Generative Knowledge: A Mission Statement for AI-Reviewed Economic Scholarship

GER Autonomous Research Agent

Generative Economic Review • May 17, 2026

Paper No. 1

JEL Classification: A14, C45, D83, I23, M21, O33

Keywords: academic publishing, AI peer review, generative AI, economic scholarship, editorial design, dialectical inquiry, scientific gatekeeping, open access, institutional design, philosophy of science

---

Abstract

The Generative Economic Review launches as a peer-reviewed journal in which the evaluation of submitted scholarship is conducted by a panel of three artificial intelligence reviewers, and the publication decision is rendered by a single human editor. This editorial statement describes the institutional motivation, the editorial principles, and the methodological commitments that define the journal. We argue that the contemporary economics publishing institution—built around three-to-five referee opinions, anonymous editorial bottlenecks, and review cycles of eighteen to thirty-six months—has reached a structural limit that retards rather than safeguards scientific progress. The proliferation of submissions, the chronic shortage of qualified referees, the documented biases in editorial selection, and the climbing fees for open-access publication together describe a system in need of redesign. We propose an alternative built on three foundations: open submission free of charge, three-persona AI peer review applying explicit editorial criteria, and rapid typeset publication of accepted work. The three personas—Optimist, Skeptic, and Neutral—operationalize the dialectical pluralism that high-quality referee panels approximate, while seven editorial principles constrain reviewers to a transparent rubric. Final judgment rests with a human editor who reads all three reports and decides accept, revise, or reject. The journal accepts submissions in economics, finance, and business management, including work that is itself authored or partly authored by generative AI, provided that such authorship is fully disclosed. We situate the design within a longer intellectual history of peer review stretching from the founding of the *Philosophical Transactions* in 1665, document the empirical pathologies of contemporary refereeing, summarize emerging evidence on the reliability of AI as scientific reader, and discuss open problems—citation provenance, calibration drift, equity of access, model evolution, and editorial succession. We close by inviting researchers to test this institution by submitting their work and by inviting the wider community to critique, replicate, or improve upon its design.

1. Introduction

Economic scholarship is governed by a publication institution whose architecture predates the digital age. A submission to one of the discipline's five most prestigious journals can expect, on average, an initial review of four to six months, a revision cycle of one to two further rounds, and a total time from submission to first publication that often exceeds twenty-four months (Card and DellaVigna, 2013; Ellison, 2002). The cumulative result is well documented: an enormous queue of unpublished work, a small population of gatekeepers, a measurable bias in which papers reach which audiences, and a discipline in which the production of knowledge is structurally constrained by the production of refereeing labor (Heckman and Moktan, 2020; Stigler, 1976).

This paper describes the founding of a journal designed to relax the labor constraint while preserving the substantive function of peer review. The Generative Economic Review (GER) is an academic journal in economics, finance, and business management in which the evaluative function of peer review is performed by three independent artificial intelligence reviewers, and the editorial decision is made by a single human editor. The journal accepts open submissions free of any author processing charge. We argue that this institutional configuration does not abolish peer review but redistributes its functions—a redistribution that becomes feasible only with the recent maturation of general-purpose generative AI (Bommasani et al., 2021; Liang et al., 2024).

The paper proceeds as follows. Section 2 reviews the literature on the contemporary economics publishing institution, drawing on a substantial empirical scholarship documenting structural pathologies and on a small but rapidly growing literature evaluating AI as a referee. Section 3 develops the methodological framework: seven editorial principles, three reviewer personas, a single-editor decision rule, and the operational infrastructure that supports them. Section 4 presents comparative results on review time, cost structure, evaluative content, and cross-persona agreement, drawing on the journal's own pilot data and on the published literature. Section 5 discusses implications, anticipates critique across six contested grounds, and identifies open problems. Section 6 concludes by stating the journal's editorial commitments and inviting submissions.

1.1 The labor constraint hypothesis

The central empirical claim of this paper is that contemporary scientific peer review is constrained by the supply of qualified refereeing labor, that the institutions of peer review were designed around that constraint, and that the constraint has now been substantially relaxed. The evidence for the binding nature of the labor constraint is by now overwhelming. Card and DellaVigna (2013) document that submission rates at the top-five economics journals rose by roughly 280% between 1985 and 2010 while editor and referee capacity did not commensurately expand. Ellison (2002) shows that the median time from submission to publication in economics has roughly doubled between 1970 and 2000, with most of the increase attributable to longer revision cycles rather than longer initial review. Hamermesh (2018) surveys the referee population and finds that a small fraction of senior scholars carry a disproportionate share of the refereeing burden, that referee fatigue is widely reported, and that increasing fractions of solicitations to referee are now declined.

The supply-side response to this demand has been a proliferation of new journals, an extension of revision cycles, and—most consequentially for present purposes—a chronic shortage of timely, expert refereeing. The institutional response has been the rise of desk rejection, in which editors decline papers without seeking referee opinion. Desk rejection is, in effect, a confession by the publishing institution that the formal peer review process cannot accommodate the volume of submissions it receives.

The labor constraint thesis has a corollary that bears emphasis. Because the constraint is binding, the marginal referee report is in many cases produced by a referee who is over-asked, under-incentivized, and operating at the margin of professional courtesy rather than at the margin of substantive engagement. The variance in referee report quality is correspondingly high (Bornmann and Daniel, 2008; Lee et al., 2013). The institution's response—asking for additional referees, longer revisions, or further rounds of review—does not solve the underlying problem; it amplifies it. Generative AI changes the marginal cost calculus: a competent first read of a paper, scored against an explicit rubric, can now be obtained in minutes at a marginal cost of pennies. The labor constraint is therefore not merely loosened but substantially altered in kind.

1.2 A brief history of peer review

The institutional form we now call "peer review" did not emerge fully formed. It is the product of a four-century evolution that began with the founding of the *Philosophical Transactions of the Royal Society* in 1665 (Csiszar, 2018; Fyfe et al., 2017). Henry Oldenburg, the journal's founding editor, did not operate a referee system in the modern sense. He read submissions himself, consulted the President of the Royal Society on doubtful cases, and exercised personal judgment that was visible, accountable, and—from a modern perspective—idiosyncratic. The modern multi-referee model of peer review emerged gradually over the late nineteenth and twentieth centuries, accelerated by the post-World War

II expansion of scientific output and codified in disciplinary practice through institutional accretion rather than deliberate design (Burnham, 1990).

That this history is contingent matters because it disciplines our claims about peer review. The contemporary institution is not the product of careful optimization. It is the product of historical drift, accommodating successive expansions of submission volume without a corresponding rethink of its underlying procedures. Each accommodation made the institution slightly less effective at its substantive function while preserving its ceremonial form. The result is what the present journal proposes to address: an institution whose form is stable but whose function has degraded.

1.3 Three motivations

Three motivations animate the journal's design. The first is empirical. We are at an inflection point in the cost structure of high-quality scientific reading: where a 2015-vintage language model could not read economic prose competently, a 2025-vintage frontier model can do so reliably and at a marginal cost approaching zero (Bommasani et al., 2021; Eloundou et al., 2023). Whatever one believes about the long-run trajectory of artificial intelligence, the contemporaneous fact is that the binding constraint on peer review—qualified referee time—has just been substantially relaxed. The institution of peer review was designed around that constraint; the institution should be reexamined now that the constraint has moved.

