Mitigating Bias In Academic Publishing: Towards Responsible (Gen)Ai-Augmentation In Peer-Rewiew Processes

Christian Meske
Ruhr University Bochum, christian.meske@rub.de

Daniel Eisenhardt
Ruhr University Bochum, daniel.eisenhardt@rub.de

Dunja Šešelja
Ruhr University Bochum, dunja.seselja@rub.de

Christian Straßer
Ruhr University Bochum, christian.strasser@rub.de

Johannes Schneider
University of Liechtenstein, johannes.schneider@uni.li

MITIGATING BIAS IN ACADEMIC PUBLISHING: TOWARDS RESPONSIBLE AI-AUGMENTATION IN PEER-REWIEW PROCESSES

Meske, Christian, Ruhr University Bochum, Bochum, Germany, christian.meske@rub.de
Eisenhardt, Daniel, Ruhr University Bochum, Bochum, Germany, daniel.eisenhardt@rub.de
Šešelja, Dunja, Ruhr University Bochum, Bochum, Germany, dunja.seselja@rub.de
Straßer, Christian, Ruhr University Bochum, Bochum, Germany, christian.strasser@rub.de
Schneider, Johannes, University of Liechtenstein, Liechtenstein, johannes.schneider@uni.li

Research-in-Progress

Abstract

The anonymized peer review process, fundamental to maintaining the quality of scientific publications, has been both praised and criticized. Concerns over various biases, including those related to authors' affiliations, gender, and groundbreaking ideas, have led to calls for critical reflection. In addition, scientific output and hence the number of necessary reviews has increased tremendously. In that context, machine learning models, and large language models such as ChatGPT more specifically, have been explored as potential solutions to enhance the reviewing process. Yet, AI can be biased itself. Thus, a systematic approach to designing AI systems that mitigate bias in peer review is lacking. Our study hence aims to address this gap by formulating design principles for (gen)AI-augmented review systems. Utilizing an echeloned design science research (DSR) methodology, the project seeks to develop new design knowledge and create a prototype system incorporating these principles.

Keywords: Generative AI, Artificial Intelligence, Scientific Review, Bias

1 Introduction

To determine the quality of a scientific manuscript and whether it is good enough for publication, the process of anonymized peer review has become established across domains. While this review process has advantages such as anonymity, helping to maintain a standard of quality and providing feedback to authors, it is both poorly understood (Tennant and Ross-Hellauer, 2020) and has faced criticism. Weber (2024) gathers examples of criticism expressed over the last 20 years, containing concerns about reviewers' competence (Resnik et al., 2008), low interrater agreement (Jackson et al., 2011; Lee, 2012), the true confidentiality of a submission (Fire and Guestrin, 2019; Johnson et al., 2018), reviewers failing to provide timely constructive and high quality reviews (Huisman and Smits, 2017; Smith, 2006), and so forth. Two particularly pressing issues resulting from this discussion are a growing number of publications demanding timely reviews and biases in peer review, each of which poses challenges to the reviewing process.

On the one hand, the timely pressure to provide a high quality and contributing review becomes more and more challenging with an increasing number of submissions (Bornmann, et al., 2021; Sun et al. 2021). Publons (2018) reports a 6.1% annual increase in submissions since 2013. Van Noorden (2014) reports an annual increase of 8-9% in publications in 2014, and UNESCO (2021) claims that scientific output in 2019 was 21% higher than in 2015. With increasing submission numbers, it becomes more and more difficult for editors to manage their reviews and, for both editors and reviewers, to keep up with the submission rate (Ghosal et al., 2018). The quality of reviews could suffer due to the time pressure on editors and reviewers. This, in turn, could lead to reviewers providing feedback that is less understandable or appropriate for authors and, as a result, not improving the quality of the manuscript. Reviewers may fail to identify errors, possibly leading to the publication of mistaken results. This is particularly problematic given the high increase in retraction rates (van Noorden, 2023).

On the other hand, it has long been recognized that peer review can suffer from various forms of biases, undermining its function of quality control even in the absence of timely pressures. Biases may originate in a differential treatment of authors depending, for instance, on their affiliation, gender, nationality or language (Lee et al., 2013). For example, in a classic study by Peters and Ceci (1982), when previously published articles by authors from prestigious institutions were resubmitted under fictitious names and less prestigious affiliations, the majority ended up rejected. What is more, peer review may suffer from a bias against groundbreaking ideas, rooted in reviewers' conservatism (Braben, 2004; Katzav and Vaesen, 2017, Stanford, 2019, while findings in Teplitskiy et al, 2022, do not support this bias), the publication bias – a tendency to publish studies showing positive rather than negative findings (Fanelli, 2010; Marks-Anglin and Chen, 2020), a bias in reviewer selection that equips a small group of reviewers with gatekeeping powers overs a significant part of submissions (Qin et al, 2014), or biases originating in editors' undisclosed conflicts of interest (Dal-Ré et al. 2019; Teixeira da Silva and Dobránszki 2015). These concerns have led to calls for the traditional system of anonymized peer review to be altered (Matt et al, 2017), or even entirely abandoned (Heesen and Bright, 2021).

