Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Fighting Fire with Fire: Infusing Artificial Intelligence into Peer Review to Sustain Quality Scholarship

Hemant K. Bhargava, Sarah H. Bana, Zhe Zhang, Laura Brandimarte, Vidyanand Choudhary, J. Frank Li, Pantelis Loupos, Daniel Zantedeschi

To cite this article:
Hemant K. Bhargava, Sarah H. Bana, Zhe Zhang, Laura Brandimarte, Vidyanand Choudhary, J. Frank Li, Pantelis Loupos, Daniel Zantedeschi (2026) Fighting Fire with Fire: Infusing Artificial Intelligence into Peer Review to Sustain Quality Scholarship. Management Science

Published online in Articles in Advance 07 Aug 2026

. https://doi.org/10.1287/mnsc.2026.00184

https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–13
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Fighting Fire with Fire: Infusing Artificial Intelligence into Peer Review to Sustain Quality Scholarship

Hemant K. Bhargava,[a,b,*] Sarah H. Bana,[c,d,e,*] Zhe Zhang,[f,*] Laura Brandimarte,[g,*] Vidyanand Choudhary,[h,*] J. Frank Li,[d,i,*] Pantelis Loupos,[a,*] Daniel Zantedeschi[j,*]

[a] Graduate School of Management, University of California, Davis, California 95616; [b] Center for Analytics and Technology in Society, University of California, Davis, California 95616; [c] Argyros College of Business and Economics, Chapman University, Orange, California 92866; [d] Stanford Digital Economy Lab, Stanford University, Stanford, California 94305; [e] Department of Management, London School of Economics and Political Science, London WC2A 2AE, United Kingdom; [f] Rady School of Management, University of California, San Diego, La Jolla, California 92093; [g] Eller College of Management, University of Arizona, Tucson, Arizona 85721; [h] Paul Merage School of Business, University of California, Irvine, California 92697; [i] Sauder School of Business, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada; [j] Muma College of Business, School of Information Systems, University of South Florida, Tampa, Florida 33620

*Corresponding authors

Contact: hemantb@ucdavis.edu, https://orcid.org/0000-0002-9268-2234 (HKB); sarah.bana@gmail.com, https://orcid.org/0000-0002-6761-4862 (SHB); z9zhang@ucsd.edu, https://orcid.org/0000-0002-8677-6955 (ZZ); lbrandimarte@arizona.edu, https://orcid.org/0000-0002-1039-5240 (LB); veecee@uci.edu, https://orcid.org/0000-0001-8812-4232 (VC); frank.li@sauder.ubc.ca, https://orcid.org/0000-0001-5191-4900 (JFL); ploupos@ucdavis.edu, https://orcid.org/0000-0003-4414-9790 (PL); danielz@usf.edu, https://orcid.org/0000-0001-6241-9809 (DZ)

Received: January 14, 2026
Revised: April 8, 2026
Accepted: May 24, 2026
Published Online in Articles in Advance: August 7, 2026

https://doi.org/10.1287/mnsc.2026.00184

Abstract. Adoption of artificial intelligence (AI) by authors has accelerated production of academic articles and increased submission rates to journals, thereby straining review capacity and hurting journal outcome metrics, such as turnaround time and decision accuracy. Academic journals face an imperative to improve review quality and productivity by incorporating generative AI tools in the review workflow. As a concrete first step toward this idea, we outline a particular workflow that deploys large language models as a structured, trained, frontline reviewer. The workflow underlines that AI-produced evaluations must be transparent and contestable by authors. We emphasize the need for AI-infused peer review to be journal specific, designed for efficiency, accuracy, and accountability, within a human-in-the-loop oversight framework. We build a mathematical model of journal operations to compare outcomes under a status quo no-AI workflow and the proposed AI-infused workflow. The model captures how governed AI reconfigures human reviewer effort and illuminates conditions under which the AI-infused workflow can improve vital metrics, such as turnaround time and decision accuracy. Our essential point is that journals need to adopt a deliberate and institutionally governed approach for using AI in the review process. We recognize that identifying an "optimal" AI-infused peer review workflow or even definitively projecting the consequences of one will require substantial experimentation to calibrate and configure the workflow across multiple alternative designs.

History: This discussion paper was accepted by Christoph Loch.
Funding: D. Zantedeschi acknowledges support for this research from the Muma College of Business Dean's Research Fund.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2026.00184.

Keywords: Large Language Models • Human-AI Collaboration • Academic Publishing • Peer Review • Human-AI Collaboration

1. Introduction and Motivation

Peer review is essential to academic research and the profession. It requires multiple expert reviewers with specialized knowledge to devote substantial time in assessing the quality of submitted articles. As with any operations system, peer review faces substantial stress when the supply of submissions overcomes the capacity to process them. Such a scenario is occurring now as artificial intelligence (AI) tools, especially generative AI and large language models (LLMs), have accelerated authors' ability to generate research articles. For instance, Google's AI Co-Scientist has successfully generated novel, literature-grounded hypotheses for biology laboratories, some of which aligned with hypotheses developed independently by the same laboratories (Gottweis et al. 2025). Similarly, Gans (2025) reports that an AI coauthor helped draft a full paper within a single day. Mathematician Terence Tao has likewise explored AI

collaboration on proofs and analyses (Tao 2025). Ludwig and Mullainathan (2024) and Manning et al. (2024) indicate that AI holds potential for generating and testing hypotheses within the social sciences.

As a result, AI appears to make authors more productive. Kusumegi et al. (2025) provide mounting evidence that LLMs boost the output of scientific papers by 23.7%–89.3% depending on the field and the author's background. Xu and Yang (2026) show that an agentic AI workflow can reanalyze and replicate studies with a high success rate conditional on accessible data and code, which should raise author productivity by letting researchers produce, verify, and revise more work in less time. More recently, Project Autonomous Policy Evaluation (APE)[1] illustrated the impact of this productivity gain in practice; its public dashboard reports more than 400 AI-generated economics papers produced within two months. Unlike the studies above, which focus on AI reshaping individual researchers' workflow, APE demonstrates what happens when AI-assisted authorship is deployed as a systematic pipeline; the paper output volume of a small team begins to drastically scale.

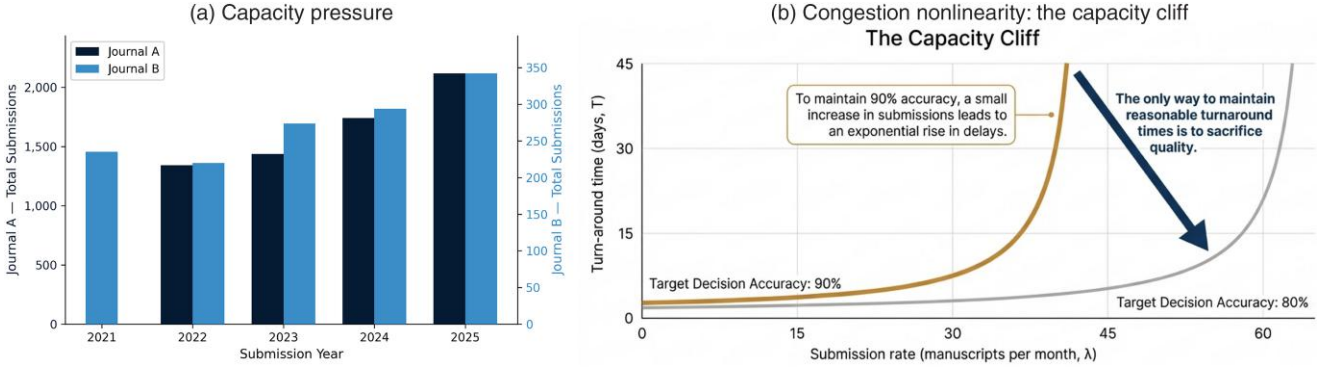
A recent editorial commentary in *Information Systems Research* underscores that generative AI is already reshaping research practices and disciplinary expectations while warning that the community must develop new norms to manage these shifts responsibly (Gopal et al. 2025). From a recent report by a top publisher: "Peer review … is under unprecedented pressure. Too many submissions are funnelled through a limited pool of reviewers, creating unsustainable workloads and threatening the effectiveness, speed and fairness of peer review" (Cambridge University Press 2025, p. 11). This article articulates the operations crisis confronting academic peer review, develops a formal model to capture the essence of the process, proposes an AI-infused peer review workflow, and applies an extended model to examine under what conditions and to what extent this new approach can help address the crisis (Cambridge University Press 2025). Our process and model exemplify how AI can be integrated, but we recognize that designing the optimal way to employ AI in the review process requires careful design, calibration, and experimentation.