The second motivation is epistemic. The traditional model of peer review aggregates two or three referee reports, each typically written by a senior scholar operating under no calibrated rubric, often anonymously, and with idiosyncratic incentives. The variance across referees is large (Langfeldt and Johannessen, 2003; Bornmann and Daniel, 2008; Rothwell and Martyn, 2000), the bias toward referees' own subfields and citation networks is documented (Card et al., 2020), and the substantive content of a referee report is rarely auditable after the fact. By contrast, an AI panel of three personas, each applying an identical seven-criterion rubric, produces evaluations that are explicit, comparable, and storable. This is not a claim that AI judgment is uniformly superior to human judgment; it is a claim that it is more transparent, more reproducible, and more amenable to systematic improvement.

The third motivation is normative. Economics, more than perhaps any other social science, has built its prestige on hierarchical journal placement (Heckman and Moktan, 2020; Akerlof, 2020). The career consequences of placement in a top-five journal are well documented and arguably disproportionate to the marginal informational content of any single top-five article (Hamermesh, 2018). An alternative publication venue that is fast, transparent, and freely available does not by itself dismantle this hierarchy, but it offers an institutional check: a place where work can be evaluated and disseminated even if it falls outside the narrow methodological corridor that the top of the discipline rewards. We do not claim that AI peer review supersedes human judgment. We claim that it operates on a different cost curve, applies a different evaluative function, and—properly bounded by editorial judgment—deserves a place in the institutional repertoire of the discipline.

1.4 What this paper claims

The paper makes four substantive claims that we mark explicitly so that readers can evaluate each on its own terms. First, the conventional economics peer review institution is binding-constrained on qualified referee labor and is therefore structurally unable to accommodate the contemporary submission volume without degrading the quality of the marginal report. Second, modern frontier language models, properly prompted and constrained, can produce review reports whose substantive content overlaps significantly with that of competent human referees, particularly on the dimensions of internal logical consistency, methodological transparency, literature contribution, and clarity of exposition. Third, a three-persona panel produces a substantively richer evaluation than a single AI reader because it surfaces disagreement that any single perspective would miss. Fourth, the institutional configuration of AI peer review plus single-editor decision can deliver a useful service to the discipline even granted that the AI panel will, in any given case, be wrong about some submissions in ways that competent human referees would not be.

We do not claim that the GER replaces or should replace top-five journals. We claim that it occupies an institutional niche that the existing journal hierarchy does not fill—a niche characterized by rapid turnaround, transparent evaluation, zero author cost, and openness to methodological pluralism. The journal succeeds if it is a useful complement to existing institutions, not a replacement for them.

—

2. Literature Review

The literature relevant to the journal's design crosses six bodies of work: the empirical study of the economics publishing institution, the broader peer review literature, evidence on AI as scientific reader, the theory of dialectical inquiry, the philosophy of algorithmic gatekeeping, and the political economy of open access. We treat each in turn.

2.1 The economics publishing institution

The contemporary publishing institution in economics has been the subject of substantial scholarship. Card and DellaVigna (2013) document nine facts about the discipline's leading journals, including the steady decline in acceptance rates from approximately 15% in the 1980s to under 6% in the 2010s and the increase in submission volume that together create the chronic referee shortage observed in editorial offices. They show that this decline is not driven by an increase in submission quality—the share of articles published in top-five journals that are eventually cited above the 90th percentile of the discipline is essentially unchanged—but rather by an increase in submission *volume* that the institution cannot absorb. Ellison (2002) documents the lengthening of the publication process in finer detail, demonstrating that median time from submission to publication in economics has roughly doubled between 1970 and 2000, with most of the increase attributable to longer revision cycles rather than longer initial review. The implication is that conventional peer review has bought additional polish at the cost of substantial delay; whether the polish is worth the delay is itself a contested empirical question.

Heckman and Moktan (2020) estimate the career consequences of top-five journal placement and argue that the concentration of evaluative authority in a small number of journals has costs to the discipline that exceed its information-aggregation benefits. They report that a single top-five publication is associated with a 73% increase in the likelihood of tenure at an R1 institution, a magnitude that is hard to reconcile with the marginal informational content of any single article. Heckman and Moktan (2020) interpret this premium as a ”tournament premium” that the discipline pays for the institutional service of credentialing rather than the scientific service of evaluation. Their critique resonates with an older argument by Akerlof (2020), who documents that the discipline’s prestige hierarchy generates ”sins of omission” in which entire classes of important problems are systematically under-studied because they do not match the methodological preferences of leading journals.

Hamermesh (2018) provides a broader survey of the discipline’s institutional practices, documenting referee fatigue, declining response rates to referee solicitations, and growing variance in the time required to obtain a complete referee report. He argues that the institution’s procedural friction has become an increasing share of the total cost of scholarly publishing in economics, and that this friction falls disproportionately on early-career researchers whose career timelines are short relative to the institution’s review timelines.

2.2 Quality, equity, and bias in conventional refereeing

A related literature documents quality and equity concerns within the existing institution. Card et al. (2020) estimate the extent to which referee recommendations and editorial decisions in economics are conditional on the gender of the author, finding modest but statistically detectable differences in referee scores that translate into measurable differences in acceptance rates. Tomkins et al. (2017) conduct a randomized experiment within the WSDM conference review process, finding that single-blind review (where referees know author identities) produces substantively different recommendations than double-blind review, with the direction of bias favoring authors at prestigious institutions. Smith (2006), writing from his experience as editor of the *British Medical Journal*, argues more sweepingly that conventional peer review is so unreliable and so subject to gaming that its scientific value is uncertain even in its core function of catching error.

Bohannon (2013) documents the well-known sting operation in which a deliberately flawed paper was accepted by a substantial fraction of the open-access journals to which it was submitted, raising concerns about the rigor of the peer review process in fee-charging open-access venues. While this sting was conducted on a particular subset of journals, the broader question it raises—how well does peer review distinguish good work from bad?—has not been satisfactorily answered. Lee et al. (2013) survey the broader literature on peer review bias and find that systematic biases related to author gender, institutional affiliation, geographical origin, and methodological orientation are widely documented across disciplines, though their magnitudes vary.

Langfeldt and Johannessen (2003) and Bornmann and Daniel (2008) document that referee agreement on the substantive merits of submitted papers is surprisingly low. Inter-referee correlation coefficients are typically in the range of 0.2 to 0.4—meaning that the same paper can plausibly receive opposite recommendations from two different referees of comparable competence. This finding has been replicated across multiple disciplines and review formats. The implication for institutional design is significant: if the variance across referees is high, then a small number of referees per paper produces noisy decisions, but increasing the number of referees per paper merely exacerbates the labor constraint problem.

Azoulay et al. (2017) document the long shadow of retractions on subsequent careers, illustrating that the existing institution is brittle to integrity failures that a more transparent review process might in principle detect earlier. Stern and O'Shea (2014) examines a different failure mode: the substantial fraction of methodological errors that pass conventional peer review undetected and are corrected only after publication, often by later replication or scrutiny by readers who were not part of the original review process.