To address these problems, recent scholarship has turned to machine learning models as a potential solution. As artificial intelligence (AI) is rapidly improving, the academic community has identified opportunities for an AI-augmented reviewing process for different stakeholders such as authors (to receive feedback before submission), reviewers and editors (Bao et al., 2021; Checco et al., 2021; Drori and Te'eni, 2024; Liu and Sha, 2023). With the emergence of ChatGPT, the previous focus of classification models used for supporting review processes has shifted to testing the limits and experimenting with large language models (LLM). Additional technical improvements to enhance the performance of LLMs, such as retrieval augmented generation (RAG), are helping to drive new approaches. To support the review process, machine learning has been used for various subtasks such as format checking, plagiarism detection, language quality, scope and relevance, to name a few (Kankanhalli, 2024). While initial proposals for such models have been made (e.g. Ghosal et al., 2018; Li, 2022), we lack a systematic approach to their design, aimed at impartial reviewing processes, which minimise the intrusion of biases. Developing design principles is essential in order to evaluate whether and how peer review can be augmented with AI technologies in a socially and epistemically responsible way. We aim to address this research gap by addressing the following research question:

What are design principles to mitigate the risk of bias in (gen)AI-based systems used to augment scientific review processes?

This short paper is organised as follows. We first present the research background, focusing on AI in review processes and bias in AI (Section 2). We then elaborate on the applied methodology, in our case the echeloned Design Science Research (DSR) approach (Section 3). After showing preliminary results (Section 4), the paper ends with a discussion and conclusion (Section 5).

2 Research Background

2.1 AI in review processes

Recent advancements in AI have highlighted its potential to support various aspects of scientific research. Razack et al. (2021) identify a wide variety of AI-based systems, which they group into the following six categories according to their areas of application within the research process: Literature search and systematic review, writing and editing, references/citation, review and workflow, plagiarism detection, and journal selection. With regard to the review process, there is a trend towards the development of AI-based systems that are specifically designed to assist subtasks in the preparation of a review. Examples of AI-based systems for such subtasks are spelling and grammar checks, which can be performed with AIRA (Frontiers, 2020), assisting with editorial processes (Mrowinski et al, 2017), a format check that inspects whether the manuscript meets the journal's requirements, based on Penelope.ai (Kankanhalli, 2024), and the review of statistical tests, such as StatCheck (Nuijten and Polanin, 2020) and StatReview (Shanahan, 2016).

Moreover, Kankanhalli (2024) identifies additional subtasks, including plagiarism detection, manuscript-reviewer matching, scope/relevance, novelty, significance, writing and integrating reviews, and reproducibility check, many of which are based on LLMs. Looking more closely at the subtasks in terms of their previously mentioned solution approaches, a connection between the different complexity of the tasks and the corresponding solutions becomes apparent. Kankanhalli (2014) also divides the tasks into pre-peer review screening and peer review tasks. Pre-peer review tasks have characteristics of tasks that are easier to structure, such as format checks, plagiarism detection and language quality checks (Kankanhalli, 2024). These tasks can usually be solved by recognising and comparing patterns. Looking at the tasks at the peer review stage, it is clear that the focus is more on content-related factors, i.e. drawing conclusions, making connections, assessing originality and impact. Crossley et al. (2023) write that the peer review tasks mentioned are more concerned with cohesion and logic. While Susarla et al. (2023) write that analyses of scope and relevance still require human intervention and control and thus offer only limited potential for automation, Drori and Te'eni (2024) already report promising results in the creation of reviews and their content with ChatGPT and, to a large extent, agreement with humans in the acceptance and rejection decision. Nevertheless, LLMs should be used with caution due to possible biases and potential hallucinations of facts (Agrawal et al, 2023; Ji et al, 2023; Shmueli and Soumya, 2024; Susarla et al, 2023). As the development of LLMs such as ChatGPT or Llama has only recently reached a level of maturity, their use in the peer review process needs to be further explored, and ways to mitigate any potential risks are needed.

2.2 Bias in artificial intelligence

In the context of machine learning, biases can occur at different stages. While most of the literature on biases in ML focuses on data generation and preprocessing, Suresh and Guttag (2021) identify seven categories of biases at different stages of the ML pipeline. These types of biases can be divided into biases from the data generation process, which are historical bias, representation bias, and measurement bias, and biases that occur during model building and implementation, which are aggregation bias, learning bias, evaluation bias, and deployment bias (Suresh and Guttag, 2021). The effects of bias depend heavily on the context of the application, the decision-making power of the system and its degree of automation, as well as the available data. For example, an application can refuse to grant credit to people on the basis of discrimination (Raghavan et al., 2020; De-Artega et al., 2019), but can also overlook potentially carcinogenic skin lesions (Narla et al., 2018).

Due to the high susceptibility to biases within the ML pipeline and the possible severe consequences, several approaches have already been developed to minimise biases within the ML system. Bias minimisation approaches include specific feature selection (Grgić-Hlača et al., 2018), data preprocessing (Feldman et al., 2015; Calmon et al., 2017; Kamiran and Calders, 2012), model tuning (Kamishima et al., 2012, Zhang et al., 2018), and postprocessing (Kim et al., 2019; Hardt et al., 2016; Kamiran et al., 2018). In Natural Language Processing (NLP) systems, biases are often caused by word associations that are naturally anchored by humans in the training data (Caliskan et al., 2017; Garg et al., 2018). For example, gender bias can be caused by word embeddings that represent stereotypically charged word associations, such as the association of women and men with social and technical professions (Caliskan et al., 2017). Solutions have also been developed to minimise biases in NLP systems. Bolukbasi et al. (2016) developed a post-processing approach and Zhao et al. (2018) an approach that influences the training process. As noted earlier, NLP systems like ChatGPT are still maturing, so more research is needed to understand and prevent bias in various contexts and interpretations.

