Industrial and Organizational Psychology (2026), **19**, 32–36
doi:10.1017/iop.2025.10048

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

ReviewGPT: Reducing subjectivity in the review process using AI

Melissa Robertson and Thamengie Richard

Department of Psychology, University of Georgia, Athens, GA, USA
Corresponding author: Melissa Robertson; Email: mrobertson@uga.edu

(Received 1 September 2025; revised 30 September 2025; accepted 8 October 2025)

There is widespread, multidisciplinary evidence that the peer review process often reflects the idiosyncratic perceptions of reviewers, resulting in low interrater reliability and concern regarding the extent to which reviews and journal decisions accurately reflect the scientific contribution of a manuscript. These issues emerge because the peer review process is highly subjective, which makes it rife with opportunities for the expression of various human biases and errors. For example, reviewers may evaluate manuscripts more harshly when they disagree with their own views or preferred theoretical or methodological approaches, place disproportionate emphasis on a single aspect of the manuscript rather than evaluating it in its entirety, or increase the critical tone of their review in an effort to impress the editor. In addition to various cognitive biases, reviewers are also subject to human error and gaps in knowledge. As a result, reviewers may misread or misunderstand authors' work, make inaccurate statements, or suggest improper practices. Because of these biases and errors, relying solely on subjective reviewer comments in editorial decision-making is likely to introduce bias and error into the review process, resulting in the rejection of meritorious scientific contributions, unnecessary revisions, and the preservation of the status quo. As a result, academic careers are derailed, research is slow to disseminate, and scientific innovation is stalled.

We fully agree with Allen et al. (2026) that the subjectivity of reviewers creates significant problems for science that must be addressed. However, in our view, subjectivity is an issue not only for reviewers, but also for authors and editors. As a result, changes to the peer review process that increase author and editorial discretion may inadvertently exacerbate the impact of human error and bias in the peer review process. In this commentary, we describe our concerns with increasing author and editorial discretion, and propose increasing objectivity as a solution to the problem of error and bias in peer review. To facilitate the adoption of this more objective path and address Allen et al.'s recommendation to explore how AI can be used to improve the peer review process, we provide access to two custom GPTs aimed at improving manuscript and review quality—*Friendly Reviewer* and *Review Checker*.

Problems with increasing author and editor discretion in peer review

To address the problems with reviewer bias and error, Allen et al. offer several recommendations aimed at redistributing power from reviewers to authors and editors. Examples include allowing authors to evaluate reviewers, increasing author utilization of appeals, reducing the involvement of

© The Author(s), 2026. Published by Cambridge University Press on behalf of Society for Industrial and Organizational Psychology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/iop.2025.10048 Published online by Cambridge University Press

reviewers in revisions, and empowering editors to discern the importance and weight of reviewer comments. These discretionary actions could potentially be used to screen out low-quality reviewers, mitigate the impact of reviewer error on manuscript decision-making, or protect authors from engaging in inappropriate or unnecessary revisions suggested by reviewers.

Despite these potentially useful applications, we argue that increasing author and editor discretion is likely to exacerbate, rather than mitigate, problems with peer review. This is because authors and editors—like reviewers—are subject to human motivations, biases, and errors that shape their perceptions and decision-making. As a result, adopting procedures that rely on their subjective evaluations is likely to exacerbate, rather than mitigate, bias and error in the peer review process. For example, authors whose papers are rejected are motivated to protect their positive views of themselves and their work, cope with the pain of rejection, and ultimately get their work accepted. As a result, author perceptions of reviewer incompetence or bias or the fairness of the review process are likely to be contaminated by the outcome of the process and the extent of criticism received. These sources of contamination undermine the validity and utility of authors' subjective perceptions (e.g., ratings of reviewers) and decisions based on these perceptions (e.g., decisions to appeal) while also raising the risk of incentivizing problematic reviewer behavior (e.g., providing a positive review while privately recommending rejection, or overlooking significant problems to receive a positive rating). Similarly, editors are subject to error and biases that affect their perceptions and decision-making. For example, editors likely have developed strong heuristics regarding what scholars produce "good work" or are "good reviewers", in addition to professional relationships with many authors and reviewers. Because authors and reviewers are typically not anonymous to the editor, these subjective perceptions and relational motivations may color editors' evaluations, feedback, and decisions. For example, editors may implicitly judge manuscripts written by well-known scholars as better quality or more capable of improvement than those written by scholars they do not know (i.e., familiarity bias). Likewise, editors may implicitly weigh the comments of well-known reviewers more heavily in their decision-making, regardless of the quality of their comments.