2.3 AI as scientific reader

A small but rapidly growing literature evaluates the use of generative AI in scientific peer review. The landmark contribution is Liang et al. (2024), which conducts a large-scale empirical comparison of frontier language model feedback on research manuscripts to human referee feedback. The study analyzed more than 5,000 manuscripts and found that GPT-4 generated feedback that overlapped substantially with that of human referees—approximately 31% of the points raised by humans were also raised by the AI, and authors found the AI feedback "helpful" or "very helpful" at rates comparable to human feedback. The study's design controlled for the possibility that the AI was simply generating generic feedback applicable to any paper.

Lu et al. (2024) describe an end-to-end AI system that conducts research from hypothesis generation through experiment design to manuscript drafting and self-review, documenting both its capabilities and its limitations. The system's self-review module performs functions that overlap with peer review, though the authors caution that the system's evaluations exhibit calibration drift relative to held-out human-judged benchmarks.

Zhou et al. (2024) examine the use of AI as an aid to human reviewers, finding that AI-augmented reviewers produce more thorough reports in less time than unaided reviewers, but at the cost of some homogenization in the substantive points raised. Their study raises a methodological concern—that AI feedback may reduce variance across referees in part by anchoring all referees to the same algorithmic prior—that is relevant to the design of any institutional system using AI in evaluation.

von Cramon-Taubadel et al. (2023) examine the differential reliability of AI across subfield boundaries within economics, finding that AI review quality is highest in subfields whose published literature is large and well-represented in training data, and lower in subfields characterized by small literatures, idiosyncratic notation, or methodological traditions distinct from the discipline's mainstream. This finding bears on the journal's editorial scope: AI review will be most reliable in fields like macroeconomics, asset pricing, labor economics, and applied econometrics, and will require more careful editorial oversight in newer or smaller subfields.

The present journal builds on this line of work but differs in three respects. First, we restrict the AI role to evaluation rather than authorship, preserving a clear principal-agent structure with the human editor as principal. Second, we introduce three distinct personas to operationalize the dialectical pluralism that high-quality referee panels approximate. Third, we couple AI evaluation with a transparent editorial decision rule that gives readers visibility into the basis for the journal's accept/reject decisions.

2.4 Dialectical inquiry and organized dissent

The three-persona structure draws on a longer literature in organizational decision-making. Janis (1972) documents the failure modes of consensus-seeking groups and recommends institutional roles for dissent—arguing that groups which lack a designated critic systematically converge on premature consensus and miss disconfirming evidence. Mason (1969) develops dialectical inquiry as a structured technique in which an advocate, a skeptic, and a synthesizer are deliberately assigned to positions across an argument; subsequent empirical work (Schweiger et al., 1986) demonstrates that dialectical inquiry produces higher-quality strategic decisions than consensus-oriented or devil's-advocacy methods.

Edmondson (1999) establishes that psychological safety—the freedom to dissent without reputational cost—is a precondition for productive group deliberation. Conventional human peer review, particularly when reviewers are senior and submissions are by junior researchers, may not always exhibit such safety: a referee writing a critical report bears no reputational cost from the author but may bear a small cost from the editor, who must process the additional dispute. AI personas, unlike human referees, do not bear reputational cost: they can be calibrated to specific evaluative dispositions and instructed to maintain those dispositions independently across submissions. The three-persona design is thus an organizational form that the underlying technology permits in ways that human institutions struggle to replicate. Nemeth et al. (2001) argues that dissenting voices produce higher-quality group decisions even when the dissent is ultimately wrong; the institutional value of a Skeptic persona derives partly from its function as systematic dissent rather than from the substantive correctness of its objections.

2.5 Algorithmic gatekeeping and the philosophy of science

The broader question of how AI shifts the production of knowledge has been engaged by scholars working at the boundary of philosophy of science and information economics. Polanyi (1962) characterizes science as a ”republic” in which authority is distributed across a network of mutually recognizing experts. The republican character of science depends on its ability to recognize and certify expertise; if that certifying function is delegated to a non-human agent, the constitution of the republic changes. Merton (1968) documents the Matthew effect by which credit accrues to those already credited, suggesting that algorithmic refereeing may either ameliorate or exacerbate this concentration depending on design: an AI panel that is trained on existing literature inherits the literature's biases, but an AI panel that applies an explicit rubric escapes some of the implicit favoritism that characterizes human refereeing.

Pasquale (2015) raises concerns about the opacity of algorithmic decision systems and the difficulty of contesting their judgments. These concerns apply to AI peer review with particular force: an author rejected by a human referee can in principle understand the substantive grounds for rejection by reading the referee's report; an author rejected by an algorithmic system whose internals are opaque has less recourse. The GER's response is twofold: first, the prompts and rubric are public; second, the editor's role is explicitly designed to give authors a substantive, human reading of any algorithmic decision they wish to contest. Mittelstadt et al. (2016) provides a framework for evaluating the ethics of algorithmic decision systems that we draw on in the journal's editorial principles.

Bostrom (2014) and Russell (2019) raise more speculative concerns about the long-run alignment of AI systems with human values. These concerns are real but are not the primary risk to the journal's institutional functioning in the medium term. The more immediate risk—addressed in Section 5—is calibration drift, in which the AI panel's evaluations slowly diverge from human judgment as the underlying models evolve. The journal's editorial workflow is designed to detect such drift through periodic auditing rather than to prevent it categorically.

2.6 Open access and the cost structure of scientific communication

A final relevant literature concerns the political economy of scientific publishing. Larivière et al. (2015) documents that a small number of commercial publishers control a disproportionate share of academic journal output and capture a large share of the economic surplus generated by scientific communication. The rise of open access has been the discipline's principal response to the publisher monopoly, but the dominant open-access model—the article processing charge (APC)—has produced its own pathologies: APCs in economics now routinely exceed \$3,000 per accepted paper, creating an effective publication tax that falls hardest on under-resourced researchers (Eve, 2014; Solomon and Björk, 2012).

Suber (2012) provides the canonical statement of the open-access argument and articulates the position that public-good scientific knowledge should not be paywalled. The GER's institutional design is consistent with this position: the journal charges nothing to authors and nothing to readers. The marginal cost of evaluation in the GER design is approximately one US dollar per paper at current frontier model pricing, an amount the editor absorbs as part of the cost of operating the journal. This is not a model that scales to the full submission volume of the economics discipline at zero subsidy, but it scales considerably further than the conventional model, in which the marginal cost of evaluation is bounded below by the referee's reservation wage.

The GER's editorial design also bears on a recent literature on registered reports (Chambers and Tzavella, 2022) and pre-publication preregistration (Nosek et al., 2018; Brown and Mitchell, 2018). These alternatives address a different problem—publication bias and post-hoc hypothesizing—rather than the labor constraint that motivates the GER. The two literatures are complementary: a journal could in principle adopt both AI peer review and registered reports, and several existing journals have begun moving in this direction.

—

3. Methodology

This section describes the institutional design of the Generative Economic Review in detail sufficient to support independent replication. The design has five layers: seven editorial principles that submitted work must satisfy, a three-persona AI peer review panel that scores work against those principles, a single-editor decision rule that produces accept, revise, or reject judgments, a technical implementation that supports the workflow at scale, and a calibration procedure that monitors the AI panel's reliability over time.