3 Methodology

For our research project, we apply the echeloned DSR (eDSR) method (Tuunanen et al., 2024) to establish design principles for (gen)AI systems used in peer review processes to mitigate affiliation bias, gender bias, language bias, and bias against novelty. eDSR is based on the proposed DSR methodology by Peffers et al. (2007), suggesting six individual phases within a DSR project: 1) identifying the problem and motivation; 2) defining the objectives; 3) designing; 4) demonstrating; 5) evaluating; and 6) communication. Tuunanen et al. (2024) extend the approach of Peffers et al. (2007) to include the possibility of focusing on several sub-systems (echelons), which are interdependent on each other in individual phases. With their reciprocal characteristics, echelons cumulate independent endeavors, contributing to the superordinate system (Tuunanen et al., 2024). Using echelons allows researchers to derive design knowledge about a broader set of goals, properties and features while focusing more specifically on their contributions to the primary system. Accordingly, a design process can gain breadth and depth in designing the artefact and deriving design knowledge. The use of echelons in the DSR project is intended to help researchers cope with the high complexity of socio-technical system design (Tuunanen et al., 2024).

Developing a generative AI-based system to complement the scientific peer review process and support multiple stakeholders is highly complex due to the multifaceted nature of the project. Accordingly, the eDSR approach allows us to address all the requirements and objectives of this project rather than focusing on a single perspective that would risk ignoring important relationships between levels.

In the following paragraph, we describe our activities for each of the design phases, as summarized in Figure 1 below. Since this is a research-in-progress paper, we only describe our specific activities up to the derivation of the first set of design requirements (see section 4). For the remaining phases, we will present our planned activities to describe how we intend to fulfil our objectives.

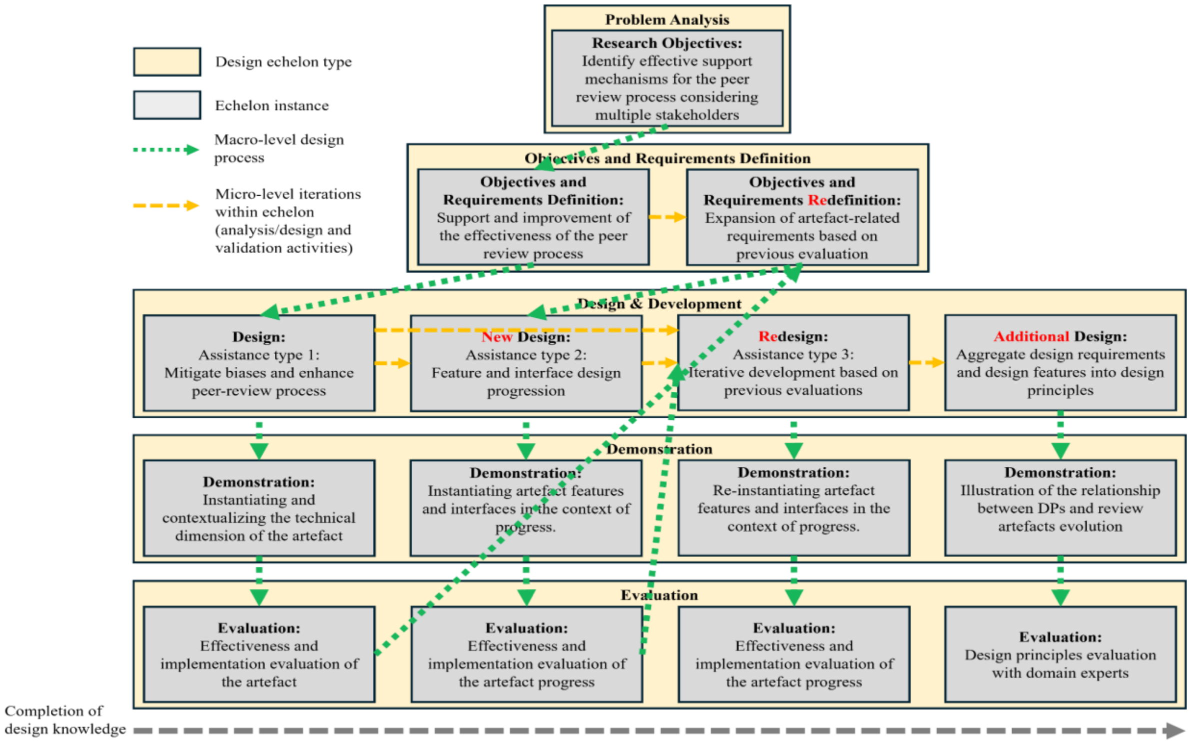

Figure 1. eDSR Methodology approach.

In the first step, problem analysis, a literature review was conducted to define the problem space and the current state of the art for our project. The focus was on previous technical implementations and the identification of problems from practice (relevance cycle) as well as on theories (rigour cycle) that can provide the basis for extracting design knowledge in the project context. The project is currently still in the first echelon instance (objective requirements definition) within the second design echelon type. Meta-requirements were derived from the identified challenges in literature to support and inform the further design process and can therefore be seen as input knowledge (Meth et al., 2015). To add to the derived meta-requirements we plan to conduct expert interviews with stakeholders from the peer-review process. From these interviews, we want to extract more user-specific and practical requirements, which will be developed into design requirements. Based on the total set of requirements, we then derive design principles. These principles will guide the design process of the actual artefact. Once the first prototype is developed, it will be demonstrated to the aforementioned stakeholder groups and evaluated to measure the effectiveness of the instantiation.

