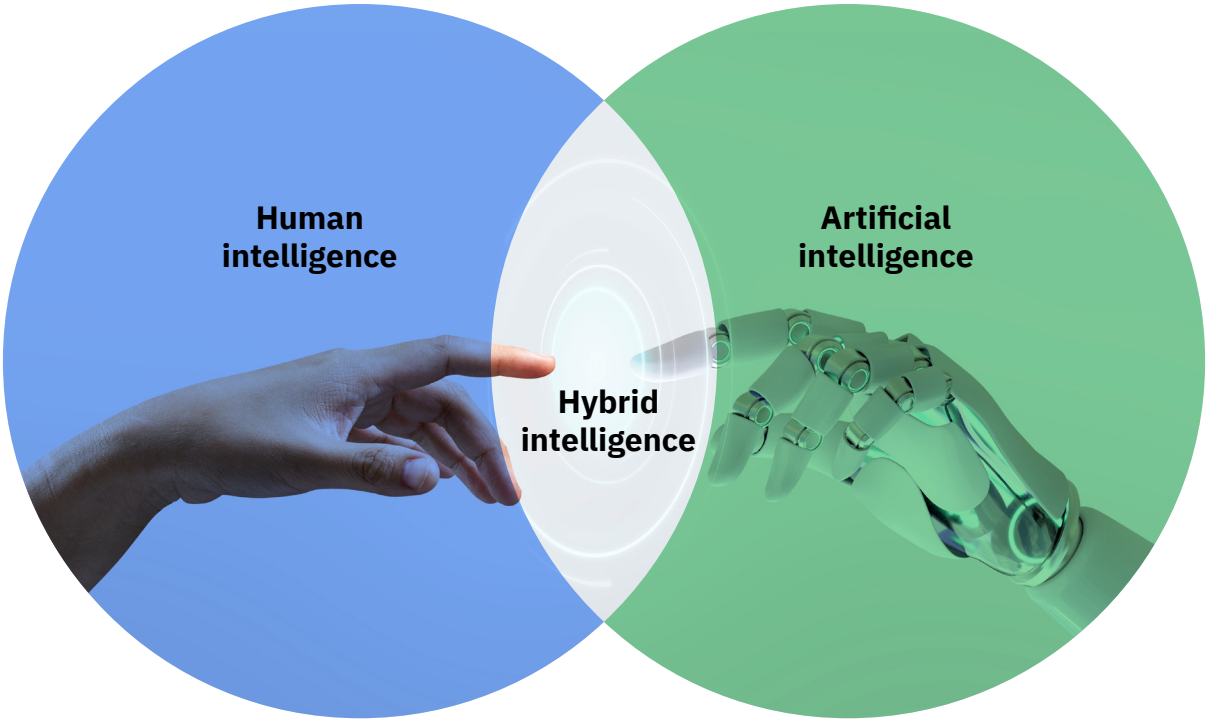
*Figure 3: Comparison of checks offered by leading research integrity solutions today**

The Hybrid Approach:
AI checks + Human Review

The hybrid AI + human checks approach emerges as a practical solution for tackling the problem of fraudulent research by addressing the inherent shortcomings of an exclusively AI-powered solution. By incorporating a human loop, not only does this model bring in adaptability, critical thinking, and contextual understanding to the table, but it also serves as a vital audit for AI checks.

Human reviewers possess the capability to discern new patterns and trends in fraudulent activity, enhancing the system's capability to evolve with emerging deceptive strategies. Moreover, the human loop plays a critical role in validating accuracy and reducing bias, thereby ensuring that the model consistently improves and refines its ability to detect and deter fraudulent research over time.

Combining AI checks and human expertise helps in creating a flexible and dynamic approach to research integrity checks, overcoming the limitations of exclusively AI-powered solutions.



Benefits of Using a Hybrid Approach

The adoption of a Hybrid AI + Human Checks approach presents many advantages, reshaping the landscape of research integrity and publication practices.

Early Detection of Fraudulent Research: By integrating AI and Human checks, the Hybrid approach identifies and filters out fraudulent manuscripts even before they undergo the traditional peer review process. This helps save an enormous amount of time and resource that would otherwise be allocated towards detecting compromised submissions. This early detection not only safeguards the reputation of journals but also upholds editorial integrity, fostering a culture of ethical research practices. The collaborative effort ensures a more robust defense against deceptive submissions, contributing to the overall reliability of scientific literature.