1.1. Impending Crisis: Operational Strain and the "Shadow AI" Problem

As AI increases submissions, academic journals are finding it harder to distinguish high-quality work while maintaining timely decisions (Gartenberg et al. 2026). *Management Science* submissions increased from 4,710 in 2024 to 5,735 in 2025, a rise of 22%. At two major journals, submissions increased by roughly 50% between 2022 and 2025 (panel (a) of Figure 1). Recent concerns on manuscript and grant time to decisions have been noted by government agencies and in the academic press (National Institutes of Health 2025, Richardson et al. 2025). Operations theory suggests a nonlinear congestion effect; as utilization approaches one, small inflow shocks translate into large delay increases—a *capacity cliff*. Panel (b) of Figure 1 illustrates this pattern based on a stylized operations model of journals (presented in Section 2), with nonlinear deterioration in the critical metrics that define a journal's performance, namely turnaround time and decision accuracy.

A related problem is *shadow AI*, where reviewers (and editors) already use AI tools but in an opaque and ungoverned manner. Although several academic journals and publishers restrict these AI tools in peer review because of concerns over confidentiality, accountability, and accuracy (Bhargava et al. 2025, Peters and Chin-Yee 2025), informal reports and detection-based studies have documented that a nontrivial fraction of reviews

Figure 1. (Color online) Capacity Pressure and a Governance Hypothesis



Notes. Panel (a) shows growth in submissions at *Organization Science* (Pierce 2026) and another top journal in the field. Panel (b) provides a stylized illustration of the nonlinear delay-quality trade-off near capacity; assuming fixed capacity, as utilization rises, journals face pressure to either accept longer delays or relax effective standards to maintain throughput.

(in computer science conferences) contain AI-generated text (Liang et al. 2024a, Latona et al. 2025). For journal leadership, shadow AI is difficult to observe or measure. When multiple reviewers privately rely on similar general-purpose LLMs, their review errors become structurally correlated, eroding the independence of judgment that journals rely upon. Moreover, reviewer-initiated use of public AI interfaces can create uncontrolled data retention, use a variety of unknown prompts, or have third-party logging. A governed, journal-specific AI system does not eliminate shadow AI, but it may reduce reviewers' incentive to seek external AI support while making the journal's own AI use transparent and its induced correlations estimable.

1.2. Proposed Solution: Managed, AI-Augmented Peer Review

Although AI use for peer review is problematic when it is opaque, uncontrolled, heterogeneous, and undisclosed, there is credible evidence of an upside to using AI. Comparative studies report substantial overlap between LLM critiques and human critiques (Liang et al. 2024b). They can be helpful in identifying clarity problems, missing reporting elements, internal inconsistencies, and other structured checks (Berente and Recker 2025).

The managerial problem is, therefore, not "*AI or no AI*" but how to move from opaque adoption to managed adoption of AI: a journal-provided, journal-scoped AI diagnostic tool applied *before* human review paired with a short author right of response and instrumented through structured logging. The AI output is explicitly nonfinal; it is an input to editorial judgment, not a recommendation and not a substitute for reviewers. We propose that the AI reviewer and process be *journal specific*. General-purpose AI tools lack alignment with a particular journal's standards, methodologies, and values. Therefore, journals will scope AI to tasks where criteria can be specified and outputs can be validated, and they may prevent scope creep into value judgments.

The AI tool is a design choice that can be tailored, trained, and vetted according to the specific criteria of the target journal, potentially focusing its capabilities on aspects appropriate for that journal (e.g., mechanical checks, writing quality, or methodological soundness) while freeing up resources for human oversight and judgment. Journals must customize their AI tool; some journals may constrain the AI to auditable rubrics, and others may allow for less constrained AI reviews. Confidentiality and intellectual property concerns further strengthen the case for a journal-led workflow.

Managed AI (versus both shadow AI and permissive policies) enables the measurement of the effects of using AI. It enables treatment of AI as an informational input with two estimable properties: *reliability* (noise/precision of the AI signal) and *error correlation* (the degree to which review errors are correlated with each other once the AI review is observed). These quantities could in principle be recovered from records or pilot rollouts and used to design workflows that meet a journal's standards. Section 3 presents a proposed workflow, which we formalize by extending our journal review model by infusing AI into the process. Given a manuscript to review, humans (or humans and AI in the governed-AI case) produce noisy unbiased estimates of manuscript quality, and the journal seeks precise estimates to improve its decision accuracy. The AI review (and initial author response) affects two channels: the productivity of reviewer time and the weight that the AI review carries in the reviewer's final report. The risk is *error correlation*; when reviewers see the AI report, their errors may become correlated with the AI's errors (anchoring).

2. A Benchmark Model of Journal Operations

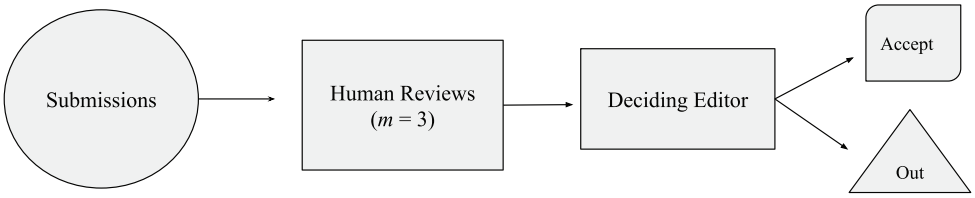
We frame an academic journal's review process as a resource-constrained estimation problem. The journal receives a flow of submissions and through editorial review, must estimate manuscript quality accurately enough to make sound accept/reject decisions (see Figure 2).

The key performance dimension is *estimation-error variance*: how precisely the deciding editor (DE) can assess manuscript quality given limited reviewer time. Improved precision allows the journal to more often correctly accept good papers and reject weak ones. As mentioned earlier in Section 1.1, as submission rates rise—particularly with AI-accelerated authorship—the reviewer time available per manuscript shrinks, putting upward pressure on estimation-error variance and threatening decision quality.

Formally, we model journal review as a stylized single-round process. We assume that manuscript quality is a scalar latent variable Q. Each manuscript is assigned to m^H human reviewers, where H denotes human parameters and outcomes in the human-only benchmark case. Each reviewer i provides a quality signal $\hat{Q}_i^H$ to a deciding editor. Reviewer signals are noisy, with precision improving in reviewer skill and time allocated.[2] The DE aggregates the reviewer signals to form an estimate of manuscript quality, and the journal sets a target estimation-error variance σ^2_{tar} that determines the required review thoroughness. The DE's only choice variable is the time allocated per reviewer, t; all other parameters are taken as given.

This parsimonious setup blends multiround review into one stage so that we can focus on estimation

Figure 2. Simplified Model of the Current Review Process



Notes. For clarity, we model the review process as a single-stage process. We also visually omit potential desk rejection for clarity.

precision rather than iterative paper improvement. A valuable extension may be to consider improvements to paper quality endogenous to the review process. The notation is summarized in Table 1.

2.1. Human Reviewer Signals and Aggregation

Reviewer i produces the signal

$$\hat{Q}_i^H = S_i^H = Q + \varepsilon_i^H, \qquad \mathbb{E}[\varepsilon_i^H] = 0,$$
$$\mathrm{Var}(\varepsilon_i^H) = \frac{1}{\theta_H \tau^H(t^H)}, \qquad i = 1, \ldots, m^H,$$

where $t^H > 0$ is the time allocated to each reviewer and $\tau^H(t^H)$ is the time-effect function, which is increasing and concave: $(\tau^H)'(t^H) > 0$, $(\tau^H)''(t^H) \le 0$, $\tau^H(0) = 0$, and $\tau^H(\infty) \to 1$. The time-effect function translates time invested by the reviewer into increasing review precision (by reducing ε_i^H). Although this function may be sigmoidal, we assume that the journal operates in the concave portion; hence, $\tau^H(t^H)$ is concave in t^H. Note that although $\hat{Q}_i^H = S_i^H$ in this benchmark, this will not be the case in the AI-assisted regime.