After the first design cycle, we are planning a second cycle by focussing on aspects and potentials that were identified during the first evaluation. We then conduct a second cycle, focusing on issues and opportunities identified during the first evaluation. Experience has shown that new objectives and requirements can be defined here, which form a further echelon instance. The extensions of the second cycle can be interface design and feature-based. After defining the newly derived objectives, we turn our attention, as in the previous cycle, to the design, demonstration and evaluation phases. As a result, a further cycle is possible, which starts with a redesign phase. The plan is to use the first evaluation to identify major deficits in the system and to rectify these in the second cycle. The design will then be optimised in smaller steps in a funnel-like manner. For this reason, the existing artefact will only be changed superficially in the redesigning phase, and no further target-changing features will be added. After the artefact has been redesigned and instantiated one last time, it will be evaluated. The results of the last evaluation are used to draw lessons for the requirements and design principles that were aggregated from the various phases. The design principles are then finalised and evaluated with domain experts.

4 Preliminary Results

There is already a knowledge base about the peer review process and the impact that AI could have on it. However, to our knowledge, there is still no explicit design knowledge for AI-augmentation in that regard. Therefore, in this section, we have summarised preliminary identified challenges and the resulting design requirements identified in the literature. We expect further challenges and design requirements to be identified. Figure 2 details all of our challenges and how they are mapped to the corresponding design requirements. The following passages give further insight into the origin of the design knowledge elicited.

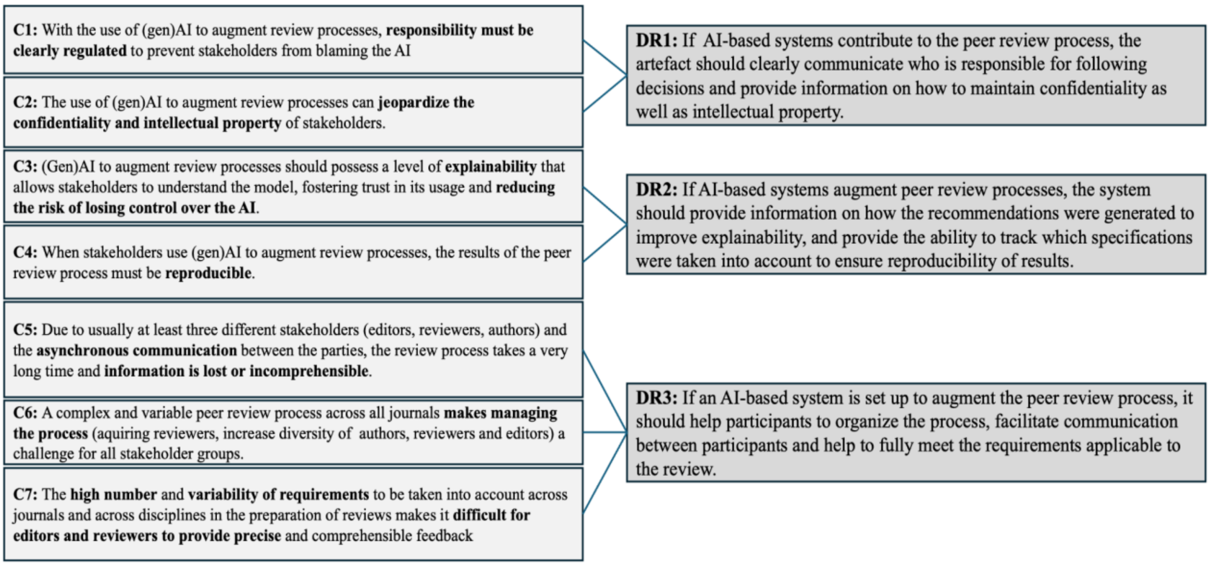

Figure 2. Initially identified challenges and requirements for (gen)AI in peer-review processes.

A key issue in the use of AI-based systems is that of control and accountability (Shneiderman, 2020). This issue also arises in the context of the peer review process, and the question of the allocation of responsibilities and accountability (C1) (Drori and Te'eni, 2024; Inam et al., 2024; Weber, 2024). The possibility of interacting with generative AI also raises questions of integrity and data protection. The input of data into generative AI raises fears of loss of intellectual property (C2) in a scientific context (Kankanhalli, 2024; Shmueli and Soumya, 2024). The two problems mentioned above (C1, C2) can be summarised as a first requirement (DR1) for the system. The focus here is on maintaining control over the generated output (submission, review) and clearly assigning responsibility for it. Another challenge in working with AI is the traceability of the results generated by the AI (Hutson, 2018; Korteling et al., 2021, p. 7; Meske et al., 2022; Wang et al., 2022).

If the underlying reasons for an AI decision are unclear to the stakeholders, trust in the system can be damaged (Hattke et al, 2018). Moreover, the added value of a detailed evaluation with AI is missing because the decision cannot be understood, risking a loss of control (C3) (Thelwall et al., 2020; Duine, 2023; Drori and Te'eni, 2024; Garcia, 2024; Kankanhalli, 2024). In order to understand AI, you also need to be able to retrace and reproduce results. Additional difficulties arise here, especially when using generative AI, as it can often deliver different results with the same input (C4) (Hutson, 2018; Weber, 2024). Based on these problems, the requirements for traceability and comprehensible presentation of decisions and outputs can be defined (DR2).