Reduced Burden on Peer Reviewers and Editors: Catching fraudulent research early in the publication process alleviates the burden on peer reviewers and editors, thereby streamlining the evaluation process and leading to faster publication times. By automating certain aspects of the initial screening, reviewers can focus on more nuanced and critical assessments, enhancing the efficiency of the overall publication workflow.

Faster Publication Timelines: Reduction of workload for peer reviewers and editors, frees up their time to focus on processing more manuscripts and leads to quicker dissemination of valid and trustworthy research findings. This would be a huge win for both authors and publishers struggling with the current pace of publishing timelines.

Scalable Option for Larger Publishers: A Hybrid approach allows larger publishers to scale up research integrity initiatives on demand. For example, typically, a publisher might not be able to expand the scope of their research integrity operations from 50 to (say) 100 journals, without hiring significant additional resources. This changes with a Hybrid approach because suspected manuscripts red-flagged by AI + Human checks, can be rejected right at the outset, freeing up time for editorial staff.

Improved Brand and Reputation for Publishers: A renewed confidence in the integrity of scientific publications is pivotal for enhancing the brand value and overall credibility of publishers. Sound and strong ethical publishing practices can go a long way in building trust and safeguarding the reputation of publishers struggling in the face of an exponential increase in the number of retractions.

Increased Research Integrity and Trust in Scientific Publications: Ultimately, a Hybrid approach can contribute to an overarching enhancement of research integrity and trust in scientific publications. By combining the strengths of AI and human expertise, the system reinforces the credibility of the scholarly ecosystem, reassuring both researchers and the public that published research has undergone a thorough and reliable vetting process. This, in turn, propels scientific advancement and leads to faster solutions and breakthroughs, which is at the core of all scientific endeavors.

Paperpal Preflight for Editorial Desk (PPE):

An example of the hybrid approach

Realizing the benefits of a hybrid approach for effectively upholding research integrity in scientific publications, CACTUS has developed an in-house solution for publishers and societies to support their editorial desks – Paperpal Preflight for Editorial Desk (PPE).

PPE provides automated research integrity checks, complemented by IntegrityGUARD human reviews through a global team of PhD-level specialists, trained in identifying fraudulent behaviors. This CACTUS offering showcases how AI efficiencies and human expertise can be combined to provide a scalable, accurate, and holistic solution to detect fraudulent activity in scientific manuscripts.

- Preflight's AI assesses the manuscript's research integrity, language quality, and technical compliance with submission guidelines.
- If results of the integrity checks are inconclusive ("warning"), the editor can request an IntegrityGUARD human check by expert PhD reviewers for a conclusive rating of "pass" or "critical".
- After analyzing the results, editors can record their assessment and download an edited version of the submission. This supports clear and informed decision making about whether or not to take a manuscript ahead for peer review.

