AI Misconduct and Prevention in Biomedical Research

Group 6

MEDSCIEN9505

Dr. Tom Drysdale

July 12, 2024

Interdisciplinary Medical Sciences

Schulich School of Medicine and Dentistry

Western University

Assignment Acknowledgement

The Turnitin originality score was 22%. ChatGPT was used for preliminary brainstorming, primarily by testing prompts within the program to understand its algorithm and learned responses. Grammarly was also used for grammatical correctness.

Author Contributions

Student Name	Student Number	Contributions
Alyssa Fryer	251191894	Created Word Document, Notes on Google Doc, Strategy 3, Countermeasure for Strategy 3, References, Editing
Preetama Badyal	251163246	Developed overview and brainstorming notes on Google Doc, Strategy 2, Countermeasure for Strategy 2, Conclusion, and overall editing of the paper.
Rishika Sharma	251008648	Created Google Doc, Notes on Doc, Created Brief Overview, Strategy 1, Countermeasure 1, Rationale and Robust-ness of Strategy and Countermeasure 1, Editing

Part One

Overview

Artificial intelligence (AI) is a transformative power that is significantly changing society through the works of automation, increasing decision-making ability and rapid accessibility. The use of AI is enhancing many industries from marketing to healthcare. Several types of AI models are used to achieve various tasks within these sectors and more. Machine learning is an example of a model composed of a computer algorithm designed to analyze inputted data to gear predictions. The evolving nature of AI has advanced to a state where this technology can be modelled after the human brain. An example of such an AI model is neural networks, a subset of machine learning that conducts signal integration through interconnected artificial neurons. An additional subset of machine learning is deep learning through which vast computations form deep neural networks. These deep neural networks can learn expansive unstructured data. All these models play a role in quickly revolutionizing the landscape of biomedical research.

AI is proving to be an efficient resource in different aspects of research such as project design, data collection, data analysis, and more. Machine learning algorithms can be developed to cater to a specific research need. For example, Dr. Dijk of the Weizmann Institute of Science in Israel targeted an algorithm to measure the spread of breast cancer cells to novel areas in the body (Reitman, 2022). Alternatively, machine learning algorithms can currently analyze three-dimensional images for novel phenotypes of cardiac ailments (Reitman, 2022). Although deep learning uses neural networks, one may perceive the two to be the same model, however, the difference is neural networks are simpler and deep learning is complex. Deep learning analyzes complex datasets and extrapolates patterns; inputted data includes medical images, genomic sequences, protein structure and more (Cao et al., 2018). Neural networks analyze data, assist in the diagnosis and prediction of disease, drug discovery and image analysis (Fonseca, 2024).

Although AI is prompting significant advancements in biomedical research, it has also increased ethical misconduct. In an assessment of 7771 articles, 4.2% were self-reported and 27.9% studies were not self-reported (Phogat et al., 2023). Examples of AI misconduct include plagiarism, image generation, and data falsification and fabrication (Phogat et al., 2023). AI misconduct entails fabricating non-existent data which can then serve as a foundation to falsify scientific results (Elali & Rachid, 2023). Fabrication of data also includes altering or incorrectly generating complex scientific images. Moreover, to prove primary outcomes AI can plagiarize results from previous studies as support for a hypothesis (Elali & Rachid, 2023). Additionally, specific AI models such as neural networks are ambiguous regarding the decision-making process used to garner output results. This lack of transparency limits accountability for misconduct in the scientific community.

Strategy 1: AI Bias Through Biased Input, Omission of Outliers, and Skewed Analysis

Primary forms of misconduct detected are data fabrication and falsification. Current practices for the detection of AI misconduct surrounding data are statistical analyses used for data verification (Eckhartt & Ruxton, 2023). The scientific community identifies anomalies in presented data through Benford's law, which describes specific patterns of different order digits in numerical datasets, which can be grounds for further investigation (Eckhartt & Ruxton, 2023). Another method applied for detection is cross-verification through comparative analyses between the raw data inputted and the results. With the presence of many different detection methods, through software or statistical analysis, data fabrication has become a relatively easier misconduct to catch. This strategy surrounds unorthodox data falsification without fabrication. As mentioned previously, machine learning algorithms are developed depending on the task, and algorithm design can be skewed to emphasize certain traits over others resulting in biased

output. For example, a predictive AI model may be developed to weigh certain socioeconomic statuses or demographics over others resulting in affirmation of positive results for a correlation that may not be generalizable to a broader population. AI models like deep learning, require training based on data to extrapolate results. Therefore, if the training data propagates a bias, then the algorithm itself will perpetuate this bias in outputted results. Through unblinded algorithmic design and development, training of the algorithm using biased data, and inputting biased data in the trained model with the omission of outliers, one can safely perpetuate a bias via data falsification.