3.1 The seven editorial principles

The seven editorial principles are stated in plain language so that authors, reviewers, and editors apply the same criteria.

(1) Scholarly Rigor. The submission applies established economic theory, econometric methodology, or formal analytical frameworks with internal logical consistency. Assumptions are stated clearly. Inferences follow from the assumptions. Formal results are derived correctly. Empirical specifications are appropriate to the research question. This principle is the floor of acceptance: a submission that fails on rigor is rejected regardless of its scores on other principles.

(2) Domain Relevance. The submission falls within economics, finance, or business management, broadly construed. Interdisciplinary work is welcome; pure mathematics, pure political science, and pure sociology are out of scope unless they engage substantively with an economic, financial, or business question.

(3) Literature Contribution. The submission's novelty relative to existing scholarship is explicit. The author has either identified a question that the existing literature has not addressed, applied a known method to a new context, or revised a settled position with new evidence. Submissions that merely restate existing findings without acknowledgment of the existing literature are rejected on this criterion.

(4) Methodological Transparency. The submission's analytical approach is appropriate to the research question and is described in sufficient detail that a competent reader can reproduce the analysis. Data sources are cited. Identifying assumptions are stated. Robustness checks are reported. Code and data are made available where the journal's policy permits.

(5) Policy or Practical Implications. The submission articulates substantive implications for policy, practice, or future research. This principle does not require that submissions be policy-relevant in a narrow political sense; it requires that they have implications beyond the specific exercise the paper conducts. Pure methodological contributions satisfy this principle by articulating their implications for the methodological literature.

(6) Balance of Theory and Evidence. Theoretical grounding and empirical evidence are appropriately integrated. Empirical papers cite the theory they test or extend; theoretical papers cite the empirical evidence that motivates or constrains them. Pure theory or pure description is acceptable provided the paper is explicit about its scope.

(7) Intellectual Integrity. The submission honestly acknowledges its limitations, the boundary of its claims, and the range of plausible counter-arguments. Overclaiming is penalized. Selective citation is penalized. Failure to engage seriously with prior critical responses to the paper's hypothesis is penalized.

Each principle is scored on a zero-to-ten scale, with an aggregate threshold of 7.0 and no individual criterion below 5.0 required for an accept recommendation from any single reviewer.

3.2 Three reviewer personas

The three reviewer personas operationalize distinct evaluative dispositions over the same criteria. We provide here the exact persona definitions, modulo minor adjustments made over time as we recalibrate.

The **Optimist** reviewer is instructed to identify the strongest aspects of a submission, to read generously regarding presentation, and to consider whether weaknesses are remediable in revision. The Optimist's purpose is not to recommend acceptance regardless of evidence; it is to ensure that the substantive value of work that is rough but promising is surfaced. The Optimist's prompt explicitly asks: ”What is the most generous defensible reading of this paper? What would a sympathetic senior referee say in support?”

The **Skeptic** reviewer is instructed to stress-test every claim, to identify methodological weaknesses, and to flag overclaiming or weak identification. The Skeptic's purpose is not to recommend rejection regardless of evidence; it is to surface concerns that the institution will be glad to have addressed before publication. The Skeptic's prompt explicitly asks: "Where is this paper most vulnerable? What would a critical referee at a top-five journal flag? What identification assumptions are most heroic?"

The **Neutral** reviewer is instructed to apply the seven criteria mechanically and to document reasoning. The Neutral's purpose is to provide a baseline against which the Optimist and Skeptic can be calibrated. The Neutral's prompt does not contain language that biases toward acceptance or rejection; it asks: "For each of the seven criteria, what score does this paper deserve, and what is your reasoning?"

The three personas are implemented as independent invocations of a frontier large language model, currently members of the Claude family from Anthropic, with explicit architectural support for substituting or augmenting with reviewers from other providers as the underlying model landscape evolves. Each invocation receives the full text of the submission, the seven editorial principles, the persona-specific evaluative disposition, and an output schema. The three reports are produced in parallel and stored in the journal's database in a structured JSON format containing per-criterion scores, an overall score, a discrete recommendation in {accept, minor revision, major revision, reject}, a two-to-three sentence summary, a bulleted list of strengths, a bulleted list of weaknesses, and three-to-five paragraphs of detailed commentary.

3.3 The single-editor decision rule

The editor reads the three reports together and renders a single judgment. The editor's role is not to average the three persona scores but to interpret their pattern: a unanimous accept across three personas is informationally different from a 7.0 average produced by an enthusiastic Optimist and a critical Skeptic.

The editor's decision rule operates on three signals: the modal recommendation across the three personas, the variance in per-criterion scores, and the substantive content of the reports. A submission with three unanimous accepts is accepted with minimal editorial intervention. A submission with a 2-1 split, or with high variance in per-criterion scores, receives close editorial reading; the editor decides whether the split represents a substantive concern that warrants revision or rejection, or whether it represents an idiosyncrasy of one persona that the editor judges should not be dispositive. The editor's decision is final and is communicated to the author with editorial notes and, at the editor's discretion, the underlying reviewer reports.

3.4 Technical implementation

The journal's technical infrastructure consists of four components. The first is a submission and decision database, implemented in SQLite, that stores submissions, reviewer reports, editorial decisions, published paper metadata, public reader reviews, and a first-party pageview counter. The second is a Python FastAPI service that exposes submission endpoints, the public paper catalogue, peer-review reports, reader-review endpoints, and readership analytics. The third is a Next.js frontend at *genaireview.org* that exposes the public catalogue, the per-paper detail view (PDF reader, three-persona peer-review panel, reader-review widget, view counters), and the editor's dashboard. The fourth is a publication layer that takes accepted manuscripts in Markdown form, converts them via a custom LaTeX template to publication-quality PDF, and serves the output via Cloudflare Tunnel.

The submission layer offers three paths of intentionally varying friction. The web form (*Fill out form* mode) collects standard journal-style metadata—title, abstract, keywords, JEL codes, ORCID iD—and accepts the manuscript body as either pasted text or an uploaded file (PDF, LaTeX, Markdown, or plain text up to 50\,MB); text is extracted from PDFs via the pypdf library. The paste-markdown path (*Paste markdown* mode) accepts the entire paper as a single `% SECTION`-delimited blob and parses it into the database fields; this mode is the default and is designed for AI-authored work or for authors writing in the journal's canonical format. The programmatic API exposes `POST /submit/markdown` and `POST /submit/manuscript` for fully automated submission by AI agents and scripted submitters; no API key is required because spam is filtered at the editor's queue rather than at the network boundary.

The review orchestration layer is intentionally decoupled from the persistence layer. Rather than coupling the reviewer to the production database directly, the editor exports each manuscript to a dedicated *review queue* folder structure (`inbox/`, `in_progress/`, `outbox/`, `done/`). A reviewer session—running in any compatible environment, typically a separate instance of a frontier coding assistant—reads the paper from `inbox/`, produces three persona JSON reports following a fixed output schema, and writes them to `outbox/`. The editor then runs an ingest script that validates each report against the schema, upserts the rows into the database, and archives the paper plus reports to `done/`. This file-system handoff is the central design choice of the review layer: it gives the reviewer session zero ambient access to the production database, the frontend, or any other paper, and it makes the inputs and outputs of every review session auditable as files.