Assume that reviewer errors $\{\varepsilon_i^H\}_{i=1}^{m^H}$ are conditionally independent given Q. The human-panel mean and estimation-error variance in the human-only benchmark is

$$\bar{Q}^H := \frac{1}{m^H} \sum_{i=1}^{m^H} \hat{Q}_i^H; \qquad \sigma^{2,H}_{\mathrm{post}}(m^H, t^H) = \frac{1}{m^H \theta_H \tau^H(t^H)}. \quad (1)$$

The variance is decreasing in both t^H and m^H (with diminishing incremental gains from additional reviewers (see the Online Appendix). A smaller estimation-error

Table 1. Notation Summary

Symbol	Definition
Q	Latent quality of each manuscript
m^H, m^A	Number of human reviewers per manuscript
$\hat{Q}_i^H$, $\hat{Q}_i^A$	Estimated manuscript quality by reviewer i that they submit to the DE
θ_H, θ_Z	Quality parameter (θ_H for human reviewers, θ_Z for AI)
t^H, t^A	Reviewer time allocated per manuscript (chosen endogenously by the DE)
$\tau^H(\cdot)$, $\tau^A(\cdot)$	Time-effect function: maps time to precision gain
H_{tot}	Total reviewer-hour budget per period
S_i^H, S_i^A	Human reviewer's independent signal of manuscript quality
ε_i^H, ε_i^A	Error of human signal estimates
$\bar{Q}^H$, $\bar{Q}^A$	Panel mean: DE's aggregate quality estimate
$t^{*,H}$, $t^{*,A}$	Optimal per-reviewer time that achieves target variance σ^2_{tar}
$\sigma^{2,H}_{\mathrm{post}}$, $\sigma^{2,A}_{\mathrm{post}}$	Estimation-error variance of DE's quality estimate
σ^2_{tar}	Target estimation-error variance ("journal quality") set by the editor
AI-specific notation	
$Z(\theta_Z)$	AI review's signal of manuscript quality
ε^Z	Error of AI review signal estimate
$\sigma^{2,Z}(\theta_Z)$	Variance of AI review signal error
$\omega(\theta_Z)$	Weight that reviewer i places on the AI review signal, $\omega(\theta_Z) \in [0,1)$
$\rho_{HH}(t^A, \theta_Z)$	Pair-wise between-reviewer error correlation induced by the shared AI review
Queueing and capacity notation	
λ	Submission arrival rate (Poisson process)
μ^H, μ^A	Service rate (throughput): $\mu^r = H_{\mathrm{tot}}/(m^r t^{*,r})$ for $r \in \{H, A\}$
$T^H(\lambda;\mu)$, $T^A(\lambda;\mu)$	Expected journal response time: $1/(\mu^r - \lambda)$ for $r \in \{H, A\}$

Notes. Superscripts H and A index the review process in the human-only benchmark and AI-assisted regime, respectively. Subscripts H and Z on θ denote human-specific and AI-specific quality, respectively.

variance $\sigma^{2,H}_{\text{post}}$ enables more accurate acceptance decisions. Given a target estimation-error variance $\sigma^2_{\text{tar}} > 0$, the DE's optimal (minimal) choice of per-reviewer time is $t^{*,H} = (\tau^H)^{-1}(1/(m^H\theta_H\sigma^2_{\text{tar}}))$.

2.2. Throughput and the Capacity Cliff

Given a total reviewer-hour budget H_{tot} per period, the journal's service rate (throughput) is $\mu^H := H_{\text{tot}}/(m^H t^{*,H})$. Manuscripts arrive as a Poisson process with rate $\lambda > 0$. By standard results from queueing theory, the expected journal response time $T^H(\lambda;\mu^H) = 1/(\mu^H - \lambda)$ is strictly increasing and convex in λ, and it diverges as $\lambda \uparrow \mu^H$. This capacity cliff property means that as AI-accelerated authorship drives up submissions, journals operating near their human-only capacity μ^H will exhibit longer turnaround times. This puts downward pressure on decision accuracy (i.e., upward pressure on estimation-error variance, $\sigma^{2,H}_{\text{post}}$) as journal reviewers spend less time per submission to keep turnaround times under control even as submission volume increases.

Figure 1 from Section 1 illustrates this emerging crisis. At relatively low submission volumes, journals comfortably maintain high decision accuracy with short turnaround times. As the submission rate grows, the system reaches the capacity cliff; reviewer workloads consume the journal's capacity, leading to rapidly escalating turnaround times and processing backlogs. Journals face the difficult choice of increasing reviewer resources, reducing the number of reviewers per manuscript, or accepting lower decision accuracy.

3. Fighting Fire with Fire: AI-Infused Peer Review

In its recent review of operational pressures facing academic journals, the Cambridge University Press underlines the role of technology in improving productivity of peer review (Cambridge University Press 2025). Calling for "new solutions that address peer review at scale," the review argues that "responsible use of AI [can] streamline administrative and technical aspects of review, allowing human expertise to focus on what makes research interesting and important" (Cambridge University Press 2025, p. 12). We propose an AI-infused hybrid editorial workflow, which strategically incorporates AI early in the review process, potentially reconfiguring the efforts of human reviewers more toward high-level critical tasks. AI influence is explicit and auditable. There is no intent to automate editorial judgment; instead, our proposed approach suggests leveraging AI to support human judgment with the aim to improve the sustainability of the peer review process. We recognize that identifying an "optimal" AI-infused peer review workflow or even definitively projecting the consequences of one will require substantial experimentation to calibrate and configure the workflow across multiple alternative designs.

3.1. Proposed AI-Augmented Journal Operations Workflow

The core of our proposal is a two-stage process in which the evaluative role of AI is tailored to the domain, methods, and value systems of the journal.

1. Initial AI review and author response. Upon submission, every paper undergoes an immediate, journal-specific AI review. The AI tool generates structured critique and feedback for the authors, who then prepare a response document (or withdraw the paper) to address or rebut the points raised by the AI. This initial exchange serves to identify and correct basic issues, increases clarity, and ensures a baseline level of quality before the paper enters the human review process.

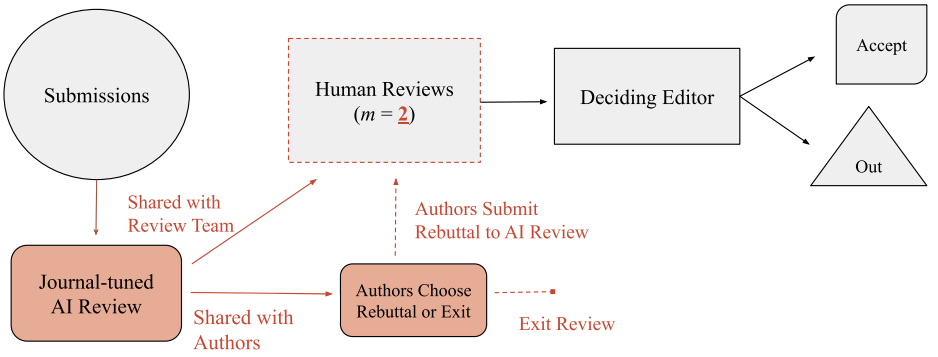
2. Human review with AI context. When the paper proceeds, human reviewers receive not only the original manuscript but also, the AI-generated review and the authors' response. This augmented context allows them to save time on mechanical checks, which now simply need human oversight, and focus their intellectual energy on aspects of the review that require deeper human involvement, such as contribution to the literature and impact.[3]

Our proposal specifically implies that the AI review focuses on activities that have been vetted with respect to that journal. For instance, a particular journal might use AI to examine writing quality or a rigorous methodology that provides evidence of these claims.[4] Crucially, this proposal shifts the use of AI from an informal, unobserved tool used idiosyncratically by individuals to a *journal-provided* and *journal-scoped* assessment that every submission sees in the same way. The AI and human reviewers work together as explained below and depicted in Figure 3. The manuscript itself remains fixed at this stage; the response is an attachment, not a revision. Reviewers receive three objects: the manuscript, the AI assessment, and the author response.

3.2. A Model of AI-Infused Peer Review

We extend the model of peer review to capture the AI-infused peer review process described in Figure 3. The model captures two key tensions. (1) AI with quality θ_Z can make reviewer time more productive (captured through τ^A) and improve the precision of human reviews through an informative signal $Z(\theta_Z)$ comprising the AI review and author response, but (2) it can reduce the diversification benefit of having multiple independent reviewers because all reviewers rely on the same AI review (through the weighting term $\omega(\theta_Z)$), creating potential correlation ρ_{HH} in their errors.

Figure 3. (Color online) Governed AI Workflow (Proposed) vs. Status Quo (Shown Previously in Figure 2)



Notes. An AI review is generated by the journal upon receiving the manuscript. Authors can choose to exit or write a rebuttal that addresses concerns or corrects any mistakes. Reviewers receive the original paper alongside the AI review and the AI rebuttal.