This is a successful method as due to the inputted dataset being obtained values, it can escape Benford's Law. This method also surpasses the initial cross-verification method as the inputted raw data into the AI model will translate to the results received. Moreover, even if the raw data were to be inputted into a similar algorithm that is trained without bias because the inputted data omits outliers, there will be a propagation of bias in the results. However, cross-verification with a different algorithm will not be sufficient evidence to declare misconduct as it may make assumptions of similar nature which leads to checks that are not independent of each other.

Strategy 2: Advanced AI Language Models for Plagiarized Text

Another potential strategy uses AI to write and subsequently publish plagiarized and/or nonsensical biomedical research. Through large language models (LLMs) and natural language processing models (NLPs), text can be generated without accurate citation by summarizing or paraphrasing data it has been trained on (Clusmann et al., 2023). An AI model specifically tailored to a lab's area of research that continuously accounts for feedback and advancements in the field can efficiently and effectively fabricate scientific articles (Clusmann et al., 2023). Fabrication can be done by training the algorithm on available datasets and documents to learn techniques associated with natural language that can be mistaken as human (Gruetzemacher, 2022; Kedia et al., 2024). This strategy is more successful than current attempts because it specifically addresses and bypasses detection strategies proposed in current literature. For example, the detection techniques developed by Cabanac and Labbé (2021) to identify papers from grammar-based generators such as SCIgen included analyzing the text for word choice and grammatical structure found in generated sentences (also known as "fingerprints"). By retrieving information from established sources written by humans in a certain research discipline, subtleties about the field such as typical sample size or recent developments will be better informed and expressed (Gao et al., 2023; Kedia et al., 2024). Therefore, understood "telltale" signs such as erroneous citations and nonsensical claims will be less frequent and apparent, posing difficulties in identifying this misconduct (Bhargava et al., 2023; Else, 2021). This will ensure unethical researchers are still provided with the career benefits of a successful publication (resume or graduate school application content, a higher h-index, job, or reviewer opportunities within their field, etc.) and enable the 'publish or perish' mindset (Cabanac & Labbé, 2021).

Strategy 3: AI for Image Fabrication

Images for scientific publications can be falsified to meet the intended results of the research using an advanced generative AI model. Generative Adversarial Networks (GANs) generate and modify images to portray desired research outcomes (L. Wang et al., 2022). GANs consist of two deep neural networks, a generator that creates the images and a discriminator that evaluates these images based on real images to improve the outputs of the generator (Kim et al., 2024). The GAN in this strategy uses a dual discriminator model, a system with two discriminators, to stimulate the behaviour of AI-detection systems (F. Wang et al., 2023). This stimulation will provide feedback on image improvement to avoid being detected as fake. By modifying inputs into the trained model, the GAN can generate a fake image that

maintains the favourable aspects of the original image(s) but alters the unwanted aspects. This generated image can bypass traditional detection methods as the changes are subtle and still contain natural patterns and textures (L. Wang et al., 2022). Figure 1 shows images created by Gu et al. (2022) using GANs to highlight the potential risk these generative models have on the future of image fraud in science publications.

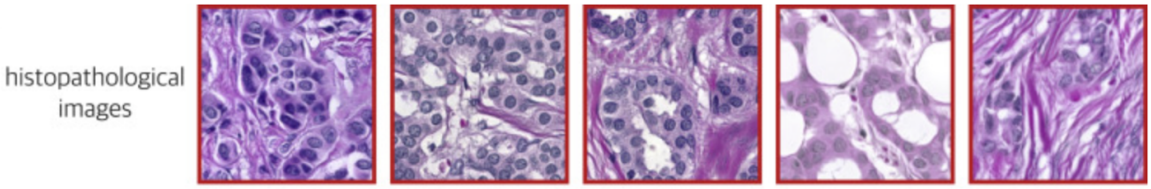

Figure 1 Fake histopathological images generated by GANs (Gu et al., 2022)

Despite the alarming presence of image misconduct in scientific publications, there have been few cases explicitly attributed to AI. However, it is predicted that many cases of fraudulent images have used AI without authors' disclosure or that AI-generated images have gone undetected (Bik et al., 2016). An unsuccessful case of AI-generated figures was identified where Western blots from different manuscripts and researchers all displayed identical backgrounds, regular spacing, and consistent band shapes (Christopher, 2018). This raised concerns about the authenticity of the research and data, ultimately leading to rejection by the editor. As reviewers are starting to identify common manipulation techniques and more publishers adopt detection software, a more robust strategy is needed to avoid detection. Using specifically trained GANs to falsify images will be the key to successfully committing image misconduct. These images are designed to match subtle and complex details that exist in real images, making it difficult to identify inconsistencies with manual examination techniques (L. Wang et al., 2022). Automated detection screening tools are another way to catch fraudulent images, however GANs can bypass these as well. These screening tools are trained to recognize image duplication and splicing modifications however, since the GAN images are entirely new this tool is ineffective in stopping the proposed strategy (Hosseini & Resnik, 2024).