Several issues have been identified in relation to the tasks and requirements to be considered as part of the peer review process. The number of stakeholders in the peer review process and the asynchronous interaction between the individual stakeholders often impair the comprehensibility and information content of the communication between them (C5) (Shmueli and Soumya, 2024). Moreover, complex and different review processes across disciplines and journals make the organisation of the review process more difficult for everyone involved in the process (C6) (Thelwall et al., 2020; Huh, 2023; Kankanhalli, 2024; Shmueli and Soumya, 2024). In addition to organisational problems, it is becoming increasingly difficult for reviewers and editors to provide accurate and understandable feedback due to the different requirements of journals (C7) (Kankanhalli, 2024; Shmueli and Soumya, 2024). Taken together, issues C5-C7 require support for communication and organisation within the peer review process, as well as support for meeting journal-specific requirements (DR3).

5 Discussion and Conclusion

Introducing AI into peer review practices offers a promising way of improving the reviewing process by addressing the timely pressure on reviewers in light of the increasing number of publications, controlling for biases introduced by human reviewers and lowering the risk of retractions of published articles. At the same time, without a thorough development of design principles to guide this process, we risk introducing new problems into the reviewing process, including inaccurate or opaque feedback and novel forms of biases (Sullivan 2022). This paper provides the initial step towards this goal. Going forward, it would be valuable to develop design principles for including (gen)AI tools that can augment the assessment of the human-based reviews and their quality. For instance, developing (gen)AI tools to inform editors whether reviews fail to offer constructive feedback, whether they may suffer from bias against novelty or whether they include patronizing or insulting remarks, can help the editors to make timely decisions, such as inviting additional reviewers. Moreover, developing tools that can flag reviews of potentially low quality can help to prevent a high number of retractions.

Another issue that deserves more research is the role of biases in peer review. While biases are generally considered as undesired within reviewing processes, the issue may be more complicated. For instance, it has been argued that opting for biased reviewers may be a strategy employed by editors to extract more information from the reviewing process (García et al, 2015). Similarly, the lack of interrater agreement may be a consequence of applying community standards in different but rational ways (Lee, 2012). Such challenges should be considered when devising design principles for (gen)AI-augmented reviewing.

Finally, the significance of this work goes beyond the development of responsible (gen)AI reviewing assistants. Another research avenue that can benefit from these insights is the study of peer-review as a complex socio-epistemic phenomenon. The practice of peer reviewing renders science as a self-regulating socio-epistemic system. The role of peer reviewing in this social and dynamic setting has been studied by means of agent-based simulation models (see Feliciani et al, 2019, for an overview). These models study the effects of evaluation and confirmation biases (Squazzoni and Gandelli, 2012a; Garcia et al., 2015, 2016a, 2016b), reciprocity between reviewers and authors and related incentive structures (Squazzoni and Gandelli, 2012b, 2013; Bianchi et al, 2018; Radzvilas et al, 2023; Righi and Takacs, 2017) or editorial strategies (Wang et al, 2016). Some models compare different variants of peer review (Grimaldo et al, 2018; Zhu et al, 2016; Radzvilas et al, 2023; Kovanis et al, 2017). By understanding risks and benefits of introducing AI into peer-review, we can examine its impact on various aspects of the process: from risks that would result from having AI-automated (in contrast to AI-augmented) reviewing process, to potential benefits of using AI tools as a mitigating factor against the self-serving attributional bias (García, 2016b).

References

Agrawal, A., Suzgun, M., Mackey, L., and Kalai, A. T. (2023). *Do language models know when they're hallucinating references?* Available at https:// arxiv.org/abs/2305.18248.

Bao, P., Hong, W., and Li, X. (2021, April). "Predicting paper acceptance via interpretable decision sets." *In Companion Proceedings of the Web Conference 2021* (pp. 461-467).

Bianchi, F., Grimaldo, F., Bravo, G. et al. (2018). “The peer review game: an agent-based model of scientists facing resource constraints and institutional pressures.” *Scientometrics* 116, 1401–1420.

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., and Kalai, A. T. (2016). „Man is to computer programmer as woman is to homemaker? debiasing word embeddings.” *Advances in neural information processing systems*, 29.

Bornmann, L., Haunschild, R., and Mutz, R. (2021). “Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases.” *Humanities and Social Sciences Communications*, 8(1), 1-15.

Braben, D.W. (2004). Pioneering research: A risk worth taking. Hoboken, NJ: Wiley-Interscience.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). “Semantics derived automatically from language corpora contain human-like biases.” *Science*, 356(6334), 183-186.

Calmon, F., Wei, D., Vinzamuri, B., Natesan Ramamurthy, K., and Varshney, K. R. (2017). “Optimized pre-processing for discrimination prevention.” *Advances in neural information processing systems*, 30.

Checco, A., Bracciale, L., Loreti, P., Pinfield, S., and Bianchi, G. (2021). “AI-assisted peer review.” *Humanities and Social Sciences Communications*, 8(1), 1-11.

Crossley, S., Heintz, A., Choi, J. S., Batchelor, J., Karimi, M., and Malatinszky, A. (2023). “A largescaled corpus for assessing text readability.” *Behavior Research Methods,* 55(2), 491-507.

