Bus Inf Syst Eng
https://doi.org/10.1007/s12599-026-01013-0

EDITORIAL

Academic Publishing in the Age of Generative AI

Björn Hanneke · Frederik Hering · Jella Pfeiffer · Hajo Reijers · Oliver Hinz

1 Introduction

Academic peer-reviewed journals turn a growing universe of ideas into credible scientific signals (Yoo 2024), based on shared norms about what counts as a contribution. This task rests on three limited resources: editorial attention, reviewer time, and audience attention to read and digest the results. However, Generative AI (GenAI) is reshaping each of them. GenAI reduces the effort required to create manuscripts. It even enables increasingly engineered research workflows that automate entire parts of the research process, including literature search, programming/implementation, analysis, and writing (Lund et al. 2024). This trend is no exception in the information systems (IS) community, which, however, usually recognizes the disruptive nature of this development faster than many other disciplines. For example, a current special issue of the Information Systems Research journal explicitly seeks papers that demonstrate "*novel research capabilities enabled by GenAI*" (Abbasi et al. 2026). However, from a journal's perspective, these changes present tremendous challenges: When the costs of producing submission-ready manuscripts decrease substantially, the work on the reviewers' side increases inversely.

In this editorial, we provide a BISE perspective on the current debate of the impact of GenAI on academic publishing and the inherent difficulties it poses for journals. We also discuss potential solutions. To do so, we distinguish three stages of AI use in academic research and publishing: 1) simple prompting, where a general-purpose model serves as an ad hoc assistant; 2) collaboration with task-specific GenAI; and 3) knowledge production via engineered multi-step agentic pipelines. While most researchers still operate in the first two stages, the third is advancing quickly, adding further pressure to the editorial process. Although authors increasingly rely on GenAI to generate outputs (Stage 1 and 2), journals require reviewers to evaluate these submissions through unassisted expert judgment. This asymmetry also limits what journals can achieve through prohibition, detection, or disclosure. As AI assistance becomes embedded in writing and research workflows, AI bans become increasingly difficult to enforce. Detection tools offer little relief because they cannot reliably distinguish AI-assisted from manual work, and disclosure requirements remain fragile when authors have incentives to underreport. The resulting challenge is how to preserve credible certification when knowledge production increasingly depends on opaque forms of human–AI collaboration.

Based on these developments, in the short run, journals face congestion and an increasing noise-to-signal problem, which we already observe at BISE. In the longer run, this shift invites a rethinking of how we, as a scientific community, certify, verify, and reward contributions to knowledge. In a world where both the generation and evaluation of knowledge can potentially be automated,

B. Hanneke (✉) · F. Hering · O. Hinz
Faculty of Economics and Business Administration, Goethe University, Frankfurt, Frankfurt am Main, Germany
e-mail: hanneke@wiwi.uni-frankfurt.de

J. Pfeiffer
Institute for Information Systems (WIN), Karlsruhe Institute of Technology, Karlsruhe, Germany

H. Reijers
Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

Published online: 29 June 2026

institutional safeguards (e.g., structured submission requirements, provenance standards, accountability rules, and incentive mechanisms) are paramount (Agrawal et al. 2026).

This editorial encourages the discussion and establishment of institutional safeguards that address three connected challenges: accountability, verification, and incentives. Agents lack *agency*. Hence, automated agentic research pipelines cannot be held responsible for scientific claims, human authors must remain accountable for the pipelines they design, supervise, and report on. At the same time, engineering that lowers authors' manuscript production costs can also reduce journals' verification costs if pipelines are designed for reviewability, traceability, and reproducibility. Finally, safeguards must be embedded in a coherent institutional strategy: journals need short-term operational responses to rising submission volumes, while the scientific community must reassess incentives for a world in which agentic knowledge production is becoming increasingly prevalent.

