

Enhancing Transparency and Research Ethics through Human-AI Techniques

Tatiana Chakravorti

Pennsylvania State University, University Park, Pennsylvania, USA
tfc5416@psu.edu

Abstract

In recent years, the empirical sciences have confronted a crisis of confidence following high-profile replication projects reporting disappointing results. The 'replication crisis' has led to widespread introspection and calls for reform across various fields. At the heart of this movement is the fundamental belief that reproducing, replicating, and interrogating published claims is essential to science. To facilitate this, many in the scientific community have emphasized the importance of openness and transparency in research, well-aligned incentives for researchers, robust peer review, and improved methods for meta-analyses over scientific corpora. Our work identifies with this movement, and in particular, highlights opportunities for emerging technologies to contribute to novel solutions in this space.

This research proposes a coherent research agenda exploring artificial intelligence (AI) and hybrid human-AI technologies to support the evaluation of the replicability of scientific findings and understanding of science as a set of social and cultural processes. We discuss work developing and testing a new class of AI algorithms, specifically, artificial prediction markets populated by algorithmic traders who buy and sell contracts representing the outcomes of replication studies of published research outcomes. Given the critical role of human intuition and experience in making it more trustworthy, we explore ways in which artificial prediction markets can be modified to include human participants, so-called hybrid prediction markets. We study these hybrid systems both in simulation and in pilot testing with human participants in a real-world scenario. In a series of studies, we engage in extensive surveys and interviews with researchers to better understand the usefulness and design requirements for both AI and hybrid human-AI technologies in the scientific workflow. We solicit feedback on artificial and hybrid markets for estimating the replicability of published claims and focus on the role of transparency and explainability in evaluating scientific integrity. We aim to situate these concerns within social and cultural contexts, making explicit the norms and incentives driving current research ecosystems.

Motivation

Reproducibility, replicability, and transparency are foundational principles of scientific inquiry, essential for ensuring that research findings are reliable and credible. Open science

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

practices, such as the sharing of data and methodologies, play a crucial role in supporting these principles by facilitating the replication and validation of findings by other researchers. Over the last decade, the open science and science of science communities have launched innovative initiatives aimed at addressing concerns related to the reproducibility, replicability, and transparency of published scientific research. These efforts have emerged in response to a growing awareness of a crisis in the scientific credibility of established findings. While there have been successes in these areas, gaps remain, as many researchers still lack awareness of these issues, the responses to them, or the best practices for conducting reproducible, replicable, and reliable research. In this era of artificial intelligence (AI), researchers need to think about how to use AI techniques to overcome this crisis at the same time consider more explainable AI techniques. The primary goal of this study is to foster a more open and accessible scientific process, wherein results are not just available but also understandable and applicable across various disciplines worldwide. To tackle the reproducibility and replication crisis, this study employed human-AI approaches. Despite notable achievements in the field of Artificial Intelligence, machine-driven technologies face significant challenges, especially in complex tasks that require creativity, planning, common sense, intuition, or learning from sparse data. These challenges underscore the need for effective human-machine collaboration strategies. The aim is to leverage human-AI approaches to predict the replication outcomes of scientific research, thereby enhancing trust and robustness in scientific findings. The goal of my research work is to address four major issues in scientific integrity:

1. Enhancing Reproducibility and Replicability through a hybrid prediction market.
2. Address and understand the perceptions of researchers regarding research integrity and the potential of the AI replicability prediction tool for scientific work.
3. Address and understand the State of Reproducibility, Replicability, and Open Science in two different cultural countries, one is from the global South and one is from the West
4. Address and understand the ethical use of large language models (like ChatGPT) and the role of Open Science in supporting scientific progress and enhancing peer review

processes from the viewpoint of reviewers, journal editors, and conference chairs who are deeply involved in the publication process.

Human-AI techniques for Research Ethics

The main goal of this work is to contribute to the ongoing efforts to solve the scientific crisis by developing and validating new AI methods for assessing scientific findings, fostering open science practices, and enhancing the collaboration between humans and AI in the scientific research process.

As we all know, AI and machine learning (ML) have seen significant advancements, achieving human-level or even superior performance in various complex tasks. Despite these achievements, the practical deployment of AI systems faces several critical challenges. AI models often lack common sense and struggle with tasks that require creativity, intuition, or the ability to learn from limited data. Furthermore, while established classification algorithms have shown success, they have not fully earned user trust for critical decision-making. These successes and limitations have spurred researchers to explore hybrid human-AI approaches, aiming to combine the strengths of both human and artificial intelligence. In this study, I have thoroughly experimented with an artificial prediction market model to understand the effects of market parameters on model performance for benchmark classification tasks. Then demonstrated, through simulation, the impact of exogenous agents in the market, where these exogenous agents represent primitive human behaviors (Chakravorti et al. 2022).