Dal-Ré, R., Caplan, A. L., and Marusic, A. (2019). “Editors’ and authors’ individual conflicts of interest disclosure and journal transparency. A cross-sectional study of high-impact medical specialty journals.” *BMJ open*, 9(7), e029796.

De-Arteaga, M., Romanov, A., Wallach, H., Chayes, J., Borgs, C., Chouldechova, A., ... and Kalai, A. T. (2019, January). “Bias in bios: A case study of semantic representation bias in a high-stakes setting.” *In proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 120-128.

Drori, I., and Te'eni, D. (2024). “Human-in-the-Loop AI Reviewing: Feasibility, Opportunities, and Risks.” *Journal of the Association for Information Systems*, 25(1), 98-109.

Duine, M. (2024). Summary Report APE 2023, 10–12 January, Berlin, Germany Berlin “Re-Visited: Building Technological Support for Scholarship and Scientific Publishing.” *Information Services and Use*, (Preprint), 1-13.

Fanelli, D. (2010). “Do pressures to publish increase scientists’ bias? An empirical support from US States data.” *PLoS ONE,* 5(4), e10271.

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., and Venkatasubramanian, S. (2015). „Certifying and removing disparate impact.” *In proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 259-268.

Feliciani, T., Luo, J., Ma, L., Lucas, P., Squazzoni, F., Marušić, A., and Shankar, K. (2019). “A scoping review of simulation models of peer review.” *Scientometrics,* 121, 555-594.

Fire, M., and Guestrin, C. (2019). “Over-optimization of academic publishing metrics: observing Goodhart’s Law in action.” *GigaScience*, 8(6), giz053.

Frontiers. (2020). *Artificial Intelligence to help meet global demand for high-quality, objective peer-review in publishing*. https://www.frontiersin.org/news/2020/07/01/artificial-intelligence-to-help-meet-global-demand-for-high-quality-objective-peer-review-in-publishing/.

García, J.A., Rodriguez-Sánchez, R. and Fdez-Valdivia, J. (2015). “Bias and effort in peer review.” *J Assn Inf Sci Tec*, 66: 2020-2030.
García, J.A., Rodriguez-Sánchez, R. and Fdez-Valdivia, J. (2016a). “Authors and reviewers who suffer from confirmatory bias.” *Scientometrics*, 109, 1377–1395.
García, J.A., Rodriguez-Sánchez, R. and Fdez-Valdivia, J. (2016b). “Why the referees’ reports I receive as an editor are so much better than the reports I receive as an author?” *Scientometrics*, 106, 967–986.
Garcia, M.B. (2024). “Using AI Tools in Writing Peer Review Reports: Should Academic Journals Embrace the Use of ChatGPT?” *Ann Biomed Eng* 52, 139–140. https://doi.org/10.1007/s10439-023-03299-7.
Garg, N., Schiebinger, L., Jurafsky, D., and Zou, J. (2018). “Word embeddings quantify 100 years of gender and ethnic stereotypes.” *Proceedings of the National Academy of Sciences*, 115(16), E3635-E3644.
Ghosal, T., Verma, R., Ekbal, A., Saha, S., and Bhattacharyya, P. (2018, May). “Investigating impact features in editorial pre-screening of research papers.” *In Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, pp. 333-334.
Grgić-Hlača, N., Zafar, M. B., Gummadi, K. P., and Weller, A. (2018). “Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning.” *In Proceedings of the AAAI conference on artificial intelligence* (32:1).
Grimaldo, F., Paolucci, M. and Sabater-Mir, J. (2018). “Reputation or peer review? The role of outliers.” *Scientometrics*, 116, 1421–1438.
Hardt, M., Price, E., and Srebro, N. (2016). “Equality of opportunity in supervised learning. Advances in neural information processing systems”, *NIPS*, 29.
Hattke, F., Bögner, I., and Vogel, R. (2018). “(Why) Do you trust your reviewers? Influence behaviors, trustworthiness, and commitment to peer review.” *Managementforschung*, (1), 61-86.
Heesen, R., and Bright, L. K. (2021). “Is peer review a good idea?.” *The British Journal for the Philosophy of Science*.
Huh, S. (2023). “Recent issues in medical journal publishing and editing policies: adoption of artificial intelligence, preprints, open peer review, model text recycling policies, best practice in scholarly publishing 4th version, and country names in titles.” *Neurointervention*, 18(1), 2.
Huisman, J., and Smits, J. (2017). “Duration and quality of the peer review process: The author’s perspective.” *Scientometrics*, 113(1), 633-650.
Hutson, M. (2018). “Artificial intelligence faces reproducibility crisis.” *Science*, 359(6377), 725726.
Inam, M., Shiekh, S., Minhas, A. M. K., Vaughan, E. M., Krittanawong, C., Samad, Z., ... and Virani, S. S. (2024). “A review of top cardiology and cardiovascular medicine journal guidelines regarding the use of generative artificial intelligence tools in scientific writing.” *Current Problems in Cardiology*, 102387.
Jackson, J. L., Srinivasan, M., Rea, J., Fletcher, K. E., and Kravitz, R. L. (2011). “The validity of peer review in a general medicine journal.” *PloS one*, 6(7), e22475.
Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, et al. (2023). “Survey of hallucination in natural language generation.” *ACM Computing Surveys*, 55(12), 1–38.
Johnson, R., Watkinson, A., and Mabe, M. (2018). *The STM Report: An overview of scientific and scholarly journal publishing (5th ed.).* STM: International Association of Scientific, Technical and Medical Publishers.