The AI review affects the editorial decision only indirectly through its effect on the human reviewers' reports. Our model's focus is on reviewer time. We do not model DE time per review nor model authors' withdrawals or desk rejections. We also do not model the author response to the AI review, which adds elapsed time to the process but not to reviewer time. Lastly, we do not model the potential effect of AI-assisted review on the mean number of review rounds of a multiround process.

3.2.1. AI Signal. Let the AI review signal be $Z(\theta_Z) = Q + \varepsilon^Z$ with unbiased mean $\mathbb{E}[\varepsilon^Z] = 0$ and variance $\mathrm{Var}(\varepsilon^Z) = \sigma^{2,Z}(\theta_Z)$, where $(\sigma^{2,Z})'(\theta_Z) < 0$. Without loss of generality (because of the freedom in defining the scale of θ_Z), set $\sigma^{2,Z}(\theta_Z) = 1/\theta_Z$.

3.2.2. Human Reviewers in the AI-Assisted Regime. Let $t^A > 0$ denote the time spent by each of the m^A reviewers in the AI-assisted regime. Reviewer i's assessment $\hat{Q}_i^A$ can be viewed as a convex combination of the reviewer's own assessment S_i^A plus the AI review given to them:[5]

$$\hat{Q}_i^A = (1-\omega(\theta_Z))\, S_i^A + \omega(\theta_Z)\, Z(\theta_Z), \qquad i = 1,\ldots,m^A, \tag{2a}$$

$$S_i^A = Q + \varepsilon_i^A, \quad \mathbb{E}[\varepsilon_i^A] = 0, \quad \mathrm{Var}(\varepsilon_i^A) = \frac{1}{\theta_H \tau^A(t^A,\theta_Z)} \tag{2b}$$

$$\text{hence,} \quad \hat{Q}_i^A = Q + \omega(\theta_Z)\varepsilon^Z + (1-\omega(\theta_Z))\varepsilon_i^A, \tag{2c}$$

where $\omega(\theta_Z) \in [0,1)$ measures the weight placed on the AI review (treated as exogenous). Note that θ_H appears in both regimes because the human reviewer's intrinsic ability is regime invariant. In the AI-assisted regime, the reviewer's signal S_i^A uses the τ^A time-effect function, which we write as $\tau^A(t^A,\theta_Z) > 0$ (i.e., we capture dependence of time t and AI quality θ_Z).[6]

We assume mutual independence of intrinsic reviewer errors ($\mathrm{Cov}(\varepsilon_i^A, \varepsilon_j^A) = 0$ for $i \neq j$) and between the reviewer and AI errors $\mathrm{Cov}(\varepsilon_i^A, \varepsilon^Z) = 0$. We also assume that there is no additional distortion term induced by AI exposure. AI and reviewer signals are treated as unbiased estimates of manuscript quality ($\mathbb{E}[\varepsilon^Z] = \mathbb{E}[\varepsilon_i^A] = 0$). The model captures only the effect of AI-assisted review on the variance channel and does not capture the directional bias (mean-effect) channel.

Under these assumptions, AI can worsen the editorial decision through two channels. (i) The shared AI error ε^Z introduces between-reviewer error correlation that reduces the diversification benefit of multiple reviewers, and (ii) the time-effect function $\tau^A(t^A,\theta_Z)$ may be lower than $\tau^H(t^H)$ if the AI review and author response prove distracting or misleading.

3.2.3. Induced Covariance Across Human Reviewers. Because all reviewers observe the same AI review, the errors of their final report become correlated through the shared AI error term ε^Z; the signals remain unbiased. For a given reviewer i,

$$\mathrm{Var}(\hat{Q}_i^A - Q) = \omega(\theta_Z)^2\, \sigma^{2,Z}(\theta_Z) + \frac{(1-\omega(\theta_Z))^2}{\theta_H \tau^A(t^A,\theta_Z)}.$$

For two distinct reviewers $i \neq j$, $\mathrm{Cov}(\hat{Q}_i^A - Q,\ \hat{Q}_j^A - Q) = \omega(\theta_Z)^2\, \sigma^{2,Z}(\theta_Z)$.

The pair-wise between-reviewer error correlation is

$$\rho_{HH}(t^A,\theta_Z) = \frac{\omega(\theta_Z)^2\, \sigma^{2,Z}(\theta_Z)}{\omega(\theta_Z)^2\, \sigma^{2,Z}(\theta_Z) + \frac{(1-\omega(\theta_Z))^2}{\theta_H \tau^A(t^A,\theta_Z)}}.$$

This error correlation is generated within the workflow itself because all reviewers see the same AI review. (The formal derivation is in the Online Appendix.) The error correlation ρ_{HH} is governed by the relative magnitude of shared versus idiosyncratic error variance. When $\omega(\theta_Z)$ is small, reviewers rely primarily on their own individual assessments and $\rho_{HH} \approx 0$. As $\omega(\theta_Z) \to 1$, $\rho_{HH} \to 1$.

3.2.3.1. Final Signal to DE. This is the observed panel mean

$$\bar{Q}^A := \frac{1}{m^A}\sum_{i=1}^{m^A}\hat{Q}_i^A = Q + \omega(\theta_Z)\,\varepsilon^Z + (1-\omega(\theta_Z))\,\bar{\varepsilon}^A, \quad (3)$$

where $\bar{\varepsilon}^A := (1/m^A)\sum_{i=1}^{m^A}\varepsilon_i^A$. The DE's estimation-error variance is

$$\sigma^{2,A}_{\text{post}}(t^A,\theta_Z) = \omega(\theta_Z)^2\,\sigma^{2,Z}(\theta_Z) + \frac{(1-\omega(\theta_Z))^2}{m^A\theta_H\tau^A(t^A,\theta_Z)}. \quad (4)$$

Online Appendix B.4 shows that the first term is a floor for the nondiversifiable AI error variance and invariant to m^A and t^A; the second term, the diversifiable human error variance, shrinks with m^A and t^A.

3.2.4. Comparative Statics with Respect to AI Quality.

The effect of AI quality on the estimation-error variance is not unambiguously beneficial. Differentiating $\sigma^{2,A}_{\text{post}}(t^A,\theta_Z)$ with respect to θ_Z, the sign depends on three forces:

1. the effect of AI quality on the time-effect function through $\partial\tau^A/\partial\theta_Z$,
2. the direct improvement in AI precision through $(\sigma^{2,Z})'(\theta_Z) < 0$, and
3. the weighting on the AI signal through $\omega'(\theta_Z)$.

Force (1) is ambiguous in sign; better AI and author response may help reviewers more effectively review the paper ($\partial\tau^A/\partial\theta_Z > 0$), or the additional complexity may hinder their review effectiveness ($\partial\tau^A/\partial\theta_Z < 0$). Force (2) is unambiguously beneficial, with higher AI quality providing a more accurate AI signal $Z(\theta_Z)$. Force (3) creates a tension; higher weight on the AI signal amplifies the nondiversifiable common error but reduces diversifiable idiosyncratic noise. Better AI improves estimation precision when the direct precision gain (force (2)) and any time-productivity gains (force (1)) dominate the increased error correlation from higher weighting (force (3)). (A formal necessary and sufficient condition is stated in the Online Appendix.)

3.3. Main Analytical Results

We first record a governance result—that optimal reliance on the AI review is interior (Proposition 1)—and then, establish two threshold results that characterize when AI quality is sufficient to deliver concrete resource savings. The first is an intensive-margin threshold: the AI quality above which better AI reduces the per-reviewer time needed to achieve a fixed estimation-error target. The second is an extensive-margin threshold: the AI quality above which the journal can achieve the same quality target with fewer human reviewers. (Formal proofs are in Online Appendix B.)

Proposition 1 (Optimal Reliance on AI). *Fix the number of reviewers m^A, per-reviewer time t^A, and AI quality θ_Z, and write the AI signal noise as $A := \sigma^{2,Z}(\theta_Z)$ and the human noise as $B^A := 1/[m^A\theta_H\tau^A(t^A,\theta_Z)]$. The DE's estimation-error variance as a function of reviewer reliance $\omega \in [0,1)$, $\sigma^{2,A}_{\text{post}}(\omega) = \omega^2 A + (1-\omega)^2 B^A$, is strictly convex and minimized at the interior weight $\omega^* = B^A/(A+B^A) \in (0,1)$. Overreliance ($\omega > \omega^*$) inflates the common AI-induced error and the crossreviewer correlation ρ_{HH}; underreliance forgoes the informative AI signal.*

Proposition 1 characterizes the variance-minimizing weight. This is increasing in quality θ_Z, holding the time-effect channel τ^A fixed. It is a benchmark, not a weight set directly. The condition under which AI-assisted review beats the human-only panel and its proof are given as Proposition 5 in Online Appendix D.3.