2 What is Cognitive Automation and Agentic Knowledge Production?

Cognitive automation refers to the automation of tasks that have traditionally been considered cognitive work (Engel et al. 2022). In academic research, these tasks include (amongst others) searching and summarizing literature, formulating hypotheses, designing analyses, generating code, writing and editing text, running statistical analyses, producing figures, and preparing responses to reviewers. Not every task can be automated as of today; however, the feasible frontier is moving fast and, as a result, the marginal cost of producing a "submission-ready" manuscript is falling rapidly. To structure the discussion, we distinguish three stages of AI use in academic research and publishing (summarized in Table 1).

2.1 Stage 1: Ad hoc GenAI Prompting

In Stage 1, researchers prompt AI models to serve as ad hoc assistants. Researchers actively direct the GenAI production process to generate text, outlines, code snippets, and polished prose, to brainstorm ideas, to produce tables or figures, and to write first-pass analysis code.

This stage accelerates research workflows. Experimental evidence beyond academia documents sizeable productivity gains from basic prompting. In a preregistered experiment with 453 college-educated professionals, ChatGPT reduced writing time by 40% and raised output quality by 18% (Noy and Zhang 2023). A field deployment across 5172 customer-support agents found a 15% average productivity gain (Brynjolfsson et al. 2025). In a field experiment with management consultants, GPT-4 access raised task completion by 12.2% and quality by 40% inside the AI frontier while degrading performance on tasks just outside it (Dell'Acqua et al. 2026). Within research settings, AI-assisted reanalysis of published economics studies recovers original results faster while surfacing methodological inconsistencies that human reviewers missed (Brodeur et al. 2025). Hence, researchers can reduce the time required to produce drafts and decrease the marginal cost of revisions.

However, these benefits introduce risks. Authors can easily polish text without enhancing conceptual clarity, methodological rigor, or empirical contribution. AI-generated content may propagate errors, including hallucinated sources, incorrect claims, and plausible/executable but flawed code. Authors may also misrepresent AI-generated content as independently verified scholarly work. These risks reflect how current models operate. Foundation models match probabilistic patterns instead of reasoning or understanding (Gopal et al. 2025; Mitchell and Krakauer 2023). They encode information stylistically, not faithfully (Riemer and Peter 2024), and they lack epistemic agency despite producing fluent conversational output (Peter et al. 2025). There is a risk that these errors compound across studies, systematically degrading research quality (Gopal et al. 2025; Ji et al. 2023). To address these risks, authors must verify claims and sources, validate code and outputs, and take responsibility for errors.

From a journal's perspective, Stage 1 GenAI usage undermines writing quality as a screening signal. Editors encounter submissions that appear competent at first glance yet remain unready for publication. Hence, submissions focused on superficial writing quality increase screening loads (i.e., more time needed to screen), while at the same time eventually driving up desk rejections (i.e., poor quality submissions), and eventually strain reviewer capacity. Journals should clarify disclosure *expectations* and enforce accountability through strict desk-reject policies for violations such as hallucinated sources, e.g. one error that is likely to stem from hallucinations automatically leading to a rejection (this practice is already partly enforced by dissertation committees at German universities) and eventually a submission ban (as recently clarified by arXiv) (Chawla 2026).

2.2 Stage 2: Collaboration with Task-Specific GenAI

In Stage 2, researchers employ specialized tools or GenAI agents that execute defined functions within partly standardized workflows or processes. Unlike Stage 1, these systems support specific process steps rather than single prompt-response interactions. For example, researchers

Table 1 Three stages of AI use in research and publishing

	Stage 1: ad hoc GenAI prompting	Stage 2: task-specific GenAI	Stage 3: agentic knowledge production
Human objective	Improve research content and speed	Improve the research process	Design a research system
AI autonomy	Low	Process-specific	High
Manual human input	High	Medium	Very low
Typical artifacts	Prose, code snippets, summaries	Paper paragraphs, review reports, larger code bodies	Manuscripts, evidence and reproducibility packages
Potential effects	More output, potentially better quality	More output at better quality	Increasingly more output, at consistently better quality, at decreasing cost
Major risks	Hallucinated results and sources, AI slop	Shallow contribution, responsibility diffusion	Hidden structural risks and biases, degradation of scientific findings
Core requirements	Disclosure, basic scholarly hygiene	Structured evidence linking, bounded tool use, and validated outputs	Auditability, provenance, reproducibility

apply structured style improvement, perform citation consistency checks, map claims to evidence, generate robustness checks, and use tools that simulate peer review as internal quality gates before submission (Van Dijcke 2026). Researchers already employ general-purpose coding assistants for causal inference and applied economics workflows (Brunnermeier and Goldsmith-Pinkham 2026; Cunningham 2026), and field-level evidence documents adoption across finance research teams (Novy-Marx and Velikov 2026).