Kamiran, F., and Calders, T. (2012). “Data preprocessing techniques for classification without discrimination.” *Knowledge and information systems*, 33(1), 1-33.

Kamiran, F., Mansha, S., Karim, A., and Zhang, X. (2018). “Exploiting reject option in classification for social discrimination control.” *Information Sciences*, 425, 18-33.

Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J. (2012). “Fairness-aware classifier with prejudice remover regularizer.” *In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012*, Bristol, UK, September 24-28, 2012. Proceedings, Part II 23 (pp. 35-50). Springer Berlin Heidelberg.

Kankanhalli, Atreyi (2024). “Peer Review in the Age of Generative AI”. *Journal of the Association for Information Systems*, 25(1), 76-84. DOI: 10.17705/1jais.00865.

Katzav, J. K., and Vaesen, K. (2017). “Pluralism and peer review in philosophy.” *Philosophers' Imprint*, 17(19), 1-20.

Kim, M. P., Ghorbani, A., and Zou, J. (2019). “Multiaccuracy: Black-box post-processing for fairness in classification.” *In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 247-254.

Korteling, J. H., van de Boer-Visschedijk, G. C., Blankendaal, R. A., Boonekamp, R. C., and Eikelboom, A. R. (2021). “Human-versus artificial intelligence.” *Frontiers in artificial intelligence*, 4, 622364.

Kovanis, M., Trinquart, L., Ravaud, P. et al. (2017). “Evaluating alternative systems of peer review: a large-scale agent-based modelling approach to scientific publication.” *Scientometrics* 113, 651–671.

Lee, C. J. (2012). “A Kuhnian critique of psychometric research on peer review.” *Philosophy of Science*, 79(5), 859-870.

Lee, C. J., Sugimoto, C. R., Zhang, G., and Cronin, B. (2013). “Bias in peer review.” *Journal of the American Society for information Science and Technology*, 64(1), 2-17.

Li, M. (2022). *Desk rejection of submissions to academic journals: an efficient screening process* (Doctoral dissertation, Doctoral dissertation Georgetown University).

Liu, R., and Shah, N. B. (2023). *Reviewergpt? an exploratory study on using large language models for paper reviewing*. arXiv preprint arXiv:2306.00622.

Marks-Anglin, A., and Chen, Y. (2020). “A historical review of publication bias.” *Research synthesis methods*, 11(6), 725-742.

Matt, C., Hoerndlein, C. and Hess, T. (2017). "Let the crowd be my peers? How researchers assess the prospects of social peer review.” *Electron Markets*, 27, 111–124.

Meske, C., Bunde, E., Schneider, J., and Gersch, M. (2022). „Explainable artificial intelligence: objectives, stakeholders, and future research opportunities.” *Information Systems Management*, 39(1), 53-63.

Meth, H., Mueller, B., and Maedche, A. (2015). “Designing a requirement mining system.” *Journal of the Association for Information Systems*, 16(9), 2.

Mrowinski MJ, Fronczak P, Fronczak A, Ausloos M, Nedic O (2017). “Artificial intelligence in
peer review: How can evolutionary computation support journal editors?” *PLoS ONE* 12(9).

Narla, A., Kuprel, B., Sarin, K., Novoa, R., and Ko, J. (2018). “Automated classification of skin lesions: from pixels to practice.” *Journal of Investigative Dermatology*, 138(10), 2108-2110.