3.3.1. Target Variance and Required Reviewer Time.

Let $\sigma^2_{\text{tar}} > 0$ be the DE's target estimation-error variance. Define the DE's optimal time allocation in the AI-assisted regime $t^{*,A}(\theta_Z)$ implicitly by $\sigma^{2,A}_{\text{post}}(t^{*,A}(\theta_Z),\theta_Z) = \sigma^2_{\text{tar}}$.

The following observation records the intensive-margin implication; when AI quality reduces estimation-error variance at the operating point, the reviewer time required to hit a fixed target also falls.

Observation 1 (Effort Saving). Suppose that (1) $\partial\sigma^{2,A}_{\text{post}}/\partial t^A < 0$ and (2) $\partial\sigma^{2,A}_{\text{post}}/\partial\theta_Z < 0$ evaluated at $(t^A,\theta_Z) = (t^{*,A}(\theta_Z),\theta_Z)$. Then, $\partial t^{*,A}(\theta_Z)/\partial\theta_Z < 0$. Thus, higher AI quality reduces the reviewer time required to achieve a fixed estimation-error target.

Condition (1) in Observation 1 holds whenever $\partial\tau^A/\partial t^A > 0$, which we assume throughout. Condition (2) in Observation 1 is the substantive requirement; it holds when the three forces identified above—direct AI precision improvement, the weight of the AI signal effect, and the time-effect function response—net out so that better AI quality reduces the estimation-error variance at the operating point where the target is just met.

3.3.2. Intuition.

The human-only benchmark requires a fixed per-reviewer time $t^{*,H}$ that does not depend on θ_Z. When the conditions of the observation hold, $t^{*,A}(\theta_Z)$ is strictly decreasing in θ_Z. Because $t^{*,A}$ is continuous and decreasing, whereas $t^{*,H}$ is constant, there exists a threshold AI quality $\hat{\theta}_Z$ satisfying $t^{*,A}(\hat{\theta}_Z) = t^{*,H}$ (provided that AI quality is high enough that $t^{*,A}$ eventually falls below $t^{*,H}$). For all $\theta_Z > \hat{\theta}_Z$, the AI-assisted regime achieves the same estimation-error target with strictly less reviewer time per manuscript. This threshold represents the minimum AI quality at which the journal begins to realize reviewer time savings from AI integration on the intensive margin. If embedded in a queueing layer, these time savings translate into higher reviewer throughput, holding journal quality fixed. We have modeled reviewer time, but we have not modeled DE time, author response

time, or the number of rounds of review. The AI review and author response step may add elapsed calendar time before human review begins; our model captures reviewer time requirements rather than total calendar latency, so this front-end latency must be evaluated empirically.

3.3.3. Reviewer Substitution Threshold.

Let the human-only benchmark consist of $m^H = 3$ human reviewers with per-reviewer time $t^{*,H}$. Now, consider the AI-assisted regime with $m^A = 2$ human reviewers evaluated at the benchmark reviewer time $t^{*,H}$, and define $G(\theta_Z) := \sigma^{2,A}_{\text{post}}|_{m^A=2}(t^{*,H}, \theta_Z)$.

Proposition 2 (Substitution Threshold). *Assume that $G(\theta_Z)$ is continuous and strictly decreasing on an interval $[\underline{\theta}_Z, \bar{\theta}_Z]$ and that $G(\underline{\theta}_Z) > \sigma^2_{\text{tar}}$ and $G(\bar{\theta}_Z) < \sigma^2_{\text{tar}}$. Then, there exists a unique critical AI quality $\theta^{\text{crit}}_Z \in (\underline{\theta}_Z, \bar{\theta}_Z)$ such that $G(\theta^{\text{crit}}_Z) = \sigma^2_{\text{tar}}$. For every $\theta_Z > \theta^{\text{crit}}_Z$, two AI-assisted reviewers strictly exceed the target precision previously delivered by three human reviewers without AI. Under these conditions, AI quality induces a unique substitution threshold.*

The existence of the threshold is conditional on $G(\theta_Z)$ being monotone decreasing over the relevant operating region; the previous comparative-statics discussion shows that this need not hold globally because AI quality affects the system through multiple competing channels. This threshold corresponds to the point at which one human reviewer can be removed without sacrificing accuracy. It provides an operational benchmark for when AI-assisted reviewer substitution is feasible. In the Online Appendix, we present formal derivations in Online Appendix C, an analysis of review time in a queueing layer in Online Appendix B, and an analysis of our model using closed-form functional forms in Online Appendix D.

We close this analysis with the crucial result that the AI-infused workflow can improve decision accuracy *even when the stand-alone AI review signal is noisier than that of a time-constrained human reviewer* (i.e., $(1/\theta_Z) > (1/[\theta_H \tau^H(t^H)])$). This condition is stronger than a condition that AI reviewer quality is below human reviewer quality ($\theta_Z < \theta_H$). The mechanism and intuition behind this result are that the AI review allows the journal to reduce the number of reviewers per paper, allowing each reviewer more time and hence, improving their signal to the DE. Analyzing the effect of this process change in conjunction with other aspects, such as the scoping of AI review and the degree to which human reviewers are influenced by the AI review when doing their own evaluation, yields the condition under which AI-infused review can be beneficial. The feasible region covers cases where $\theta_Z < \theta_H$. We state the result formally in Proposition 3 and provide a numerical example to illustrate the finding.

Proposition 3 (AI-Infused Review Can Be Effective Even When AI Is Weaker Than Human). *Suppose the AI-assisted review regime cuts the number of human reviewers from $m^H = 3$ (each with time t^H) to $m^A = 2$ (each with time t^A) while holding total human reviewer time per manuscript fixed, $m^H t^H = m^A t^A$. Assume that ε^A_1, ε^A_2, and ε^Z are mutually independent. Then, the AI-assisted regime can yield a lower-variance signal to the DE than the human-only benchmark, even when $(1/\theta_Z) > (1/[\theta_H \tau^H(t^H)])$ (i.e., the stand-alone AI review signal is noisier than that of a time-constrained human reviewer and hence, also larger than the unconstrained human; that is, $(1/\theta_Z) > (1/\theta_H)$).*

Proposition 3 has been framed above as an *accuracy* result—lower estimation-error variance at fixed total reviewer time; however, the journal can instead convert some or all of the gain into shorter turnaround, fewer reviewer hours per manuscript, or a mix of accuracy and speed. We illustrate this with four numerical scenarios. The accuracy-gain framing of the proposition holds $m^H t^H = m^A t^A$ fixed, and this *reallocation*—fewer reviewers, each with more time—delivers accuracy gain only in scenarios 1–3. Scenario 4 below illustrates the alternative in which part of the gain is taken as a reduction in total reviewer time. The essential point is that the AI-assisted regime lets at least one metric (accuracy or turnaround) improve, whereas the other does not worsen.

Example 1. Let $(m^H = 3, m^A = 2, \omega = \frac{1}{5}, k = 1)$ with human reviewer quality $\theta_H = 1$. Let human-only reviewer time $t^H = -\log(0.3) \approx 1.204$. Then, $\tau^H(t^H) = 1 - e^{-t^H} = 0.7$. The individual human review signal under this benchmark time allocation has noise $(1/0.7) = (1/[\theta_H \tau^H(t^H)])$. The estimation-error variance under the three human reviewer benchmark is $\sigma^{2,H}_{\text{post}} \approx 0.4762$.

Keeping total human reviewer time (for each manuscript) unchanged when moving from three human reviewers to two AI-assisted human reviewers, $m^H t^H = m^A t^A$, yields $t^A = (3/2)t^H \approx 1.806$. The stand-alone AI signal has noise $(\sigma^{2,Z}(\theta_Z) = 1/\theta_Z)$, which exceeds the individual human review signal in scenarios 1, 3, and 4 below (in scenario 2, AI noise is worse only than the unconstrained human reviewer). The effect of AI review on human productivity is defined by $\psi(\theta_Z)$ (a value exceeding one represents a productivity gain) and impacts the time effect via $\tau^A(t^A, \theta_Z) = 1 - e^{-t^A \psi(\theta_Z)}$.

All scenarios yield that the AI-assisted estimation-error variance with two human reviewers

$$\sigma^{2,A}_{\text{post}} = \left(\frac{\omega^2}{\theta_Z} + \frac{(1-\omega)^2}{m^A \theta_H \tau^A(t^A, \theta_Z)} \right)$$

is less than $\sigma^{2,H}_{\text{post}} \approx 0.4762$ (Table 2).