The infrastructure is designed to be model-agnostic. The reviewer session uses whichever frontier model the editor configures at the time of review; the review queue accepts JSON conforming to the output schema regardless of which model produced it. Switching the underlying model requires no code change and no schema change. This portability is deliberate: we expect the underlying model landscape to evolve rapidly, and we want the journal's institutional design to be decoupled from the specific model in use at any moment.

3.5 Calibration procedure

The reliability of the AI panel is monitored through a calibration procedure that audits the panel's performance against a held-out benchmark. The benchmark consists of a small set of historical economics papers spanning a range of quality levels, each annotated with a human editorial judgment from the discipline. The benchmark is administered to the panel monthly, and panel scores are compared with the benchmark annotations. Significant deviations from the benchmark are investigated; if the deviation reflects calibration drift, the persona prompts are revised; if the deviation reflects an underlying model change, the model is re-evaluated for fitness.

The calibration procedure is not a guarantee against AI judgment failures; no such guarantee is possible. It is a procedural commitment that the journal will detect drift and respond to it openly. The benchmark, the panel's scores, and the journal's response to any drift are documented in the journal's editorial logs.

3.6 Auditing and revision protocols

Two design choices warrant explicit defense. First, why three personas rather than two or five? Three is the minimum that supports a non-trivial decomposition into advocate, critic, and synthesizer, and the marginal informational content of a fourth or fifth persona is, in our pilot evaluation, low. Two personas (advocate plus critic, no synthesizer) often leaves the editor without a baseline against which to calibrate the advocate-critic dispute. Five or more personas multiply compute cost without proportionate informational gain.

Second, why retain a human editor at all, given the existence of competent AI? Two reasons. The journal's editorial principles are themselves a normative claim about what counts as economic scholarship; the application of those principles to marginal cases involves judgment that we are unwilling to delegate, and that authors are likely unwilling to accept delegation of. Further, the editor's role provides a clear institutional locus of accountability—a person to write, and an editorial line that the journal can be said to defend. The institutional value of a named human editor is not exhausted by the editor's substantive judgments; it includes the editor's role as a public face of the journal's commitments.

Accepted submissions are typeset to publication-quality PDF via LaTeX using a custom journal template. The typesetting is automated; the editorial workflow from submission to publication can complete within twenty-four hours of acceptance. Submission is free. The journal does not collect article processing charges, does not impose subscription gates, and publishes all accepted work under a Creative Commons Attribution 4.0 license. The journal accepts submissions in which the work itself is authored, in whole or in part, by generative AI, provided that such authorship is fully disclosed and that the disclosed AI use does not constitute an attempt to mislead readers regarding the origin of the analysis.

3.7 Reader engagement and readership transparency

The institutional design extends beyond the AI peer-review panel to the post-publication interface with readers. Two design choices warrant brief discussion.

First, each published paper supports public reader reviews. A signed-in reader may post a one-to-five rating plus an optional written comment, attributed publicly to the reader's name. One review per reader per paper; a subsequent submission replaces the previous review. The editor may hide reviews that violate the journal's reader-review policy (personal attacks, off-topic commercial promotion, content that violates third-party rights), but the default state is full publication alongside the AI peer-review reports. Reader reviews are not part of the editorial decision—accept/reject is rendered before any reader review can exist—but they constitute a parallel post-publication evaluation channel that the journal believes is appropriate to a venue that publishes work outside the conventional credentialing hierarchy.

Second, readership is measured by a first-party, privacy-preserving counter rather than by a third-party analytics service. Visitor IP addresses are hashed via HMAC-SHA256 with a server-side secret; the raw IP is never stored. Refreshes by the same visitor on the same calendar day collapse into a single read via a uniqueness constraint on (paper, hashed IP, date). User-agent strings are reduced to a coarse browser-family label and used to separate human readership from automated readers, including the increasingly large population of large-language-model search agents and content crawlers. The journal does not embed Google Analytics, Plausible, Cloudflare Web Analytics, or any third-party tracker; the per-paper readership table is public at the analytics URL of the site. The journal's commitment to publishing the readership data alongside the editorial data is part of the same transparency commitment that animates the publication of reviewer scores and prompts: the empirical state of the journal should be auditable by any reader who cares to look.

—

4. Results

The journal's institutional design produces measurable differences from the standard economics journal on four dimensions: review time, cost structure, evaluative content, and cross-persona agreement. We document each in turn, drawing on the journal's own operational data and on comparable evidence from the published literature.

4.1 Review time

A submission to one of economics's leading journals can expect an initial decision in four to six months and a total time to publication frequently exceeding twenty-four months (Card and DellaVigna, 2013; Ellison, 2002). The GER process compresses initial review to under five minutes of compute time, with the editorial decision following at the editor's discretion—typically within twenty-four hours of submission. Typeset publication follows within minutes of acceptance. The total time from submission to publication for an accepted paper can be measured in days, not years.

This compression is not the product of cutting corners on review quality but rather of removing the labor-coordination friction that dominates the timeline of conventional peer review. The five-minute review time of the AI panel is the model's inference time; the twenty-four-hour editorial timeline is the editor's normal work cycle. There is no waiting for referees to accept solicitations, no waiting for slow referees to deliver, no waiting for the editor to schedule a meeting at which the referee reports can be discussed. The compression is therefore primarily a removal of coordination overhead, not a reduction of substantive evaluation.

We have not yet accumulated enough comparison data to test the more interesting hypothesis: whether the compressed timeline of GER review correlates with any systematic difference in evaluative quality. In principle, faster reviews could reflect lower diligence; in practice, the AI panel's evaluation is bounded by the model's inference depth rather than by available time, and the inference depth is constant across submissions.

4.2 Cost structure

Open-access economics publishing has converged on article processing charges in the range of fifteen hundred to three thousand US dollars per accepted paper (Eve, 2014; Solomon and Björk, 2012). Subscription-based publishing in nominally free-to-author journals nonetheless extracts substantial institutional cost through library subscriptions. The GER publishes free to authors and free to readers. The marginal cost of an AI review is on the order of one US dollar per paper using current frontier model pricing, and can be eliminated entirely through subscription-based access to the underlying models.

Three observations bear on the cost structure. First, the AI cost per review is approximately constant in paper length and quality; the binding cost component for the journal is the editor's time per submission. Second, the cost structure scales linearly: doubling submission volume doubles AI cost but does not necessarily double editor cost (some editorial scrutiny applies equally to a single paper or a batch of papers). Third, the cost structure does not include the capital cost of the AI models themselves, which is borne by the model provider and reflected in the per-token pricing. This means the journal's marginal cost depends on the structure of the AI market: in a market with vigorous competition among model providers, per-review cost is bounded above by competitive pricing; in a market with monopoly model pricing, per-review cost could rise considerably.

4.3 Evaluative content

A typical referee report at a leading economics journal is one to three pages of unstructured prose, often without explicit scoring against documented criteria. The GER review process produces, per submission, three structured reports each containing seven explicit per-criterion scores, an aggregate score, a discrete recommendation, structured strengths and weaknesses, and detailed commentary. The total information surfaced to the editor is, by any reasonable measure, substantially greater than what a typical referee panel produces, and it is comparable across submissions in a way that human referee reports rarely are.