A prototype human-AI prediction market is built, showcasing its potential for facilitating meaningful human-AI collaboration (Chakravorti et al. 2023). Within these markets, AI agents (bot traders) are trained to buy outcomes of future events alongside humans. Such classification decisions can be interpreted as outcomes of these events, with the price of an asset linked to a specific classification outcome serving as an indicator of the system's confidence in that decision. By integrating human participants into these markets alongside bot traders, I amalgamate insights from both sources. This research work was conducted with prototype hybrid markets aimed at predicting the outcomes of replication studies. I explored both the challenges and opportunities presented by this approach, shared insights from semi-structured interviews with participants of the hybrid market, and provided a vision for future research and development in this area.

In this work, I focused on the formative evaluation of the replication prediction tool which is the product of my artificial prediction market (Chakravorti et al. 2024b; Wu et al. 2024). Formative evaluation is carried out during the development or improvement phase of the subject under evaluation, with the primary goal of gathering feedback for real-time enhancements. We conducted semi-structured interviews with faculty members and PhD scholars from universities across India to gauge their current research practices and their understanding of the reproducibility crisis. There are two groups of researchers, one is social science and another one is engineering to explore both perspectives. The prototype AI replication prediction tool will be introduced

as a basis for discussions on how computational tools can be designed and implemented to support their work with research ethics. By centering these perspectives, we explore the acceptance and perceptions of hybrid human-AI solutions and also how they can integrate this tool into their research.

Finally, I have made lots of efforts to understand experiences and perspectives surrounding reproducibility, replicability, and open science practices have predominantly focused on researchers and research communities in Western countries (Chakravorti, Koneru, and Rajtmajer 2024; Chakravorti et al. 2024a). This research work represents an initial step towards incorporating cultural perspectives by conducting a comparative study of researchers in the USA and India. The goal is to conduct a comparative analysis of the state of reproducibility and open science in India and the USA, identifying the challenges researchers face in these countries, the incentives necessary for fostering transparent research practices, the features researchers look for in a paper to assess its credibility, and their perceptions and experiences with the current peer review process. Finally, I have made lots of efforts to understand experiences and perspectives surrounding reproducibility, replicability, and open science, and have predominantly focused on researchers and research communities in Western countries. This research work represents an initial step towards incorporating cultural perspectives by conducting a comparative study of researchers in the USA and India. The goal is to conduct a comparative analysis of the state of reproducibility and open science in India and the USA, identifying the challenges researchers face in these countries, the incentives necessary for fostering transparent research practices, the features researchers look for in a paper to assess its credibility, and their perceptions and experiences with the current peer review process.

Conclusion

This research provides significant contributions to the fields of science of science, human-AI collaboration, and ethical research practices. It offers innovative solutions to longstanding challenges in reproducibility and peer review, advocates for the integration of cultural perspectives, and explores the frontier of human-AI collaboration in advancing scientific integrity and transparency.

References

- Chakravorti, T.; Fraleigh, R.; Fritton, T.; McLaughlin, M.; Singh, V.; Griffin, C.; Kwasnica, A.; Pennock, D.; Giles, C. L.; and Rajtmajer, S. 2023. A prototype hybrid prediction market for estimating replicability of published work. In *HAI 2023: Augmenting Human Intellect*, 300–309. IOS Press.
- Chakravorti, T.; Gautam, S.; Silverstein, P.; and Rajtmajer, S. M. 2024a. Open Science Practices by Early Career HCI Researchers: Perceptions, Challenges, and Benefits. arXiv:2410.04286.
- Chakravorti, T.; Koneru, S. D.; and Rajtmajer, S. 2024. Reproducibility, Replicability, and Transparency in Research:

What 430 Professors Think in Universities across the USA and India. arXiv:2402.08796.

Chakravorti, T.; Singh, V.; Rajtmajer, S.; McLaughlin, M.; Fraleigh, R.; Griffin, C.; Kwasnica, A.; Pennock, D.; and Giles, C. L. 2022. Artificial prediction markets present a novel opportunity for human-AI collaboration. *arXiv preprint arXiv:2211.16590*.

Chakravorti, T.; Wu, C.; Koneru, S.; and Rajtmajer, S. 2024b. Perspectives from India: Opportunities and Challenges for AI Replication Prediction to Improve Confidence in Published Research. arXiv:2310.19158.

Wu, C.; Chakravorti, T.; Carroll, J. M.; and Rajtmajer, S. 2024. Integrating measures of replicability into scholarly search: Challenges and opportunities. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18.