Two technical advances drive Stage 2. First, retrieval-augmented generation (RAG) enables GenAI to access external knowledge at query time, trace sources, and verify claims beyond static parametric recall (Gopal et al. 2025). Second, tool-using GenAI invokes external functions, such as code interpreters, statistical libraries, and APIs, integrating results into the research task (Schick et al. 2023). Neither advance removes the need for human judgment. RAG continues to hallucinate due to retrieval limitations (Ji et al. 2023), and each tool-use step introduces errors that may compound (Gopal et al. 2025). A benchmark across 19 contemporary LLMs finds that even frontier models silently corrupt roughly 25% of document content over long delegated workflows, with corruption worsening with document size and interaction length and not improving when agentic tool use is enabled (Laban et al. 2026). Hence, researchers must curate sources, read primary evidence, validate code, and decide which outputs to accept.

Compared to Stage 1, Stage 2 has the potential to enhance quality by reducing inconsistencies and improving reporting completeness (Filimonovic et al. 2025). However, Stage 2 increases capability gaps because specialized tools may be expensive or require technical infrastructure, risking a digital divide (Liang et al. 2024) which concentrates research advantages among well-funded institutions. Moreover, researchers may develop a false sense of confidence when they treat workflow aids as correctness checks. Furthermore, standardization can homogenize manuscripts and erode the quality and reliability signal of a paper. Heavy reliance on AI tools trained on existing literature may inadvertently push research toward mainstream approaches and conventional thinking, eroding the methodological diversity that enables Information Systems to address complex phenomena (Sarker et al. 2019). It may also prevent science from advancing because the process favors already observed artifacts and approaches.

For journals, Stage 2 increases submissions that meet formal requirements and appear well-structured, which improves readability and might raise the bar for meaningful contributions. Instead of using prose quality as a screening proxy, journals should employ structured early rejections that connect claims to evidence. Specifically, journals should encourage contribution statements and explicit links between claims, data, and analysis. When tools support reviewing, journals should define bounded use cases and require human accountability for the final judgment.

2.3 Stage 3: Agentic Knowledge Production Pipelines

Stage 3 introduces a fundamental shift. Researchers design and operate end-to-end agentic pipelines that generate, test, refine, and document scientific outputs (Lu et al. 2026). These systems integrate data retrieval, code generation and execution, and iterative evaluation loops through modular agents. Researchers move from writing artifacts to designing systems that produce and validate these artifacts. Current prototypes include the Automated Peer Evaluator

(APE) project (Yanagizawa-Drott 2026), E2ER (End-to-End Researcher) (Hanneke 2026), Sakana AI Scientist v2 (Yamada et al. 2025), the zeropaper project (Lopez-Lira 2026), paper orchestrator (Song et al. 2026), ARIS (Yang et al. 2026), and replicability-focused infrastructure (Kohler et al. 2026). A typical Stage 3 pipeline specifies objectives, retrieves prior work, generates hypotheses, runs analyses, stress-tests results, and drafts narrative outputs traceable to evidence.