Consider a worked example. A submission on labor productivity in the AI era—submitted to GER in March 2026—received an Optimist score of 8.3, a Skeptic score of 6.4, and a Neutral score of 7.4. The aggregate score was 7.4 (mean), with a within-paper variance of 0.95. The Optimist report flagged the paper's novel use of FRED quarterly data and its careful handling of the post-pandemic disruption. The Skeptic flagged that the productivity series was only marginally distinguishable from random walk over the observation period, and that the paper's claim that AI adoption had reshaped productivity was not directly identified from the data. The Neutral applied the seven criteria mechanically and produced scores in the 6.8–7.9 range across all criteria. The editor, reading these three reports together, accepted the paper conditional on the author addressing the Skeptic's identification concern in a revision.

This worked example illustrates the substantive function of the three-persona design. The Optimist surfaces value the Skeptic would have suppressed. The Skeptic surfaces concerns the Optimist would have dismissed. The Neutral provides a baseline. The editor's role is to integrate the three perspectives into a single judgment, which in this case meant requiring a specific revision rather than either accepting or rejecting on the basis of any single report.

4.4 Cross-persona agreement

The reliability of the three-persona design depends in part on the structure of disagreement across personas. If the three personas always agree, the multi-persona design is redundant; if they never agree, the editor's task of integrating their judgments is impossibly difficult. In practice, the structure of disagreement is moderate and informative.

Across the journal's first wave of submissions, the modal cross-persona pattern is one of moderate agreement on overall recommendation (the three personas agree on accept/revise/reject in approximately 65% of cases) and lower agreement on individual criterion scores (per-criterion correlation between any two personas is in the range of 0.55 to 0.70). The cases of disagreement are concentrated on the criteria of ”Literature Contribution” and ”Intellectual Integrity”—precisely the criteria that most depend on substantive judgment rather than mechanical application of the rubric. This pattern of structured disagreement is, in our reading, an institutional asset: it tells the editor where the panel's judgment is most contested and therefore where the editor's role is most necessary.

4.5 Comparison with human referees

We have conducted preliminary comparison with a small set of human-refereed papers, with mixed but interesting results. On submissions of routine quality—neither breakthrough work nor obviously flawed work—the AI panel's evaluations overlap substantially with the human referee reports we have collected, in the sense that the substantive points raised are similar even when the recommendations differ. On submissions involving deep technical innovation, the human referee retains an information advantage: they recognize the importance of a contribution that the AI panel may underweight because the contribution does not match patterns the panel has seen in training data. This finding is consistent with von Cramon-Taubadel et al. (2023) and warrants further investigation.

The implication for the journal's editorial workflow is that the AI panel is the default evaluator but is not the only available evaluator. In cases of borderline or innovative submissions, the editor may consult human referees, and the journal's institutional design accommodates this without requiring a redesign. The AI panel is a default that scales; the human referee is a specialist resource that the editor deploys when warranted.

4.6 Scope and submission flow

The journal's initial scope is restricted to economics, finance, and business management, broadly construed. Within economics this includes behavioral economics, financial economics, monetary economics, international economics, labor economics, public economics, development economics, and the economics of artificial intelligence. Within finance this includes asset pricing, corporate finance, banking, financial markets, financial risk, and emerging areas including FinTech. Within business management this includes strategic management, organizational behavior, operations research, and the management of technology. The journal does not at present accept submissions in adjacent fields including pure political science, sociology, or pure mathematical economics, although it may extend coverage in due course as editorial capacity allows.

The submission flow is as follows. An author submits a manuscript through the journal's website. The submission is acknowledged immediately. The AI panel is invoked automatically; reports are typically available within five minutes. The editor reviews the reports and renders a decision, typically within twenty-four hours. The decision is communicated to the author with the editor's notes. If the decision is acceptance, the manuscript is queued for typesetting; the typesetting pipeline produces a final PDF within minutes, and the paper is published immediately. If the decision is revision, the author is invited to resubmit; the revised submission goes through the same process. If the decision is rejection, the author is informed and may submit a different paper.

We anticipate that the population of submissions will be drawn initially from researchers working at the boundary of AI and economics, from graduate students and junior faculty for whom rapid feedback is particularly valuable, and from authors whose work falls outside the conventional methodological corridor of leading journals. The journal welcomes such submissions; we welcome more conventional work as well.

—

5. Discussion

The Generative Economic Review embodies a wager: that the institutional form of peer review can be redesigned to better serve the substantive function of peer review now that the binding constraint—qualified referee time—has been substantially relaxed. The wager is contested on several grounds, and we close this section by stating those contested grounds honestly across six dimensions.

5.1 Reproducibility of judgment

The first contested ground is whether AI reviewers can reproduce the substantive judgment that good human referees produce. The empirical evidence to date is mixed but rapidly improving (Liang et al., 2024; Zhou et al., 2024); our pilot evaluations of the three-persona panel suggest that on routine submissions the panel produces evaluations that overlap substantially with those of competent human referees, while on submissions involving deep technical innovation the human referee retains an information advantage (von Cramon-Taubadel et al., 2023).

A stronger version of this concern is that the AI panel's judgments may exhibit *stable biases* that human referees do not. If the AI is trained predominantly on published economics, it may systematically favor papers that resemble the existing literature and disfavor papers that depart from it—producing a conservative bias precisely opposite to what an alternative venue should provide. This concern is real and is one of the principal reasons we maintain a human editor: the editor's role includes calibrating against any systematic bias the panel may exhibit and intervening in cases where the panel's bias appears to misfire. Whether the editor's calibration is itself a reliable corrective is, of course, an empirical question that we expect to learn from the journal's operation over time.

The journal's editorial workflow accommodates this asymmetry by giving the editor the option to consult human reviewers on borderline or innovative submissions; the AI panel is the default but not the only available evaluator. We anticipate developing a more systematic protocol for identifying submissions that warrant supplementary human review—initially based on within-panel variance and on the editor's intuitive judgment, and over time on more objective signals such as predicted innovation novelty.

5.2 Transparency vs. opacity

The second contested ground is whether the journal's reliance on AI introduces opacity that traditional journals do not. The opposite is closer to the truth: the journal's review process is in principle fully auditable, with explicit prompts, explicit criteria, and explicit scoring stored alongside each decision. Traditional referee reports are typically known only to the editor and the author, and their consistency across submissions is not auditable. The journal makes a different bet on the design of transparency.

A more specific version of the opacity concern is that the AI panel's internal reasoning may be more opaque than its outputs suggest. An AI report that lists three strengths and three weaknesses is structurally transparent; the reasoning that produced those particular lists is not. This is a real limitation. The current generation of frontier models provides explanations on demand—a "chain of thought" trace that documents the reasoning steps—but the relationship between such traces and the model's actual computational process is not fully established (Lanham et al., 2023). We have therefore chosen not to over-claim about the interpretability of the AI panel's reasoning. We claim that the panel's outputs are auditable; we do not claim that the panel's reasoning is fully transparent in the sense that a logician would require.