How PPE Works

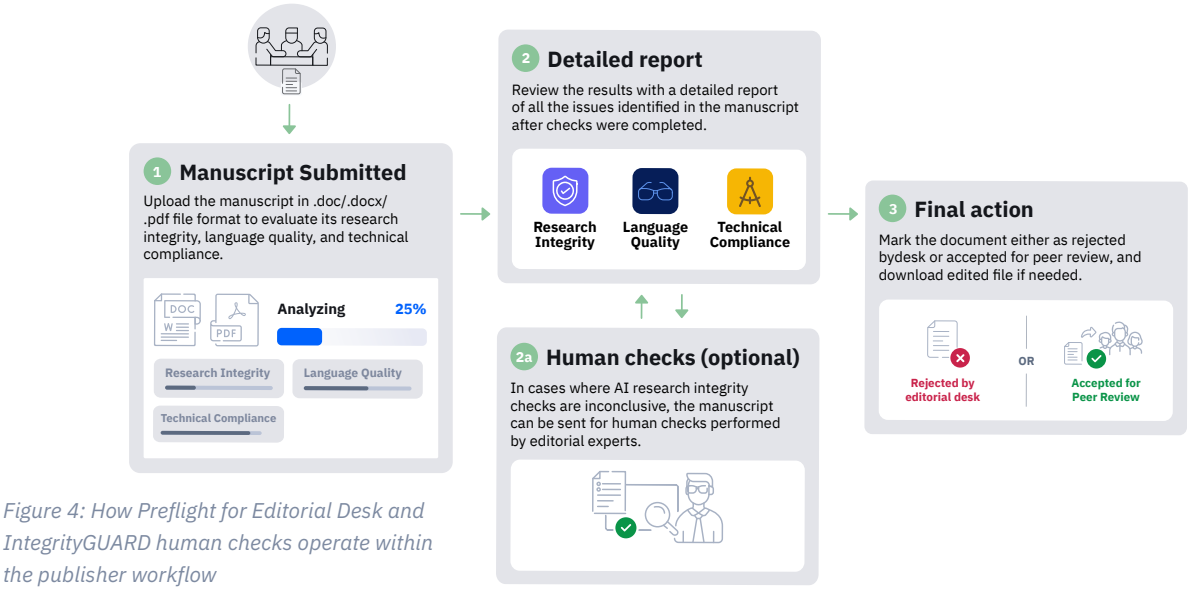


Figure 4: How Preflight for Editorial Desk and IntegrityGUARD human checks operate within the publisher workflow

The hybrid solution from Paperpal Preflight for Editorial Desk and IntegrityGUARD can be used for abstracts, conference proceedings, and journal workflows, to detect compromised content and support publishers with maintaining the highest standards of integrity in scholarly communication.

PPE automated Research Integrity checks have been developed in-house at Cactus Communications. The tool runs 20+ checks for multiple markers of research integrity (including papermill detection), assigns a weighted score to the result of each check, and generates a composite score at the end – which conclusively states whether a manuscript has an integrity rating of Pass/Warning/Critical.

Paperpal Preflight for Editorial Desk: Research Integrity Checks

Transparency	Papermill Activity	Science
Ensures research openness by verifying data access, ethical practices, and funding sources (need, presence and completeness).	*Identifies potential papermill activity through analysis of authorship, fabricated content, citations, and image duplication.*	*Analyzes article structure, content match between sections, journal scope match, and use of inappropriate language.*
• Data Availability Statement • Ethics Approval Statement • Patient/ participant consent • Funding information • Author contributions • Consent for images	• **Authorship:** Affiliations, previous retractions, number, geographic distribution • **Fabricated content:** AI Gen text, tortured phrases, nonsense phrases, LLM response trace • **Citations:** Excessive citations to the same lab/group, to unrelated work, to retracted work, and self-citations; reference URL validity • **Image Integrity:** Detect duplication of parts/whole images within the manuscript, garbled text within images	• Journal subject area match • Appropriate article sections • Abstract and article content match • Reference age distribution • Inappropriate language (sexism, racism)
Composite score of 6 Transparency markers	Composite score of 11 Research Integrity markers	Composite score of 5 Science markers

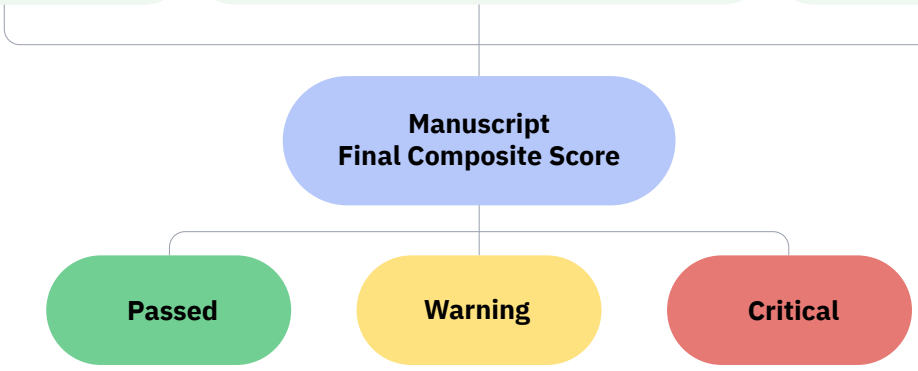


Figure 5: Checks included within each category and how these are scored to generate an overall integrity rating for the manuscript.

The hybrid approach, combining AI and human expertise, seems most effective for ensuring research integrity. AI can efficiently flag potential issues in submissions, while human judgment remains integral for thorough review and investigating suspected fraud. Leveraging AI enables human experts to focus efforts more efficiently, thereby effectively filtering out compromised research.

Paperpal Preflight is a game-changer for publishers and societies seeking to maintain research integrity while improving turnaround times. Our hybrid AI and human-powered solution empowers editorial teams to work more efficiently and effectively, ensuring that high-quality research findings reach the global scientific community with greater speed.