Two key developments described in the AI capability literature underpin the technical foundations of Stage 3. First, autonomous GenAI agents extend tool-using systems by integrating reasoning (decomposing complex tasks into steps), memory (maintaining context across iterations), and tool use (interacting with APIs, databases, and code interpreters) (Gopal et al. 2025; Wu et al. 2025). Second, multi-agent ecosystems coordinate several specialized GenAI agents, for example, one as a searcher, another as a coder, and another as a critic (Park et al. 2023; Wu et al. 2025). Gopal et al. (2025) warn that the complexity of such systems creates new hazards: error cascades across steps, emergent but unverifiable behaviors, and diffusion of accountability. Researchers have questioned the faithfulness of GenAI's self-generated reasoning traces, since explanations can diverge from the models' underlying decision processes (Chen et al. 2025; Matton et al. 2025). As a result, researchers must pair agentic configurations with strong constraints, such as explicit success criteria, data and citation provenance, and checkpoints at which humans read the primary evidence themselves. Stage 3 has the potential to increase throughput and integrity if pipelines preserve provenance, enforce documentation, reduce untracked manual steps, and produce verifiable artifacts.

From a journal perspective, Stage 3 amplifies risks dramatically. Researchers that optimize for output can flood journals with submissions and introduce new failure modes through hidden pipeline assumptions, breakable tool chains, and incentives to overfit to benchmarks. Automated components encounter security concerns when adversarial inputs challenge them. As contributions increasingly reflect system design rather than manual writing, authorship norms become more complex**.**

Without institutional safeguards, Stage 3 increases journal congestion as researchers can produce increasing amounts of submission-ready manuscripts. With safeguards and intentional design, Stage 3 outputs may potentially lower journals' verification costs while contributing to knowledge generation. Therefore, journals should require auditability, provenance, and reproducibility, including evidence packages and clear descriptions of what each team automated versus performed manually. Additionally, editorial policies should reward verification outputs and discourage volume-maximizing behavior. Even when pipelines operate with high autonomy, human authors retain accountability for designing, supervising, and reporting results generated by these pipelines.

While in Stage 3 humans design the pipeline, in Stage 3*, the pipeline modifies itself through reinforcement learning, automated component selection, or agents that propose and test new configurations. Researchers specify the meta-workflow that learns and updates the workflow. Stage 3* amplifies every Stage 3 issue, for instance researchers face greater challenges in reproducibility when the pipeline at submission differs from the one that produced earlier results.

Overall, the stage distinction matters because most debates about GenAI in research implicitly assume Stage 1 or 2. However, we must recognize that concerns about shallow text generation, hallucinations, and misrepresentation remain real but insufficient for evaluating the impacts of Stages 3 and 3*. Because Stage 3 primarily involves engineering, the relevant question shifts from whether text appears human to how agentic pipelines affect the journal's core role of separating signal from noise. Researchers and institutions must align engineering practices with research quality rather than allowing engineered outputs to flood the journal system, worsening the congestion problem, which we can already observe from Stage 1 and 2.

3 Effects of Cognitive Automation on Academic Publishing

Across fields, cognitive automation impacts journals through four channels, namely, submission volume, screening load, reviewer capacity, and integrity. We characterize each using BISE submission data (Tables 2 and 3) alongside cross-journal evidence.

Submission volume. As drafting costs fall, Submission pressure rises. At BISE, the number of submissions grew from 96 in 2010 to 602 in 2025, roughly doubling between 2020 and 2025. Organization Science reports a 42% volume increase after ChatGPT's late-2022 release, with AI-generated writing accounting for much of the growth (Gartenberg et al. 2026). The submission volume at BISE had already grown before the arrival of ChatGPT, so that the increase cannot be attributed solely to lower production costs, but also to a growing research community, an increasing popularity of the journal, and so on.

Screening load. As more manuscripts clear basic readability, surface cues lose value. BISE's desk-reject share has exceeded 70% since 2019 and acceptance fell from 11.3% in 2020 to 3.8% in 2025; the 2019 jump is procedural, the post-2022 acceleration may be linked to

Table 2 BISE submissions and editorial decisions, 2010–2025. EIC = Editor-in-Chief; DE = Department Editor. Decisions are reported as a share of the year's total submissions