We are also concerned about the impact of recommendations that increase reliance on a single individual's subjective opinion. For example, Allen et al. suggest that only editors should be able to read the revised manuscript after initial review, and that reviewers should focus their comments on aspects of the manuscript for which they have expertise. Although we agree that these changes may increase the efficiency of peer review, they may also increase error and bias. As noted by Allen et al., authors and editors are unlikely to have depth of expertise regarding all aspects of a manuscript. The involvement of multiple reviewers is intended to address potential gaps in editors' and authors' knowledge, and make them aware of problems, solutions, assumptions, and contributions that they may not have otherwise considered. When decisions rely on a single decision maker, there is a risk that problems and assumptions may go unseen (resulting in acceptance of work that should be revised or at least challenged) or that important solutions and contributions may go unarticulated (resulting in rejection of work that has high potential for meaningful contribution).

We are also concerned that increasing reliance on editors shifts decision-making power toward individuals who are most well-versed in the implicit norms of the field and the rules of "the publication game"—in other words, those with the highest levels of mastery over the "hidden curriculum" of academic life. This depth of implicit knowledge may make even minor norm violations highly salient, triggering implicit evaluations that a submission is low quality. For example, citation style errors or "overexplanations" of widely held assumptions (e.g., defining a *p*-value or justifying the use of a 7-point response scale) are minor norm violations that may be disproportionately engaged in by less experienced authors or by those who have been socialized into different academic norms. Yet, although these norm violations are highly correctable, they may implicitly signal competence or quality to those most familiar with the conventions of the field, "tipping the scales" toward rejection. In addition, well-established scholars often share

https://doi.org/10.1017/iop.2025.10048 Published online by Cambridge University Press

taken-for-granted assumptions and terminology, which can undermine the introduction of new approaches or ways of thinking and the accessibility of the research to a broader audience. For these reasons, we are concerned that increasing reliance on the editor in the review process may further undermine the diversity, novelty, reach, and accessibility of our science, raising concerns that organizational research will only be understood by a small pool of peers with similar assumptions, training, and worldviews.

We agree with Allen et al., that reviewer bias and error are significant problems plaguing our science. Yet, the voluminous literature on questionable research practices, inaccurate citation, and methodological errors suggests that authors are *also* subject to bias and error, and that these errors often go uncorrected by editors and reviewers alike (Banks et al., 2016; Kepes et al., 2022). Likewise, there is strong multidisciplinary evidence that human decision-makers—regardless of expertise—are subject to a variety of biases that distort their judgment and undermine the fairness and validity of decisions (Hilbert, 2012; Kahneman, 2013). Addressing these issues requires less, not more, reliance on subjective evaluations in peer review.

Reducing subjectivity in the peer review process using AI

Addressing the problems associated with subjectivity in the review process requires increasing objectivity. To accomplish this, we suggest the same approach used in other performance appraisal contexts: The review process can become more objective *and* developmental by clarifying the criterion, developing evaluation measures that align with the criterion, and providing developmental feedback on the extent to which performance aligns with criterion dimensions (DeNisi & Murphy, 2017). Indeed, the importance of objective review criteria is already partially recognized through the use of quantitative dimension-based ratings in the review process and in journal instructions to reviewers. However, quantitative dimensions are often imprecisely defined in review portals, reviewers' qualitative comments hold considerably more weight in the review process relative to their quantitative ratings, and there are no formal systems in place to ensure that reviewers are following journal instructions. These practices have the effect of increasing the potential for error and bias in the review process; mitigating them requires careful attention to performance criteria for both manuscripts and reviewers. Thankfully, comprehensive criteria for high-quality manuscripts (e.g., Aguinis et al., 2018; Appelbaum et al., 2018) and peer reviews (Köhler et al., 2020) already exist. However, without tools to reinforce these standards, we run the risk that they will be ignored. Given constraints on author, reviewer, and editor time and attention, and the breadth of criteria proposed for manuscript and reviewer evaluation, we suggest the use of artificial intelligence (AI) as a tool to aid in the process of structured, criterion-based appraisal and feedback during peer review. Using ChatGPT, we created two custom GPTs to aid authors, reviewers, and editors in the review process: *Friendly Reviewer* and *Review Checker*. Links to the models, example use cases, and checklists are available on the Open Science Framework: https://osf.io/nbcqy/.