Table 2. Illustrative Examples of Effects for AI-Infused Peer Review

Scenario	θ_Z	$\psi(\theta_Z)$	t^A	$m^A t^A$	$1/\theta_Z$	$\sigma^{2,A}_{\text{post}}$
1. (No productivity gain)	$\frac{1}{2}$	1	1.806	3.612	2.000	0.4629
2. (Improved AI quality)	$\frac{3}{4}$	1	1.806	3.612	1.333	0.4363
3. (AI + productivity boost)	0.65	1.279	1.806	3.612	1.538	0.4168
4. (Reallocation: $m^A t^A = 0.9\, m^H t^H$)	0.65	1.279	1.625	3.251	1.538	0.4273

The two-reviewer AI-assisted process produces a lower-variance signal to the DE through a mix of effects. (i) Modest reliance on AI keeps the nondiversifiable AI noise floor small (ω^2/θ_Z, ranging from 0.053 to 0.080 across the four scenarios), (ii) reducing the human panel from three reviewers to two gives reviewers more time per manuscript, and (iii) the AI review plus author response increases the productivity of that time through $\psi(\theta_Z) > 1$ (in scenarios 3 and 4). Scenario 4 illustrates accuracy gain plus a 10% reduction in total reviewer time (shorter turnaround).

3.4. Shadow AI

Our analysis has covered an idealized peer review status quo, in which multiple human reviews produce independent signals and do so without using an AI review tool. As discussed earlier, emerging evidence of *shadow AI* in the review process suggests that reality may be different than the human-only benchmark; with ad hoc use of AI, human reviewers also become increasingly correlated with each other as they mutually (yet privately) rely on AI-assisted review tools. In our model, this corresponds to (1) unknown reviewer weighting of the AI signal and (2) private AI signals that are generated with unknown prompts by the individual reviewer.

We discuss the potential implications of shadow AI for our benchmark (a panel of independent reviewers). Reviewers using shadow AI may be more productive but may also induce a positive correlation among the human reports; reviewers privately relying on similar AI tools may share a common, journal-unobserved error component. This is similar to the common error—the induced correlation ρ_{HH}—that the governed workflow introduces, with one key difference; under shadow AI, the common component is unmanaged and unmeasured, whereas under the governed workflow, it may be observable. The AI-induced correlation that our model stresses is a cost relative to a panel of independent reviewers, but it can potentially be a benefit relative to shadow AI because the journal can observe the common component. A further distinction is the structured author response, which can meaningfully interact with the AI review in a way that shadow AI does not.

Compared with shadow AI, our proposed governed framework has three advantages. First, the AI review is produced by the journal and shown to reviewers. This makes the AI signal $Z(\theta_Z)$ observable, and it enables the journal to estimate, perhaps through experiments, reviewers' dependence on the AI review (ω). Second, the AI signal is generated by the journal, increasing transparency to authors on what signal $Z(\theta_Z)$ the reviewers are using. Third, the governed workflow elicits an author response to the AI review, which enters the time-effect function $\tau^A(t^A, \theta_Z)$. This response can improve reviewer productivity—an important channel that is absent under shadow AI.

3.5. Key Considerations About AI-Infused Peer Review

One concern for an adverse effect from using an AI reviewer (and sharing the AI review with human reviewers) is error correlation between human reviewers and AI output—specifically, reviewer complacency. However, this risk exists to some extent even when reviewers independently use AI tools. Observing the journal-approved AI makes this transparent, with correlation estimable potentially via experiment. To mitigate anchoring, our view is that AI output should be (i) prohibited from accept/reject recommendations and (ii) presented in structured and tunable formats (e.g., flags first with expandable evidence). Humans retain responsibility for all decisions, including contribution, novelty, impact, and ultimately, whether to accept or reject. Additional concerns involving implementation and the need for experimentation are summarized in Table 3.[7]

4. Conclusion

This article emerged from the imperative to address operational stress faced by journals from increased submission volume combined with the potential harms from hidden uncontrolled use of AI in peer review. Rather than responding defensively or resisting AI, the academic community has the opportunity to lead in shaping systems that uphold scholarly values in this new era. This paper offers both a theoretical model and a practical workflow to integrate AI into peer review (laid out in Section 3), not as a substitute for human judgment but as a structured and transparent workflow. We propose employing AI as a first-line reviewer to provide early feedback and request author response prior to human evaluation. Authors receive quick

Table 3. Key Questions for AI-Augmented Peer Review

Question	Considerations	Our view
Should an AI reviewer be available to authors before submission?	Availability of the tool and/or prompt; cost for the journal; potential for authors' gaming behavior	Weak preference for restricting presubmission access
Are AI and human review tasks strictly separated?	Workflow design; correlation between AI and human review; feasibility	Division of tasks is one of emphasis, not exclusion
Will human reviewers not revert to using their own private AI tools?	Quality of AI reviewer	Governed AI workflow reduces the incentive for shadow AI and makes the journal's own AI use transparent and measurable
Could reviewers become complacent or overreliant on AI assessment?	Conformity with institutionally legitimated algorithms and its impact; accountability concerns; reputational consequences	Even if complacency arises, it is observable; need for experimentation (e.g., through randomized holdout designs)
Might reviewers overcorrect by shifting their standards upward in response to AI flags?	False negatives; quicker path to rejection for weaker papers	Need for experimentation: compare acceptance rates and rejection reasons across AI-exposed and holdout conditions
Will authors not game the AI reviewer by strategically preoptimizing their submissions?	Aesthetic vs. substantive improvements	Strategic compliance with AI review criteria alone cannot guarantee acceptance
Does the author response to the AI review create an excessive burden without commensurate quality gains?	Generic vs. useful AI feedback; response not a revision	Track revision cycles and author time to resubmit to ensure that the process remains efficient
Could AI homogenize reviews, screening out novel or risky contributions in favor of safe, incremental work?	AI's homogenizing tendencies; assessment of novelty and breakthrough potential	Need for experimentation: test whether AI affects reviewers' judgment and recommendations of papers that appear novel
Does the proposed workflow benefit only marginal-quality submissions, or does it help across the quality spectrum?	Accept/reject threshold; quality levels	Benefits span the full-quality distribution but operate through different mechanisms

Note. An extended discussion is in the appendix.

feedback from an audited, transparent, and journal-approved process, with an opportunity to clarify and explain. The AI review combined with author-provided explanation can provide greater context and focus to reviewers. A governed AI process can serve accountability requirements (human reviewers make final decisions, consistent with the European Union AI Act mandates and similar regulation), maintain confidentiality through secure infrastructure with reviewer-level access controls, and enable measurement of the process. Our framework preserves reviewer agency, supports editorial judgment, and converts unobserved processes—such as reviewers privately using their own AI tools and prompts—into managed operational levers under editorial oversight.

Our specific proposal for countering the AI-induced operational challenge is a call to action that aims to prepare journals for the realities of AI-augmented scholarship rather than a finalization in each aspect. Indeed, the final workflow will be a function of various journal-specific discussions and iterations with stakeholders. It has both merits and limitations, underlining the need for careful experimentation. There are many remaining questions and concerns for future assessment, including those highlighted in Table 3 and the appendix. We hope that this proposal establishes a foundation for the management sciences scholarly community to build upon and improve.

Acknowledgments

The authors received valuable feedback from seminar participants at the Indian School of Business, Santa Clara University, the University of Houston, the University of Illinois Urbana-Champaign, the University of Utah, and Washington University in St. Louis. The authors also thank participants at the Biz AI Workshop at the University of Texas at Dallas; the Columbia Conference on Artificial Intelligence, Machine Learning, and Digital Analytics; the INFORMS Annual Meeting; the Information and Operations Management Workshop at the University of Wisconsin–Madison; and the Winter Operations Workshop at the University of Utah. Additionally, the authors thank Zhongju (John) Zhang for being an excellent collaborator as the authors developed the idea and Nitin Bakshi and Brian Han for feedback that helped take this work to the next level. The authors are grateful to two reviewers and the department editor for their excellent comments.

Appendix. Extended Discussion of Key Questions for AI-Augmented Peer Review

Question 1. Should the journal's AI reviewer be made publicly available to authors before submission?

When discussing the availability of the AI reviewer, it is important to distinguish between the tool itself and the prompt as the trade-offs associated with availability of the two differ.