Year	EIC desk-reject (%)	DE desk-reject (%)	Post-review reject (%)	Accept (%)	Total
2010	17.7	51.0	3.1	18.8	96
2011	17.1	34.2	19.7	21.4	117
2012	29.1	35.0	10.7	17.5	103
2013	5.2	64.7	5.2	18.5	173
2014	10.2	63.8	6.6	15.8	196
2015	18.2	58.1	8.9	9.7	236
2016	4.1	69.7	10.3	12.8	195
2017	13.8	52.0	16.0	14.1	269
2018	30.7	41.8	10.2	12.7	244
2019	73.4	7.9	6.7	9.0	267
2020	70.9	7.3	7.0	11.3	302
2021	59.5	13.8	11.4	11.1	333
2022	61.7	8.2	16.1	14.0	329
2023	53.3	13.6	17.9	12.2	403
2024	64.1	12.1	14.6	7.6	471
2025	70.8	8.6	7.8	3.8	602

Table 3 BISE reviewer statistics, 2010–2025. Some reviews are still in progress for submissions from 2025

Year	Reviews in due time (%)	Average number of delayed days	Reviewers agree to review (%)	Average number of days from submission to reviewer agreement[1]	Number of reviewers involved
2010	60.7	14.5	79.4	24.1	122
2011	59.5	15.7	82.7	36.1	174
2012	59.4	14.2	83.9	27.4	165
2013	62.0	15.3	70.1	22.6	222
2014	78.4	19.5	64.6	37.7	262
2015	71.1	22.3	61.5	47.9	282
2016	56.4	11.1	67.4	26.9	266
2017	63.2	11.6	67.7	30.4	348
2018	54.7	12.8	64.1	42.2	244
2019	52.4	15.0	66.0	31.4	236
2020	58.0	12.8	63.6	38.3	288
2021	61.3	13.9	62.0	43.9	373
2022	63.6	13.8	69.4	43.1	278
2023	61.1	12.9	61.2	46.7	320
2024	62.9	13.5	61.1	39.1	238
2025	55.7	13.9	57.5	43.0	270

[1]We considered only the initial review round

cognitive automation. By early 2024, LLM-modified text reached 17.5% of computer-science papers across arXiv, bioRxiv, and the Nature portfolio (Liang et al. 2024), while Organization Science reports a decline in writing quality over the same period (Gartenberg et al. 2026). At BISE, we observe submissions that look competently packaged but vary widely in substantive contribution. With absolute volume of submissions still rising, the screening effort on our side is growing.

Reviewer capacity. Reviewer time becomes the binding constraint when submission numbers increase faster than the reviewer pool. At BISE, reviewer acceptance of invitations dropped from 66.0% in 2019 to 57.5% in 2025. Time from submission to reviewer agreement rose from 31 to 43 days. On-time review completion settled at 55.7% in 2025, near the historical low. While submissions doubled between 2020 and 2025, the reviewer pool contracted from 288 to 270. These trends indicate an increasing need for

researchers who are available for reviewing. Rising desk rejections can only partially absorb this growing pressure. Therefore, we also want to express at this point our gratitude to the vibrant and motivated BISE community, who continually provide collegial and academic feedback.

Integrity. Since 2023, BISE requires authors to disclose AI tool use, though editors cannot reliably distinguish AI-generated from non-AI-generated manuscripts. While at BISE, reviewers are required not to use GenAI to create their reviews or to upload papers to GenAI, cognitive automation has also reached peer reviewing. A text-distribution analysis estimates that 6.5–16.9% of review text submitted to ICLR 2024, NeurIPS 2023, CoRL 2023, and EMNLP 2023 were substantially LLM-modified, with higher shares near submission deadlines and among less-engaged reviewers (Liang et al. 2024). At ICLR 2024, AI-assisted reviews rated paired submissions higher than human reviews in 53.4% of the time, and borderline papers receiving such a review were 4.9 percentage points more likely to be accepted (Latona et al. 2024). The vulnerability is bidirectional, as prompt injections hidden in manuscripts flip LLM-generated reviews from rejection to acceptance at success rates of above 90% (Keuper 2025). Venue experiments at ICLR 2025, AAAI 2026, and STOC 2026 confirm efficiency gains but expose shallow domain understanding, factual errors, scoring bias, and comment homogenization (Gopal et al. 2025; Naddaf 2025).

