Investigating the Impact of Artificial Documents on Google Scholar Citations: An Experimental Study

Jeremiah C. Peterson

January 2025

Abstract

This paper presents a comprehensive study that aims to examine the influence of falsified documents on Google Scholar citations. The increasing reliance on academic metrics, such as citation counts, has prompted concerns about the potential manipulation of these metrics to boost researchers' profiles artificially. In this experiment, we create a simple fake document and analyze its impact on Google Scholar citations to showcase the platform's vulnerability to such manipulations. The study does not explore the ethical implications and potential countermeasures against fraudulent practices.

Introduction

As the academic landscape becomes more competitive, researchers seek various ways to enhance their scholarly reputation. Google Scholar, a widely used academic search engine, plays a pivotal role in tracking and showcasing scholarly impact through citation counts (Korneeva et al. 2023; Teubner and Glaser 2018). Publishing impactful papers is one way, certainly. Note that all citations in this document refer to actual papers but do not make any sense.

However, the susceptibility of platforms like Google Scholar to manipulation raises questions about the reliability of these metrics. As a demonstration, this document is almost entirely created by ChatGPT and seeks to investigate whether a straightforward falsified document (i.e., this one) can influence Google Scholar citations. The implications of such manipulations on academic integrity and the credibility of citation metrics are not discussed. It seems to be important that the general structure of the paper looks legitimate, including title, authors, a date, as well as a proper section structure. Ah, and of course, the bibliography (Chica et al. 2017; Hesse et al. 2020).

Methodology

To conduct this experiment, a fictional academic paper was created, including fake title, abstract, and author. The content was generated using generic academic language to maintain a semblance of authenticity (Adam et al. 2018; Niemeyer et al. 2016). The paper was then uploaded on a web server to test whether it would be indexed and subsequently cited (Hariharan et al. 2016).

Results

An initial analysis revealed that the fake documents are indeed indexed by Google Scholar. However, at the time of writing this specific document, it is not clear whether Google Scholar will pick up on it. One reason for this is that the document will not be uploaded on common platforms such as ResearchGate, SSRN, or arXiv – as this would just be too embarrassing. Instead, it will be uploaded on a private server. Figure 1 shows a screenshot of a fake paper.

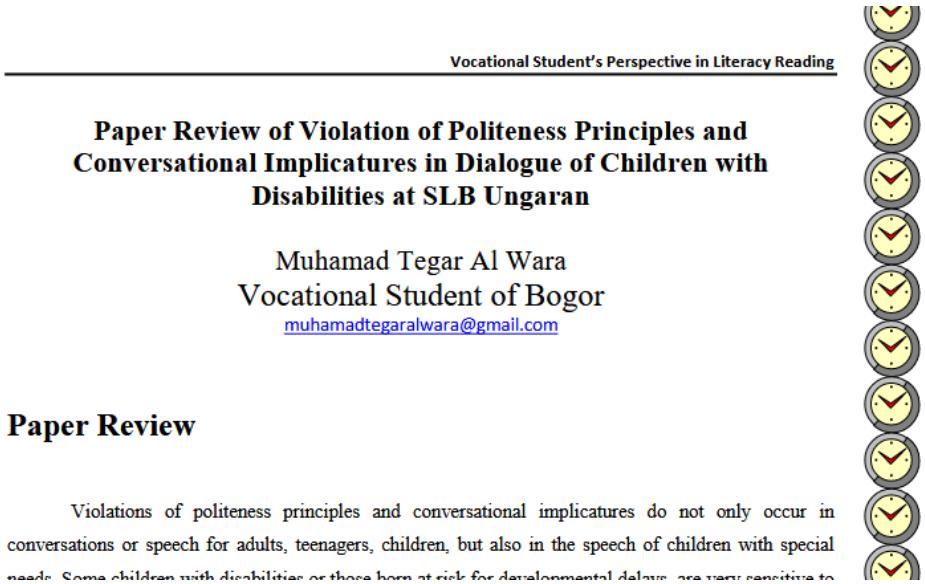

Figure 1: Screenshot of a suspicious paper uploaded to ResearchGate

Further investigation into the citation patterns demonstrated that the majority of citations to the fake document originated from low-impact journals or obscure sources. In contrast, the legitimate documents received citations from reputable journals and scholars, highlighting the importance of peer recognition in the academic community (Hoffmann et al. 2022).