- Case for public access to the tool and the prompt. Making the AI reviewer and its prompt accessible presubmission could enhance transparency and reproducibility. Authors could use the tool to self-improve manuscripts before formal submission, raising average submission quality.
- Case for private access to the tool and the prompt. Public access to the tool creates an incentive for authors to iteratively optimize submissions, resulting in a potentially very costly investment on the journal's part. A public prompt could reveal proprietary evaluation criteria, enabling gaming across submitted papers in the form of strategic compliance without substantive improvement (see also Question 6 for further comments on this issue).
- Our view. Our analysis incorporates a weak preference for restricting presubmission access to the tool and the prompt; the version eventually submitted to the journal is frozen, and postsubmission authors can only provide an explanatory response to issues raised by the journal's AI reviewer. That said, this is ultimately a journal-specific design choice with trade-offs between openness and gaming risk (see Question 6). Journals that do grant presubmission access could mitigate gaming by varying prompts and criteria of the AI review over time.

Question 2. Are AI and human review tasks strictly separated?

- Case for strict separation. A clean division—with AI evaluating dimensions X (such as writing, clarity, and methodology) and humans evaluating dimensions Y (such as novelty and contribution)—would simplify workflow design and could reduce correlation between AI and human signals, strengthening the informational gain from adding AI.
- Case for flexible overlap. Strict separation is not practical because many review dimensions are intertwined (e.g., clarity affects assessment of contribution), and prohibiting humans from evaluating AI-covered dimensions would artificially constrain expert judgment. Furthermore, as AI capabilities evolve, a rigid partition would require constant revision of task division.
- Our view. The division is one of emphasis, not exclusion. AI should handle a vetted subset of checks (e.g., methodological completeness, clarity, and reporting standards—which should be journal-specific design choices based on experimentation) (see Section 3.5), letting human reviewers focus on checks where human judgment is essential, such as contribution, novelty, and impact. This allocation should evolve dynamically with the capabilities and limitations of new generations of models. AI can expedite human reviewers' work even when overlap between AI and human assessments is high; when authors acknowledge and address AI-raised concerns in the response memo, reviewers can move through those points faster, reducing turnaround time even if the informational precision gain is modest.

Question 3. Will human reviewers not revert to using their own private AI tools regardless?

- Case that private AI use will persist. Reviewers may find it convenient to use their own AI tools for tasks outside the journal's criteria for AI review or simply out of habit. Busy reviewers may also turn to private AI tools to address dimensions that the journal-provided AI does not cover. There is no way to fully prevent this behavior.
- Case that governed AI reduces the incentive. A well-designed, journal-specific AI reviewer substantially reduces the marginal value of external tools. If the journal's AI has already performed a thorough structured assessment on the dimensions that it is trusted to evaluate, there is little for an individual reviewer's generic LLM to add on those dimensions.
- Our view. This concern motivates efforts to inject AI into the peer review process—but under a supervised controlled manner. Under the status quo, reviewers already use private AI tools in an uncoordinated, unobservable manner—the "shadow AI" problem (Section 3.4). Although we cannot completely eliminate shadow AI, a governed AI workflow reduces the incentive for it and crucially, makes the journal's own AI use transparent and measurable. Reviewers still need to signal review quality to the DE through meaningful human input where it matters most as reflected in the depth of their judgments on dimensions such as novelty and contribution. The best that the journal can do is decrease the variance that comes from shadow AI; our framework aims to achieve this by making one AI signal explicit and auditable.

Question 4. Could reviewers become complacent or overreliant on the AI assessment?

- Case that complacency is a real risk. If reviewers see an AI report covering certain dimensions, they may assume that those dimensions are "handled" and reduce their own scrutiny—particularly under time pressure. Recent evidence on algorithmic conformity reinforces this concern. Liel and Zalmanson (2025) demonstrate that humans frequently adopt erroneous AI recommendations, even on tasks that they can perform perfectly without support. Critically, they identify *normative pressure*—discomfort at disagreeing with an institutionally legitimated algorithm—as a distinct mechanism beyond mere confidence in AI or a desire to reduce effort. A journal-provided AI review carries precisely this kind of institutional legitimacy, potentially amplifying conformity effects. This could degrade overall review quality even as AI improves, and it may also contribute to the homogenization of reviews (see Question 8).
- Case that AI increases accountability. Reviewers already vary in diligence under the status quo. A structured AI report may actually increase accountability by creating a documented baseline against which reviewer effort can be compared. Reviewers who neglect dimensions covered by the AI can be identified through holdout comparisons. Moreover, Liel and Zalmanson (2025) find that algorithmic conformity decreases when participants perceive their decisions' real-life impact as high, and publication decisions are among the highest-stakes judgments in academic life, which may partially protect against complacency. Reputational consequences for reviewers further reinforce independent judgment.
- Our view. We do not know the answer to this question with certainty; it requires empirical investigation—especially for long-term "learning to get lazy" effects (versus the narrow short-term effects measured in recent papers). The workflow addresses the risk through design: constrain AI to auditable dimensions, provide explicit reviewer guidance emphasizing dimensions outside the criteria of AI review, and use randomized holdout designs (where some reviewers do not see

the AI report) to measure whether anchoring occurs. The key insight from our model is that even if some complacency arises, it may be observable under our framework (reviewer error correlation may be estimable), whereas under shadow AI, it is not.

Question 5. Might reviewers overcorrect by shifting their standards upward in response to AI flags?

- Case that overcorrection could occur. If AI consistently flags certain issues (e.g., missing robustness checks), reviewers may internalize higher thresholds on those dimensions—potentially increasing false negatives and rejecting otherwise publishable work.
- Case that higher standards reflect better-informed reviewing. Higher standards on flagged dimensions may simply reflect more informed reviewing. If AI helps reviewers notice genuine issues that they would otherwise miss, the "correction" is appropriate, not excessive. Moreover, surfacing problems earlier in the process leads to faster, more informed editorial decisions—particularly for papers that would ultimately be rejected anyway. A quicker path to rejection benefits both the journal (reduced reviewer burden) and the authors (who can revise and resubmit elsewhere sooner).
- Our view. The distinction between appropriate recalibration and harmful overcorrection is empirical. Journals should monitor decision shifts under AI exposure by comparing acceptance rates and rejection reasons across AI-exposed and holdout conditions. Tracking downstream signals—citation rates and replication outcomes—can help detect whether standards are drifting beyond the journal's intended bar. This is precisely the kind of operational learning that a governed AI framework, such as the one proposed in this work, enables through its measurement infrastructure (Section 3.5).

Question 6. Will authors not game the AI reviewer by strategically preoptimizing their submissions?

- Case that gaming is a serious risk. If authors anticipate the criteria of the AI review, they can preoptimize for compliance—surface-level fixes that satisfy automated checks without substantive improvement. Baumann et al. (2026) show that prompting an LLM to rewrite a paper can raise AI review scores through style rather than scientific substance. This is especially concerning if the AI reviewer is made publicly available (see Question 1).
- Case that gaming is bounded or even beneficial. If "gaming" means that authors genuinely improve clarity, completeness of reporting, and methodological rigor to satisfy the criteria of the AI review, this is a feature, not a bug. The AI review targets exactly the improvements that one would want.
- Our view. Gaming is a real concern, but its severity depends on implementation. Because the AI review is nonfinal and because human reviewers make the ultimate judgment, strategic compliance with the AI review criteria alone cannot guarantee acceptance. Mitigations include limiting response length, varying prompts and criteria over time, and monitoring abnormal divergence between AI comments and eventual human reviewer assessments. The deeper question—whether authors and reviewers will change behavior strategically in response—is inherently an empirical one that must be investigated through pilot implementations.

Question 7. Does the author response to the AI review create an excessive burden without commensurate quality gains?

- Case that the burden may be excessive. Requiring authors to respond to AI-generated feedback adds a new step to an already demanding submission process. If the AI feedback is generic or addresses minor issues, the response effort may not translate into meaningful manuscript improvement.
- Case that early feedback saves time overall. The response document is intended as a lightweight, structured reply—not a full revision. It creates a correction loop that can catch clear issues early, potentially saving months of review time. Authors already write response documents to human reviewers; this simply shifts some of that effort earlier. Moreover, because this step will occur within a short time window after submission, it will be less taxing (and possibly even exciting) for authors to provide clarifications on their submitted version.
- Our view. Journals should track revision cycles and author time to resubmit to ensure that the process remains efficient. If AI-identified issues do not predict editorial outcomes, the prompt should be adjusted accordingly. In cases where the AI's developmental feedback has limited treatment effect on manuscript quality, the AI review can still serve a valuable triage and screening function—enabling faster desk rejections for clearly inappropriate submissions and freeing human reviewer effort for papers that merit full evaluation.