Friendly Reviewer is a custom GPT that evaluates quantitative, deductive manuscripts against a checklist of best practices for reviewers in the organizational sciences (see OSF). We developed the checklist by incorporating best practices from a range of sources, including American Psychological Association (APA) journal reporting standards (Appelbaum et al., 2018), journal instructions (*Journal of Applied Psychology*, n.d.), published papers on rigor and contribution in the organizational sciences (Aguinis et al., 2018; Lange & Pfarrer, 2017), and our own disciplinary expertise. Users can upload PDF manuscript drafts and copy and paste the review checklist into the chat, and *Friendly Reviewer* will engage in a structured review based on the checklist, providing a categorical evaluation of whether a checklist item is present, partially present, missing, or not applicable, with the evaluation grounded in specific evidence from the manuscript. In addition, the GPT provides guidance on how to address deficiencies based on Aguinis et al. (2018) and other

https://doi.org/10.1017/iop.2025.10048 Published online by Cambridge University Press

sources. Users can also interact with the GPT for additional targeted feedback; for example, users can ask the GPT to be more critical of particular sections (e.g., theorizing and hypothesis development) or to provide examples or steps for addressing deficiencies. The aim of *Friendly Reviewer* is to improve manuscript quality and address reporting deficiencies before submitting a paper for peer review. *Friendly Reviewer* increases objectivity in the review process by evaluating manuscripts against a standard set of best practices rather than using the idiosyncratic and undefined standards adopted by human authors and reviewers. In doing so, it promotes research quality across a breadth of important criteria while reducing reliance on human knowledge and subjectivity.

Review Checker is a custom GPT that evaluates peer reviews against a checklist of best practices and provides reviewers with developmental feedback. We developed the reviewer checklist based on best practices from Kohler et al. (2020) and our own disciplinary expertise (see OSF). Users can upload their review comments as PDF documents or text into the chat, and *Review Checker* will engage in a critical structured review of the content based on the checklist. As with *Friendly Reviewer*, users can interact with *Review Checker* to revise their reviews (e.g., asking for suggestions for how to rephrase a harshly worded criticism). The aim of *Review Checker* is to improve the quality, comprehensiveness, collegiality, and objectivity of reviews prior to submission and decision recommendations.

Limitations and concerns

Although we believe that AI tools like *Friendly Reviewer* and *Review Checker* have high potential to improve the peer review process, there are also limitations and potential concerns. First, we do not yet have empirical evidence that these tools actually improve manuscript and review quality. Additional testing and validation are needed to determine whether they achieve their intended aims. We agree with Allen et al. that our field is well-positioned to contribute to this scholarly conversation, given our disciplinary expertise in performance appraisal and training, and encourage future research on the effects of AI tools on manuscript and review quality.

Second, authors may be concerned about the privacy and security of their unpublished work. Although we deselected the option to let OpenAI use conversation data to improve their models when creating the GPTs, and GPT developers do not have access to user chats, usage data are still collected and retained by OpenAI (https://openai.com/policies/row-privacy-policy/). To address potential privacy concerns, we recommend that concerned users (a) only upload de-identified content, (b) turn off the account setting "Improve the model for everyone", and (c) delete relevant chats and memories within account settings following use. In addition, it would be a violation of current APA guidelines and publisher policies for reviewers and editors to upload papers they receive for review—we therefore urge users to only submit their *own* work to *Friendly Reviewer*. However, reviewers and editors can use the *Friendly Reviewer* checklist provided on OSF to aid in manuscript review. We also urge reviewers to remove any confidential material from their reviews (e.g., quotes from manuscripts) before using *Review Checker*.