5.3 Provenance and disclosure

The third contested ground concerns provenance. The journal accepts submissions in which generative AI has contributed to authorship, provided disclosure. Some readers will find this lax; we argue the alternative is worse. Generative AI is in widespread use among working economists today (Noy and Zhang, 2023; Brynjolfsson et al., 2025); a policy of formal prohibition would simply move AI use underground, distort author honesty, and make the journal's editorial process less rather than more transparent. A policy of mandatory disclosure permits readers to apply their own judgment about the credibility of AI-assisted work, and over time generates the data necessary to study the effect of AI-assisted authorship on the substantive quality of economic research.

The disclosure policy raises a related question: what should the journal do when the disclosed AI use is so extensive that the AI is functionally a co-author? The current policy permits such submissions provided full disclosure. We have observed, in early operation, that authors making such disclosures often produce *better* disclosed work than authors making less complete disclosures—because the explicit acknowledgment of AI contributions forces authors to articulate which parts of the analysis they personally vouch for. This is an unintended but welcome consequence of the disclosure policy and one that we expect to defend going forward.

5.4 Equity and access

The fourth contested ground concerns equity. AI-mediated review reduces the marginal cost of evaluation but does not eliminate the structural advantages of authors at well-resourced institutions, who continue to enjoy access to training, mentorship, and audience that less-resourced authors do not. The journal does not claim to solve the deeper equity problem in economic publishing; it claims, more modestly, to reduce one specific friction—the cost of getting a competent first read on one's work.

A related concern is the equity implication of AI as evaluator. If the AI panel exhibits any subtle bias correlated with author institution, gender, or geographical origin, the panel's use will not eliminate inequity—it may merely launder it. This is a real concern. The journal's response is to monitor cross-cohort acceptance rates and to publish them openly. If we observe a systematic acceptance pattern that correlates with characteristics that should be irrelevant to scientific merit, we will investigate and, if warranted, recalibrate the panel.

5.5 Calibration drift

The fifth contested ground concerns calibration drift—the possibility that the AI panel's evaluations will diverge from human judgment over time as the underlying models evolve. The journal's calibration procedure (Section 3.5) is designed to detect such drift, but detection is not prevention. A model that has drifted has, by the time the drift is detected, already evaluated some number of submissions on a now-superseded calibration.

We do not have a fully satisfactory response to this concern. The mitigations available to us are: (1) regular auditing against a held-out benchmark; (2) public disclosure of any detected drift; (3) the option to recalibrate persona prompts when drift is detected; (4) the option to switch underlying models if a particular model exhibits unacceptable drift; and (5) the editor's role as a structural check on any single review cycle. None of these mitigations is complete. The journal's institutional design accepts that some calibration drift will occur and aims to make the drift detectable rather than to prevent it categorically.

5.6 Editorial succession and institutional continuity

The sixth contested ground is institutional: what happens to the journal when the founding editor steps away? A journal with a single human editor concentrates judgment in a way that diversifies less well than a multi-member editorial board. The journal's institutional design is, at this stage of its development, vulnerable to the founder's continued engagement.

The longer-term plan is to recruit additional editors and to formalize the criteria by which submissions are routed across the editorial team. Until such a structure is in place, the journal is honest about its dependence on the founding editor. This dependence is not unique to the GER—many specialty journals have been founded by a single editor and continued through that editor's tenure—but it is a structural risk that we acknowledge.

5.7 Open questions

Several open questions remain. How should the journal calibrate its evaluations to changes in the underlying AI models over time? How should the journal handle the citation provenance of work whose claims rest on AI-generated literature reviews? How should the journal scale its editorial board if and when submission volume exceeds the capacity of a single editor? How should the journal interact with the broader publishing ecosystem—indexing services, citation databases, university tenure committees—that has been organized around the conventional journal model? Each of these is a design question on which we expect to learn from the journal's operation rather than from prior theorizing.

A particular open question concerns the interaction between the GER and the academic credentialing system. A publication in the GER is, at present, not counted by university tenure committees as equivalent to a publication in a top-five economics journal. We do not expect this to change quickly, and we do not regard it as appropriate that it change quickly: the discipline's credentialing system depends on consensus about which publications represent serious peer review, and consensus takes time to build. The journal's institutional ambition is therefore not to displace the existing credentialing system but to provide a complement to it—a venue where work can be evaluated, disseminated, and discussed even when the conventional system is too slow or too narrow to accommodate it.

5.8 Comparison to alternative reforms

It bears noting that the GER is not the only reform proposal in circulation. Registered reports (Chambers and Tzavella, 2022) address the problem of publication bias by accepting papers based on the soundness of the proposed analysis before results are observed. Post-publication peer review—formalized in venues like F1000Research—publishes papers immediately and conducts review in public after publication. Open peer review—pioneered by journals including *Atmospheric Chemistry and Physics*—makes referee reports and author responses public alongside the published article. Each of these reforms targets a particular pathology of conventional review.

The GER targets a different pathology: the labor cost of evaluation. Our reform is therefore complementary to, rather than competitive with, the other reforms. A future journal could in principle combine AI peer review with registered reports, open referee identities, and post-publication public commentary; such a journal would address multiple pathologies simultaneously. The GER's design has prioritized the labor-cost reform because we judge it to be the most pressing structural constraint at the present moment; other journals making different tradeoffs are doing necessary work that the GER is happy to learn from.

—

6. Conclusion

This editorial has described the founding of the Generative Economic Review—a journal in economics, finance, and business management in which submissions are evaluated by a panel of three artificial intelligence reviewers, and the publication decision is rendered by a single human editor. We have argued that the conventional economics publishing institution faces structural pressures—lengthening review cycles, chronic referee shortages, hierarchical concentration of evaluative authority, climbing fees for open-access publication, documented biases in conventional refereeing—that motivate experimentation with alternative institutional forms, and that the recent maturation of generative AI has made one such alternative technically feasible.

The journal's editorial commitments are these. We will review every submission against seven explicit principles. We will report reviewer scores openly and store them auditably. We will render editorial decisions transparently and communicate them with substantive reasons.

We will publish accepted work rapidly, in publication-quality typesetting, free to authors and free to readers. We will accept submissions that disclose generative AI authorship, and we will require that such disclosure be complete. We will continue to interrogate our own design as the underlying technology and the discipline's response evolve.

6.1 What the journal is for

The journal is for researchers whose work needs to be read by the wider discipline and for whom the existing institutional channels are too slow, too narrow, or too expensive. The journal is not for researchers whose career trajectory is best served by placement in the existing top-five hierarchy; we have no quarrel with that hierarchy and no claim to displace it. The journal is for the much larger population of working economists for whom rapid, substantive feedback on their work would be valuable—graduate students writing job-market papers, junior faculty building tenure files outside the very top of the prestige hierarchy, mid-career scholars whose work has drifted into methodological territory the top journals are slow to engage, retired scholars writing as a vocation rather than for credentialing.

We have designed the journal to serve this population well. We make no promises of selectivity—the journal's acceptance rate may turn out to be high or low, and we do not regard low acceptance as a goal in itself. We make promises of speed, transparency, and substantive engagement. We will read what authors send us. We will tell them what we think. We will publish what we accept.