Nishchay Shah
Chief Technology Officer &
Head, Emerging Products
Cactus Communications

Conclusion & Future Outlook

In conclusion, as we navigate the evolving landscape of scientific inquiry, the importance of preserving research integrity cannot be overstated. While traditional peer review systems and research integrity teams within publisher ecosystems continue to play vital roles in safeguarding research integrity, the scale and complexity of fraudulent activity demand innovative solutions. Hybrid approaches that combine AI technology with human expertise, such as Paperpal Preflight for Editorial Desk, show great promise in addressing these challenges.

AI-powered tools have an immense potential to enhance editorial workflows and elevate the quality of scientific publications. The continuous refinement and integration of artificial intelligence into editorial processes holds promise for more efficient and rigorous vetting mechanisms, ensuring the reliability of research findings.

This collaborative effort is not just a response to the challenges of today but a proactive step toward sustaining the trust that underpins the progress of scientific knowledge and innovation. It is imperative that the academic and scholarly publishing communities work together to develop and implement robust, scalable solutions that uphold the integrity of research for generations to come.

By fostering a culture that combines the strengths of AI tools with human discernment, the scholarly publishing industry has an opportunity to collectively fortify its defenses against fraudulent research.

References

01 Van Noorden, R. (2023a). How big is science's fake-paper problem? Nature, 623(7987), 466–467. https://doi.org/10.1038/d41586-023-03464-x

02 Van Noorden, R. (2023b). More than 10,000 research papers were retracted in 2023 — a new record. Nature, 624(7992), 479–481. https://doi.org/10.1038/d41586-023-03974-8

03 McKie, R. (2024, February 3). "the situation has become appalling": Fake scientific papers push research credibility to crisis point. Available from https://www.theguardian.com/science/2024/feb/03/the-situation-has-become-appalling-fake-scientific-papers-push-research-credibility-to-crisis-point

04 Naggie, S., Boulware, D. R., Lindsell, C. J., Stewart, T. G., Gentile, N., Collins, S., McCarthy, M. W., Jayaweera, D., Castro, M., Sulkowski, M., McTigue, K., Thicklin, F., Felker, G. M., Ginde, A. A., Bramante, C. T., Slandzicki, A. J., Gabriel, A., Shah, N. S., Lenert, L. A., Dunsmore, S. E., ... Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV-6) Study Group and Investigators (2022). Effect of Ivermectin vs Placebo on Time to Sustained Recovery in Outpatients With Mild to Moderate COVID-19: A Randomized Clinical Trial. JAMA, 328(16), 1595–1603. https://doi.org/10.1001/jama.2022.18590

05 Anderson, K (2024, March 29). Supreme Retractions. The Geyser. https://www.the-geyser.com/supreme-retractions/?ref=the-geyser-newsletter

06 Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proceedings of the National Academy of Sciences of the United States of America, 109(42), 17028–17033. https://doi.org/10.1073/pnas.1212247109

07 Resnik, D. B., & Elmore, S. A. (2016). Ensuring the Quality, Fairness, and Integrity of Journal Peer Review: A Possible Role of Editors. Science and engineering ethics, 22(1), 169–188. https://doi.org/10.1007/s11948-015-9625-5

About CACTUS

Founded in 2002, Cactus Communications (cactusglobal.com/) is a science communication and technology company. We specialize in AI products and solutions that improve how research gets funded, published, communicated, and discovered. We solve problems for researchers, universities, publishers, academic societies, and life science organizations through innovative products and services developed under the brands Editage, Researcher.Life, Cactus Life Sciences, Impact Science, Mind the Graph, and Paperpal. CACTUS has offices in Princeton, London, Aarhus, Singapore, Beijing, Shanghai, Seoul, Tokyo, and Mumbai; a global workforce of over 3,000 experts; and customers from over 190 countries. CACTUS is considered a pioneer in its workplace best practices and has been consistently ranked a great place to work over the last several years.

Our Brands

Paperpal | Mind the GRAPH | Researcher.Life | Discovery By researcher.life | editage | CACTUS LIFE SCIENCES

CACTUS®

London | Princeton | Singapore | Beijing | Shanghai | Tokyo | Seoul | Aarhus | Mumbai

For more information
Visit www.cactusglobal.com or
email:shilpi.mehra@cactusglobal.com