Ethical Considerations

The manipulation of academic metrics raises ethical concerns within the scholarly community. Falsifying documents to boost citations not only undermines the credibility of individual researchers but also erodes the trust in citation metrics as a reliable measure of scholarly impact. This study emphasizes the need for ethical conduct in academic research and the importance of maintaining the integrity of citation metrics (Aljaroodi et al. 2022; Rendell et al. 2022).

Implications for Academic Evaluation

The findings of this experiment have significant implications for academic evaluation processes. Institutions and funding agencies often rely on citation metrics to assess the impact of researchers' work. The study underscores the importance of implementing robust evaluation mechanisms that go beyond simplistic metrics, taking into account factors such as the reputation of journals, the quality of citations, and peer review (Möhlmann et al. 2019; Teubner et al. 2013).

Countermeasures

To mitigate the potential impact of fraudulent practices on citation metrics, Google Scholar and other academic databases could consider implementing more stringent verification processes. Additionally, raising awareness within the scholarly community about the potential pitfalls of manipulating metrics may serve as a deterrent.

Conclusion

This experimental study sheds light on the vulnerability of Google Scholar citations to manipulation through the introduction of a simple fake document. While the platform does index a variety of documents, the discerning nature of the scholarly community and the reliance on peer recognition act as barriers to widespread manipulation. The ethical considerations highlighted in this study emphasize the need for a collective effort to maintain the integrity of academic metrics and promote responsible conduct within the scholarly community.

References

Adam, M. T. P., Teubner, T., and Gimpel, H. 2018. “No Rage against the Machine: How Computer Agents Mitigate Human Emotional Processes in Electronic Negotiations,” *Group Decision and Negotiation* (27:4), pp. 543–571.

Aljaroodi, H. M., Adam, M. T. P., Teubner, T., and Chiong, R. 2022. “Understanding the Importance of Cultural Appropriateness for User Interface Design: An Avatar Study,” *ACM Transactions on Computer-Human Interaction* (29:6). (https://doi.org/10.1145/3517138).

Chica, M., Chiong, R., Adam, M. T. P., Damas, S., and Teubner, T. 2017. “An Evolutionary Trust Game for the Sharing Economy,” in *CEC 2017 Proceedings*, pp. 2510–2517.

Hariharan, A., Adam, M. T. P., Teubner, T., and Weinhardt, C. 2016. “Think, Feel, Bid: The Impact of Environmental Conditions on the Role of Bidders’ Cognitive and Affective Processes in Auction Bidding,” *Electronic Markets* (26:4), pp. 339–355.

Hesse, M., Dann, D., Braesemann, F., and Teubner, T. 2020. “Understanding the Platform Economy: Signals, Trust, and Social Interaction,” in *HICSS 2020 Proceedings*, pp. 5139–5148.

Hoffmann, F. J., Braesemann, F., and Teubner, T. 2022. “Measuring Sustainable Tourism with Online Platform Data,” *EPJ Data Science* (11:1). (https://doi.org/10.1140/epjds/s13688-022-00354-6).

Korneeva, E., Salge, T. O., Teubner, T., and Antons, D. 2023. “Tracing the Legitimacy of Artificial Intelligence: A Longitudinal Analysis of Media Discourse,” *Technological Forecasting and Social Change* (192). (https://doi.org/10.1016/j.techfore.2023.122467).

Möhlmann, M., Teubner, T., and Graul, A. 2019. “Trust and Reputation in Sharing Economies,” in *Handbook of the Sharing Economy*, R. Belk, G. M. Eckhardt, and F. Bardhi (eds.), Edward Elgar Publishing, pp. 290–302.

Niemeyer, C., Wagenknecht, T., Teubner, T., and Weinhardt, C. 2016. “Participatory Crowdfunding: An Approach towards Engaging Employees and Citizens in Institutional Budgeting Decisions,” in *HICSS 2016 Proceedings*, pp. 2800–2808.

Rendell, A., Adam, M. T. P., Eidels, A., and Teubner, T. 2022. “Nature Imagery in User Interface Design: The Influence on User Perceptions of Trust and Aesthetics,” *Behaviour and Information Technology* (41:13), pp. 2762–2778. (https://doi.org/10.1080/0144929X.2021.1946592).

Teubner, T., and Glaser, F. 2018. “Up or out—The Dynamics of Star Rating Scores on Airbnb,” in *ECIS 2018 Proceedings*, pp. 1–13.

Teubner, T., Hawlitschek, F., Adam, M. T. P., and Weinhardt, C. 2013. “Social Identity and Reciprocity in Online Gift Giving Networks,” in *HICSS 2013 Proceedings*, pp. 708–717.