Part Two

Countermeasure to Strategy 1

To counteract skewed data input and analysis there are a few guidelines the scientific community can establish. First and foremost is disclosure of the use of AI in research. If a lab is conducting its research with the use of AI, this is information that needs to be transparently stated, along with a reason explaining the use. Failure to disclose the use of AI needs to be strictly reprimanded with sanctions placed against the lab. Such transparency and repercussions allow the scientific community to assess the research with the correct perspective of scrutiny to ensure integrity. Additionally, an electronic workbook programmed with automatically tracked changes and saved versions of the workbook is critical. This workbook serves as a store for raw data that can be used for cross-verification with data inputted into the AI model. Moreover, if a lab is developing an algorithm to conduct specific tasks, a proper project management plan of algorithm design, methods, and training data needs to be embedded into the electronic workbook. This transparency surrounding algorithmic design allows the scientific community to robustly review the algorithm and assess it for the potential perpetuation of bias. With self-disclosure of AI use in research, there should be regulatory oversight by ethical committees to ensure the integrity of data and results obtained alongside any additional tasks the AI model was designed to conduct. With the increasing prevalence of AI in biomedical research, ethical committees should develop a division to review self-disclosure of AI use and the proposed project management plan as a preventative measure against ethical

misconduct. The countermeasure to the aforementioned strategy entails transparency, accountability, data management and regulatory oversight throughout the study.

Countermeasure to Strategy 2

To prevent plagiarism using specialized large language models, a countermeasure can be implemented at the level of the publication submission and peer review process. A concern with the use of AI in scientific writing is that it is nearly impossible to identify by peer reviewers alone (Conroy, 2023). Plagiarism checkers are also not entirely accurate at detecting misuse and cannot be solely relied on (Else, 2023). A potential solution is to mandate the submission of an integrity statement or brief that details the raw data and detailed methodology involved in writing a study. This will include the search strategy used for the article content, references utilized, and the relevant information taken from each. This will prevent authors from relying on the lack of traceability that is associated with AI generated writing and provide a more objective framework for peer reviewers to reference when assessing papers for publication. Implementation can begin at the level of regulatory bodies for research (e.g., COPE) and larger publishing companies, with the gradual plan of becoming standardized across research disciplines. While resources such as the time and effort of peer reviewers are still required in this process, it is more streamlined by providing scientists with more explicit accountability for their research practices. By preventing this misconduct in scientific writing, research findings are more likely to be a result of true scientific developments.

Countermeasure to Strategy 3

Identifying AI-generated images through detection software will be an ongoing battle of outcompeting, as GANs can often bypass these tools. Additionally, detection software like Proofig has the potential for false positives, creating further problems (Hosseini & Resnik, 2024). A more effective countermeasure for maintaining the integrity of scientific images is to require authors to provide detailed raw image data for all figures included in the papers. Authors will also be required to disclose any modifications made to the raw images for publication. This will be implemented seamlessly into the submission process, ensuring all images are provided upfront for editors and reviewers, to avoid troubling the author later in the process. This countermeasure specifically targets a weakness of GANs, which is their inability to generate large, high-resolution images that have the same authenticity and features of raw data (L. Wang et al., 2022). By investigating these raw images and the publication, it will be easier for reviewers to detect any manipulations or inconsistencies. Furthermore, knowing that reviewers and publishers will have access to, and examine the raw images, will likely deter authors from manipulating their images, as any discrepancies could be identified. Integrating this requirement into the submission process helps tackle the problem of AI-generated images in publications, promoting integrity and credibility in scientific literature.

Conclusion

The combination of these countermeasures will ensure that misconduct in biomedical research using AI is properly addressed at various levels of the research process. This will not only increase the trust the research community and public have in science, but it will also ensure that health policies and clinical care decisions made from research findings are credible and meaningful for the respective patient populations. Providing guidelines alongside encouraging the ethical and transparent use of AI will support its development as an effective tool in research and reduce the harm it can pose to society.