Nuijten, M. B., and Polanin, J. R. (2020). ““Statcheck”: Automatically detect statistical reporting inconsistencies to increase reproducibility of meta-analyses.” *Research Synthesis Methods*, 11(5), 574-579.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). “A design science research methodology for information systems research.” *Journal of management information systems*, 24(3), 45-77.
Peters, D.P., and Ceci, S.J. (1982). “Peer-review research: Objections and obligations.” *Behavioral and Brain Sciences*, 5(2), 246–252.
Publons (2018). *Global state of peer review*. Available from: https:// publo ns. com/ static/ Publo
ns-Global-State-Of-Peer-Review-2018. pdf. Cited 2020 Sep 8.
Qin, S., Silaghi, M.C., Menezes, R., Cheung, W. (2014). *Negative Implications of a Power-Law Distribution: A Study on Networks of Scientific Reviewers*. In: Contucci, P., Menezes, R., Omicini, A., Poncela-Casasnovas, J. (eds) Complex Networks V. Studies in Computational Intelligence, vol 549. Springer.
Radzvilas, M., De Pretis, F., Peden, W. et al. (2023). “Incentives for Research Effort: An Evolutionary Model of Publication Markets with Double-Blind and Open Review.” *Comput Econ*, 61, 1433–1476.
Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020, January). “Mitigating bias in algorithmic hiring: Evaluating claims and practices.” *In Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 469-481.
Razack, H. I. A., Mathew, S. T., Saad, F. F. A., and Alqahtani, S. A. (2021). “Artificial intelligence-assisted tools for redefining the communication landscape of the scholarly world.” *Science Editing*, 8(2), 134-144.
Resnik, D. B., Gutierrez-Ford, C., and Peddada, S. (2008). “Perceptions of ethical problems with scientific journal peer review: an exploratory study.” *Science and engineering ethics*, 14, 305-310.
Righi, S., Takács, K. (2017). “The miracle of peer review and development in science: an agent-based model.” *Scientometrics* 113, 587–607
Shanahan, D. (2016, May 23). *A peerless review? Automating methodological and statistical review. Research in Progress Blog.* https://blogs.biomedcentral.com/bmcblog/2016/ 05/23/peerless-review-automatingmethodological-statistical-review/
Shmueli, Galit and Ray, Soumya (2024). “Reimagining the Journal Editorial Process: An AI-Augmented Versus an AI-Driven Future." *Journal of the Association for Information Systems*, 25(1). DOI: 10.17705/1jais.00864
Shneiderman, B. (2020). “Human-centered artificial intelligence: Reliable, safe and trustworthy.” *International Journal of Human–Computer Interaction*, 36(6), 495-504.
Smith, R. (2006). “Peer review: A flawed process at the heart of science and journals.” *Journal of the Royal Society of Medicine*, 99(4), 178-182.
Squazzoni F. and Gandelli C. (2012a). “Peer review under the microscope. An agent-based model of scientific collaboration.” *Proceedings of the 2012 Winter Simulation Conference (WSC)*, Berlin, Germany
Squazzoni F., and Gandelli C. (2012b). “Saint Matthew strikes again: An agent-based model of peer review and the scientific community structure.” *Journal of Informetrics*, Vol. 6, Issue 2, Pages 265-275,
Squazzoni, F. and Gandelli, C. (2013). “Opening the Black-Box of Peer Review: An Agent-Based Model of Scientist Behaviour”. *Journal of Artificial Societies and Social Simulation*, 16 (2) 3
Stanford, P. K. (2019). “Unconceived alternatives and conservatism in science: the impact of professionalization, peer-review, and Big Science.” *Synthese*, 196, 3915-3932.

Sullivan, E. (2022). “Inductive risk, understanding, and opaque machine learning models.” *Philosophy of Science*, 89(5), 1065-1074.
Sun, J., Mavrogenis, A. F., and Scarlat, M. M. (2021). “The growth of scientific publications in 2020: a bibliometric analysis based on the number of publications, keywords, and citations in orthopaedic surgery.” *International Orthopaedics*, 45, 1905-1910.
Suresh, H., and Guttag, J. (2021, October). “A framework for understanding sources of harm throughout the machine learning life cycle.” *In Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1-9.
Susarla, A., Gopal, R., Thatcher, J. B., and Sarker, S. (2023). “The Janus effect of generative AI: Charting the path for responsible conduct of scholarly activities in information systems.” *Information Systems Research*, 34(2), 399-408.
Teixeira da Silva, J. A., and Dobránszki, J. (2015). “Problems with traditional science publishing and finding a wider niche for post-publication peer review.” *Accountability in research*, 22(1), 22-40.
Tennant, J. P., and Ross-Hellauer, T. (2020). “The limitations to our understanding of peer review.” *Research Integrity and Peer Review*, 5(1), 6.
Teplitskiy, M., Peng, H., Blasco, A., and Lakhani, K. R. (2022). “Is novel research worth doing? Evidence from peer review at 49 journals.” *Proceedings of the National Academy of Sciences*, 119(47), e2118046119.
Thelwall, M., Papas, E. R., Nyakoojo, Z., Allen, L., and Weigert, V. (2020). “Automatically detecting open academic review praise and criticism.” *Online Information Review*, 44(5), 1057-1076.
Tuunanen, T., Winter, R., and vom Brocke, J. (2024). “Dealing with complexity in design science research: A methodology using design echelons.” *Management Information Systems Quarterly*, 48(2), 427-458.
UNESCO Science Report: “the race against time for smarter development.” *In: Schneegans S, Straza T, Lewis J, editors. Paris: UNESCO Publishing; 2021*. https:// www.unesco.org/reports/science/2021/ en/statistics.
van Noorden R. (2014). *Global scientific output doubles every nine years. News Blog*, Nature; http:// blogs. nature. com/ news/ 2014/ 05/ global-scien tific-output-doubl es-every-nine-years. html.
van Noorden, R. (2023). “More than 10,000 research papers were retracted in 2023 — a new record.” *Nature*, 624(7992), 479–481.
Wang, H., Li, J., Wu, H., Hovy, E., and Sun, Y. (2022). “Pre-Trained Language Models and Their Applications.” *Engineering*. https://doi.org/https://doi.org/10.1016/j.eng.20 22.04.024
Wang, W., Kong, X., Zhang, J., Chen, Z., Xia, F., and Wang, X. (2016). „Editorial behaviors in peer review.” SpringerPlus, 5, 1-11.
Weber, Ron (2024) “The Other Reviewer: RoboReviewer.” *Journal of the Association for Information*
Systems, 25(1), 85-97. DOI: 10.17705/1jais.00866.
Zhang, B. H., Lemoine, B., and Mitchell, M. (2018). “Mitigating unwanted biases with adversarial learning.” *In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 335-340.
Zhao, J., Zhou, Y., Li, Z., Wang, W., and Chang, K. W. (2018). “Learning gender-neutral word embeddings.” arXiv preprint arXiv:1809.01496.
Zhu, J., Fung, G., Wong, W.H. et al. (2016). “Evaluating the Pros and Cons of Different Peer Review Policies via Simulation.” *Sci Eng Ethics*, 22, 1073–1094.