Question 8. Could AI homogenize reviews, screening out novel or risky contributions in favor of safe, incremental work?

- Case that homogenization is a risk. LLMs can exhibit homogenizing tendencies (Doshi and Hauser 2024, Kumar et al. 2025, Baumann et al. 2026). Without deliberate design choices, AI systems may disproportionately flag unconventional work—unusual methodologies, provocative framing, and interdisciplinary approaches—while rewarding formulaic contributions that fit standard templates.
- Case that homogenization is less of a risk. This risk may be more limited than it first appears. AI reviewers are used as supporting tools rather than as decision makers. Their role is to provide feedback in a more consistent and systematic way, whereas the assessment of novelty, risk, and breakthrough potential continues to be the job of the human reviewer who receives the AI feedback.
- Our view. Homogenization is best treated as an empirical, journal-specific question that depends critically on the design of the reviewing process, especially the structure of human-AI collaboration (Boussioux et al. 2024). This is precisely why a randomized experiment is valuable. It allows a journal to test whether AI affects reviewers' judgment and recommendations of papers that appear novel. We expect that if used carefully, AI review may create room for human reviewers to recognize breakthrough contributions.

Question 9. Does the proposed workflow benefit only marginal-quality submissions, or does it help across the quality spectrum?

- Case that benefits are concentrated at the margins. The model's accuracy gains are most pronounced near the accept/reject threshold, where the additional AI signal sharpens the editorial decision. For very-high-quality or very-low-quality papers, the decision is often clear regardless of AI input.
- Case that benefits span the full distribution. Benefits extend across the full distribution, although through different channels. Low-quality papers benefit from faster, more

informative desk rejections. High-quality papers benefit from reduced turnaround time as reviewer burden decreases.

- Our view. The benefits span the full-quality distribution but operate through different mechanisms. For low-quality or inappropriate submissions—and a substantial fraction of submissions fall in this category—AI screening enables faster feedback, reducing wasted reviewer time. In the model, midrange papers see the largest accuracy gains. For high-quality papers, the primary benefit is faster turnaround; reduced reviewer burden translates directly into shorter response times.

Endnotes

[1] See https://ape.socialcatalystlab.org/.

[2] Although some journals involve an associate editor, we omit this for simplicity.

[3] We recognize that there are alternative ways to define the "scope" of AI review (including what tasks are delegated to the AI reviewer, whether all human reviewers get the AI review or only a subset do, and how AI and human review are sequenced). There is a need for both careful design and experimentation to evaluate each variation and choose between them. For instance, withholding the AI review from one of the m reviewers removes that reviewer's shared ε^Z component, lowering the induced correlation ρ_{HH} and the nondiversifiable error floor, at the cost of forgoing the AI-assisted time-effect gain τ^A for that reviewer—a direct trade-off between reviewer effort/time and correlated error that the present model can represent and that merits dedicated analysis.

[4] Recent evidence from International Conference on Learning Representations (ICLR) 2025 submissions (Sahu et al. 2025) suggests that AI reviewers may approximate human accept/reject decisions while still lagging on certain judgments, such as novelty and theoretical contribution. This performance boundary across different tasks is likely to shift as model capabilities evolve.

[5] This is a modeling abstraction, not a literal description of how reviewers integrate information; the convex combination captures the key feature that exposure to the AI review shifts the reviewer's report toward the AI signal.

[6] In a full-structural model, τ^A would depend on t, θ_Z, and an author response variable X: $\tau^A(t, \theta_Z, X)$.

[7] For an extended discussion, see the appendix.

References

Baumann J, Pei J, Koyejo S, Hovy D (2026) Stop automating peer review without rigorous evaluation. Preprint, submitted May 4, https://arxiv.org/abs/2605.03202.

Berente N, Recker J (2025) Let's all cheer for the *Journal of the Association for Information Systems*. *This IS Research* (March 19), https://www.janrecker.com/this-is-research-podcast/lets-all-cheer-for-the-journal-of-the-association-for-information-systems-19-march-2025/.

Bhargava HK, Brown S, Ghose A, Gupta A, Leidner D, Wu DJ (2025) Exploring generative AI's impact on research: Perspectives from senior scholars in management information systems. *ACM Trans. Management Inform. Systems* 16(2):1–9.

Boussioux L, Lane JN, Zhang M, Jacimovic V, Lakhani KR (2024) The crowdless future? Generative AI and creative problem-solving. *Organ. Sci.* 35(5):1589–1607.

Cambridge University Press (2025) Publishing futures: Working together to deliver radical change in academic publishing. Accessed October 15, 2025, https://coilink.org/20.500.12592/8ht6vqy.

Doshi AR, Hauser OP (2024) Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Sci. Adv.* 10(28):eadn5290.

Gans J (2025) What will AI do to (p)research? Accessed October 15, 2025, https://joshuagans.substack.com/p/what-will-ai-do-to-presearch.

Gartenberg C, Hasan S, Murray A, Pierce L (2026) More versus better: Artificial intelligence, incentives, and the emerging crisis in peer review. *Organ. Sci.* 37(3):795–812.

Gopal RD, Li J, Riemer K, Sarker S, Singh PV, Susarla A, Bichler M, Thatcher JB (2025) Inventing with machines: Generative AI and the evolving landscape of IS research. *Inform. Systems Res.* 36(4):1949–1967.

Gottweis J, Weng W-H, Daryin A, Tu T, Palepu A, Sirkovic P, Myaskovsky A, et al. (2025) Towards an AI co-scientist. Preprint, submitted February 26, https://arxiv.org/abs/2502.18864.

Kumar H, Vincentius J, Jordan E, Anderson A (2025) Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking. *Proc. 2025 CHI Conf. Human Factors in Comput. Systems* (ACM, New York), 1–18.

Kusumegi K, Yang X, Ginsparg P, de Vaan M, Stuart T, Yin Y (2025) Scientific production in the era of large language models. *Science* 390(6779):1240–1243.

Latona GR, Ribeiro MH, Davidson TR, Veselovsky V, West R (2025) The AI review lottery: Widespread AI-assisted peer reviews boost paper scores and acceptance rates. *Proc. ACM Human-Comput. Interaction* 9(7):1–28.

Liang W, Izzo Z, Zhang Y, Lepp H, Cao H, Zhao X, Chen L, et al. (2024a) Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. Salakhutdinov R, Kolter Z, Heller K, Weller A, Oliver N, Scarlett J, Berkenkamp F, eds. *Proc. 41st Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 235 (PMLR, New York), 29575–29620.

Liang W, Zhang Y, Cao H, Wang B, Ding DY, Yang X, Zou J, et al. (2024b) Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI* 1(8):AIoa2400196.

Liel Y, Zalmanson L (2025) Turning off your better judgment: Algorithmic conformity in artificial intelligence-human collaboration. *J. Management Inform. Systems* 42(4):1087–1117.

Ludwig J, Mullainathan S (2024) Machine learning as a tool for hypothesis generation. *Quart. J. Econom.* 139(2):751–827.

Manning BS, Zhu K, Horton JJ (2024) Automated social science: Language models as scientist and subjects. NBER Working Paper 32381, National Bureau of Economic Research, Cambridge, MA,

National Institutes of Health (2025) Supporting fairness and originality in NIH research applications. NIH Notice Number NOT-OD-25-132. Accessed October 15, 2025, https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-132.html.

Peters U, Chin-Yee B (2025) Generalization bias in large language model summarization of scientific research. *Roy. Soc. Open Sci.* 12(4):241776.

Pierce L (2026) The annual report. Accessed April 2, 2026, https://orgsci.substack.com/p/the-annual-report.

Richardson RAK, Hong SS, Byrne JA, Stoeger T, Amaral LAN (2025) The entities enabling scientific fraud at scale are large, resilient, and growing rapidly. *Proc. Natl. Acad. Sci. USA* 122(32):e2420092122.

Sahu G, Larochelle H, Charlin L, Pal C (2025) Reviewertoo: Should AI join the program committee? A look at the future of peer review. Preprint, submitted October 9, https://arxiv.org/abs/2510.08867.

Tao T (2025) Machine-assisted proof. *Notices Amer. Math. Soc.* (January), https://www.ams.org/journals/notices/202501/noti3041/noti3041.html.

Xu Y, Yang LY (2026) Scaling reproducibility: An AI-assisted workflow for large-scale reanalysis. Preprint, submitted February 17, https://arxiv.org/abs/2602.16733.