References

Bhargava, D. C., Jadav, D., Meshram, V. P., & Kanchan, T. (2023). ChatGPT in medical research: challenging time ahead. *Medico-Legal Journal*, *91*(4), 223–225. https://doi.org/10.1177/00258172231184548

Bik, E. M., Casadevall, A., & Fang, F. C. (2016). The prevalence of inappropriate image duplication in biomedical research publications. *MBio*, *7*(3). https://doi.org/10.1128/MBIO.00809-16

Cabanac, G., & Labbé, C. (2021). Prevalence of nonsensical algorithmically generated papers in the scientific literature. *Journal of the Association for Information Science and Technology*, *72*(12), 1461–1476. https://doi.org/10.1002/ASI.24495

Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., & Xie, Z. (2018). Deep Learning and Its Applications in Biomedicine. *Genomics, Proteomics & Bioinformatics*, *16*(1), 17–32. https://doi.org/10.1016/J.GPB.2017.07.003

Christopher, J. (2018). Systematic fabrication of scientific images revealed. *FEBS Letters*, *592*(18), 3027–3029. https://doi.org/10.1002/1873-3468.13201

Clusmann, J., Kolbinger, F. R., Muti, H. S., Carrero, Z. I., Eckardt, J.-N., Laleh, N. G., Löffler, C. M. L., Schwarzkopf, S.-C., Unger, M., Veldhuizen, G. P., Wagner, S. J., & Kather, J. N. (2023). The future landscape of large language models in medicine. *Communications Medicine*, *3*(1), 1–8. https://doi.org/10.1038/s43856-023-00370-1

Conroy, G. (2023). Scientific sleuths spot dishonest ChatGPT use in papers. *Nature*. https://doi.org/10.1038/D41586-023-02477-W

Eckhartt, G. M., & Ruxton, G. D. (2023). Investigating and preventing scientific misconduct using Benford's Law. *Research Integrity and Peer Review*, *8*(1). https://doi.org/10.1186/S41073-022-00126-W

Elali, F. R., & Rachid, L. N. (2023). AI-generated research paper fabrication and plagiarism in the scientific community. *Patterns*, *4*(3), 100706. https://doi.org/10.1016/J.PATTER.2023.100706

Else, H. (2021). "Tortured phrases" give away fabricated research papers. *Nature*, *596*(7872), 328–329. https://doi.org/10.1038/D41586-021-02134-0

Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, *613*(7944), 423. https://doi.org/10.1038/D41586-023-00056-7

Fonseca, M. (2024, January 22). *A handy guide to Bayesian Neural Networks for biomedical researchers*. Editage Insights. https://www.editage.com/insights/a-handy-guide-to-bayesian-neural-networks-for-biomedical-researchers

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *Npj Digital Medicine 2023 6:1*, *6*(1), 1–5. https://doi.org/10.1038/s41746-023-00819-6

Gruetzemacher, R. (2022, April 19). *The Power of Natural Language Processing*. Harvard Business Review. https://hbr.org/2022/04/the-power-of-natural-language-processing

Gu, J., Wang, X., Li, C., Zhao, J., Fu, W., Liang, G., & Qiu, J. (2022). AI-enabled image fraud in scientific publications. *Patterns*, *3*(7). https://doi.org/10.1016/j.patter.2022.100511

Hosseini, M., & Resnik, D. B. (2024). Guidance needed for using artificial intelligence to screen journal submissions for misconduct. *Research Ethics*. https://doi.org/10.1177/17470161241254052

Kedia, N., Sanjeev, S., Ong, J., & Chhablani, J. (2024). ChatGPT and Beyond: An overview of the growing field of large language models and their use in ophthalmology. *Eye*, *38*(7), 1252–1261. https://doi.org/10.1038/s41433-023-02915-z

Kim, J. J. H., Um, R. S., James, ·, Lee, W. Y., & Ajilore, · Olusola. (2024). Generative AI can fabricate advanced scientific visualizations: ethical implications and strategic mitigation framework. *AI and Ethics*, 1–13. https://doi.org/10.1007/S43681-024-00439-0

Phogat, R., Manjunath, B. C., Sabbarwal, B., Bhatnagar, A., Reena, & Anand, D. (2023). Misconduct in Biomedical Research: A Meta-Analysis and Systematic Review. *Journal of International Society of Preventive & Community Dentistry*, *13*(3), 185. https://doi.org/10.4103/JISPCD.JISPCD_220_22

Reitman, E. (2022, May 10). *The Role of Machine Learning Algorithms in Biomedical Discovery*. Yale School of Medicine. https://medicine.yale.edu/news-article/david-van-dijk-the-role-of-machine-learning-in-biomedical-discovery/

Wang, F., Ma, Z., Zhang, X., Li, Q., & Wang, C. (2023). DDSG-GAN: Generative Adversarial Network with Dual Discriminators and Single Generator for Black-Box Attacks. *Mathematics*, *11*(4), 1016. https://doi.org/10.3390/MATH11041016

Wang, L., Zhou, L., Yang, W., & Yu, R. (2022). Deepfakes: A new threat to image fabrication in scientific publications? *Patterns*, *3*(5). https://doi.org/10.1016/j.patter.2022.100509