Third, like human reviewers, AI is subject to error and bias and can make inaccurate conclusions or suggestions, or overlook important information. In addition, although they rely on disciplinary best practices, the GPTs may impose rigid evaluation standards that may not be appropriate for all manuscripts. For this reason, we caution users to verify the accuracy, appropriateness, and comprehensiveness of the output. In addition, for both models, users can add or revise checklist criteria by pasting a revised checklist into the chat and instructing the GPT to rely on the revised, rather than the original, checklist. Through checklist revision, the GPTs can accommodate novel methods, changes in best practices, or different types of manuscripts (e.g., registered reports, research proposals).

Fourth, these tools are currently targeted toward the development of individual authors and reviewers rather than systemic change (e.g., editors, journals, and publishers). Given privacy

https://doi.org/10.1017/iop.2025.10048 Published online by Cambridge University Press

concerns and the rapidly evolving nature of these tools and the policies surrounding them, we felt this was an appropriate first step. We note that users should be cautious before using *any* AI tool for research purposes, as different journals have different guidelines for use. For example, APA journals currently allow authors to use AI with disclosure and citation, whereas AOM journals currently only allow AI use for specific purposes (i.e., as spelling/grammar aids or in data collection/analysis; see links to current policies on OSF). Authors and reviewers should therefore carefully check journal policies prior to using any AI tool, including *Review Checker* and *Friendly Reviewer*.

Finally, some may be concerned that the ultimate outcome of using AI in peer review will be the elimination of humans altogether in the publication process, with AI-generated manuscripts being reviewed by AI, revised by AI, and summarized by AI. We do not think AI models are sufficiently "knowledgeable" regarding disciplinary standards and norms to independently accomplish these tasks at present. However, we welcome the introduction of any innovations—AI or human—that increase the rigor, reproducibility, fairness, accessibility, and impact of our science.

Competing interests. We have no conflicts of interest to disclose.

References

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals*, **12**(1), 83–110. https://doi.org/10.5465/annals.2016.0011

Allen, T. D., French, K., Avery, D. R., King, E., & Wiernik, B. M. (2026). Developmental reviewing: Is it really good for science? *Industrial and Organizational Psychology*, **19**(1), 1–15.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology the APA publications and communications board task force report. *American Psychologist*, **73**(1), 3–25. https://doi.org/10.1037/amp0000191

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology*, **31**(3), 323–338. https://doi.org/10.1007/s10869-016-9456-7

DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? *Journal of Applied Psychology*, **102**(3), 421–433. https://doi.org/10.1037/apl0000085

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, **138**(2), 211–237. https://doi.org/10.1037/a0025940

Journal of Applied Psychology. (n.d.). Manuscript preparation instructions. Retrieved August 26, 2025, from https://www.apa.org/pubs/journals/features/apl-manuscript-checklist.pdf

Kahneman, D. (2013). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kepes, S., Keener, S. K., McDaniel, M. A., & Hartman, N. S. (2022). Questionable research practices among researchers in the most research-productive management programs. *Journal of Organizational Behavior*, **43**(7), 1190–1208. https://doi.org/10.1002/job.2623

Köhler, T., González-Morales, M. G., Banks, G. C., O'Boyle, E. H., Allen, J. A., Sinha, R., Woo, S. E., & Gulick, L. M. V. (2020). Supporting robust, rigorous, and reliable reviewing as the cornerstone of our profession: Introducing a competency framework for peer review. *Industrial and Organizational Psychology*, **13**(1), 1–27. https://doi.org/10.1017/iop.2019.121

Lange, D., & Pfarrer, M. D. (2017). Editors' comments: Sense and structure—The core building blocks of an *AMR* article. *Academy of Management Review*, **42**(3), 407–416.

Cite this article: Robertson, M. & Richard, T. (2026). ReviewGPT: Reducing subjectivity in the review process using AI. *Industrial and Organizational Psychology* **19**, 32–36. https://doi.org/10.1017/iop.2025.10048

https://doi.org/10.1017/iop.2025.10048 Published online by Cambridge University Press