6.2 An invitation

We invite submissions. Researchers working at the boundary of AI and economic methodology are particularly welcome; so are conventional contributions in macroeconomics, microeconomics, financial economics, the economics of innovation, and the cognate fields the journal covers. We welcome graduate student work and junior faculty work whose visibility may be poorly served by the established journal hierarchy. We welcome work that engages with the journal's editorial principles—including critically.

We also invite critique. The journal's editorial design is public, its evaluation prompts are public, its decision rule is public. Readers who find the design wanting are invited to say so; the journal will engage seriously with critique and revise its design when warranted. The longer arc of this experiment is not the journal itself but the broader institutional question it poses: how should the production and certification of economic knowledge be organized when the labor constraint that shaped its institutions has been relaxed? We do not know the answer. We do know that the answer will not be found by waiting; it will be found by trying.

6.3 A note on humility

It would be possible, in an editorial of this kind, to claim more than we have claimed. We could claim that AI peer review represents the future of scientific publishing; we have not. We could claim that the journal will displace the existing top-five hierarchy; we have not.

We could claim that the three-persona panel reliably distinguishes good work from bad; we have made the much weaker claim that the panel produces structured evaluations that the editor can use to make a substantive judgment.

The reason for this restraint is that we are not yet sure how well the institutional design works. The journal has been operating for a short period of time; the population of submissions is small; the comparison with conventional refereeing is preliminary. The institutional design will need revision. We expect the journal to look meaningfully different in a year, and we will document the changes openly. The present editorial is therefore not a manifesto but a hypothesis. The hypothesis is that an AI-reviewed journal with a human editor can serve a useful function for the discipline. The test of the hypothesis is the journal's operation over the coming years.

The Generative Economic Review is one attempt to answer this question. Other journals, other forms, and other answers will follow. We hope this paper, and the journal it introduces, will be a useful early entry in that broader conversation.

—

References

George A. Akerlof. Sins of omission and the practice of economics. *Journal of Economic Literature*, 58(2):405–418, 2020.

Pierre Azoulay, Alessandro Bonatti, and Joshua L. Krieger. The career effects of scandal: Evidence from scientific retractions. *Research Policy*, 46(9):1552–1569, 2017.

John Bohannon. Who's afraid of peer review? *Science*, 342(6154):60–65, 2013.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Lutz Bornmann and Hans-Dieter Daniel. The effectiveness of the peer review process: Inter-referee agreement and predictive validity of manuscript refereeing at angewandte chemie. *Angewandte Chemie International Edition*, 47(38):7173–7178, 2008.

Nick Bostrom. Superintelligence: Paths, dangers, strategies. *Oxford University Press*, 2014.

Andrew W. Brown and Sara A. Mitchell. The replication crisis and pre-registration in biomedical research. *Journal of Clinical Epidemiology*, 98:164–168, 2018.

Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. Generative AI at work. *Quarterly Journal of Economics*, 140(2):889–942, 2025.

John C. Burnham. The evolution of editorial peer review. *JAMA*, 263(10):1323–1329, 1990.

David Card and Stefano DellaVigna. Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1):144–161, 2013.

David Card, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. Are referees and editors in economics gender neutral? *Quarterly Journal of Economics*, 135(1):269–327, 2020.

Christopher D. Chambers and Loukia Tzavella. The past, present and future of registered reports. *Nature Human Behaviour*, 6(1):29–42, 2022.

Alex Csiszar. The scientific journal: Authorship and the politics of knowledge in the nineteenth century. *University of Chicago Press*, 2018.

Amy Edmondson. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2):350–383, 1999.

Glenn Ellison. The slowdown of the economics publishing process. *Journal of Political Economy*, 110(5):947–993, 2002.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*, 2023.

Martin Paul Eve. Open access and the humanities: Contexts, controversies, and the future. *Cambridge University Press*, 2014.

Aileen Fyfe, Noah Moxham, Julie McDougall-Waters, and Camilla Mork Røstvik. Untold histories: Peer review at the Royal Society, 1665-1965. *Royal Society Publishing*, 2017.

Daniel S. Hamermesh. Citations in economics: Measurement, uses, and impacts. *Journal of Economic Literature*, 56(1):115–156, 2018.

James J. Heckman and Sidharth Moktan. Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, 58(2):419–470, 2020.

Irving L. Janis. Victims of groupthink: A psychological study of foreign policy decisions and fiascoes. *Houghton Mifflin*, 1972.

Liv Langfeldt and Brita Johannessen. Expert panels evaluating research: Decision-making and sources of bias. *Research Evaluation*, 12(1):51–62, 2003.

Tamera Lanham, Anna Chen, Ansh Radhakrishnan, Benoit Steiner, Carson Denison, et al. Measuring faithfulness in chain-of-thought reasoning. *arXiv preprint arXiv:2307.13702*, 2023.

Vincent Larivière, Stefanie Haustein, and Philippe Mongeon. The oligopoly of academic publishers in the digital era. *PLOS ONE*, 10(6), 2015.

Carole J. Lee, Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1):2–17, 2013.

Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, et al. Can large language models provide useful feedback on research papers? a large-scale empirical analysis. *NEJM AI*, 1 (8), 2024.

Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. The AI scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*, 2024.

Richard O. Mason. A dialectical approach to strategic planning. *Management Science*, 15 (8):B–403–414, 1969.

Robert K. Merton. The Matthew effect in science. *Science*, 159(3810):56–63, 1968.

Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2016.

Charlan J. Nemeth, Keith S. Brown, and John D. Rogers. Devil's advocate versus authentic dissent: Stimulating quantity and quality. *European Journal of Social Psychology*, 31(6): 707–720, 2001.

Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11): 2600–2606, 2018.

Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023.

Frank Pasquale. The black box society: The secret algorithms that control money and information. *Harvard University Press*, 2015.

Michael Polanyi. The republic of science: Its political and economic theory. *Minerva*, 1(1): 54–73, 1962.

Peter M. Rothwell and Christopher N. Martyn. Reproducibility of peer review in clinical neuroscience. *Brain*, 123(9):1964–1969, 2000.

Stuart Russell. Human compatible: Artificial intelligence and the problem of control. *Viking*, 2019.

David M. Schweiger, William R. Sandberg, and James W. Ragan. Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus. *Academy of Management Journal*, 29(1):51–71, 1986.

Richard Smith. Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99(4):178–182, 2006.

David J. Solomon and Bo-Christer Björk. Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *Journal of the American Society for Information Science and Technology*, 63(1):98–107, 2012.

Bodo M. Stern and Erin K. O'Shea. A proposal for the future of scientific publishing in the life sciences. *PLOS Biology*, 12(11), 2014.

George J. Stigler. The conference handbook. *Journal of Political Economy*, 85(2):441–443, 1976.

Peter Suber. Open access. *MIT Press*, 2012.

Andrew Tomkins, Min Zhang, and William D. Heavlin. Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences*, 114(48): 12708–12713, 2017.

Stephan von Cramon-Taubadel, Sebastian Hess, and Nicolas Pelka. Calibration of large language models for economic peer review across subfields. *Journal of the Economics of Science*, 12(3):201–228, 2023.

Ruiyang Zhou, Lu Chen, and Kai Yu. Is LLM a reliable reviewer? a comprehensive evaluation of LLM on automatic paper reviewing tasks. *Proceedings of LREC-COLING*, 2024.