Hence, the trajectory at BISE (i.e., rising volume, harder screening, falling acceptance) matches what has been reported in the scientific community. However, we are not yet living in a world dominated by Stage 3 pipelines, but effects consistent with Stage 1 and Stage 2 adoption are already apparent, such as rising screening load, more pressure on reviewer capacity, and greater reliance on early rejects. When the cost of preparing a manuscript decreases, the journal's constraint shifts from author time to reviewer time, and the share of papers that are competently written but not ready for publication rises. Congestion is the symptom that is visible in the submission numbers. The response requires a set of measures targeting inflow, early rejects, review capacity, and governance, which we will discuss in the next section.

4 Addressing the Congestion Problem

Congestion occurs when demand (here: submissions) surpasses capacity (here: review capacity). Journals encounter longer delays and heavier screening loads in the short term, and they risk declining certification quality in the longer term. Simply adopting better tools may not be enough to reach a better equilibrium (Gartenberg et al. 2026). During our last BISE editorial board meeting, we already brainstormed potential policy bundles that could address the congestion problem. We have summarized the following ideas to stimulate a broader discussion within our community and are providing this list as an outcome of the brainstorming session. Currently, we do not intend to implement one or all of these ideas:

Active-manuscript cooldown. BISE could limit how many manuscripts a single author can submit for review simultaneously (e.g., no more than two active submissions) or in a certain time period (e.g., two submission per year). The number of co-authors could be used to standardize this number. Once an author reaches the limit, he or she would have to pause additional submissions until a decision is taken or a new time period is reached. Persistent identifiers (e.g., ORCID) enable enforcement. Since sequential submitting to several journals could limit the effectiveness of this measure, one could also think of a joint database between different journals, e.g., the AIS basket journals or Springer journals.

Structured editorial early rejects. We should be stricter in assessing whether a manuscript's contribution is plausible, novel, and within scope before inviting reviewers. Cooldowns reduce inflow of manuscripts, and early rejects can reduce the amount of processing time wasted on them. Dedicated mandatory provision of references including DOIs could increase processing ease and speed, while allowing an easier detection of AI hallucinations.

Review capacity expansion. We could require authors to complete a peer review before submitting a new manuscript, linking submission privileges to active community participation. This mechanism expands the reviewer pool over the long term and distributes the reviewing workload more equitably. However, it also requires additional quality controls. Authors who treat mandatory reviews as an obligation to be checked off and submit low-effort evaluations undermine the system rather than strengthen it.

Portable peer review. Serial re-review of the same paper across venues wastes capacity. Platforms, such as Review Commons, show that transferring manuscripts with reviews reduces repeated referee labor and accelerates decisions. This approach raises effective capacity without requiring more volunteer labor. While we could pursue collaborations with other venues to establish a process, it will take time and will not solve the immediate congestion problem we face.

Submission fees. Journals could introduce submission fees (which is commonly done in the Finance discipline) as refundable deposits or in tiered structures with waivers, reviewer payments, or editorial support. These actions would avoid pricing without waivers, which disadvantages early-career and financially constrained researchers.

Bounded AI use. We could deploy AI for checklisting, summarization, and procedural checks in an advisory role,

following NeurIPS' checklist-assistant model. Gopal et al. (2025) recommend incorporating pluralism into such workflows. For example, algorithmic filters should consider sources beyond top venues and English-language publications. They should also give more weight to work from underrepresented regions and methods. Additionally, they should highlight conflicting findings rather than only presenting the majority view. Such mechanisms could also help to detect hallucinations and low-quality inputs that manual surface screening may miss during editorial early rejections. However, this could be a further step to the dystopian view where AI writes papers that AI reviews and which then are used to train AI while humans increasingly lose touch to research.

Alternative certification channels. We have established BISE as a respected journal in the Information Systems community. Strengthening BISE's positioning as human-first journal, we could open alternative submission channels via collaborations or redirects to other outlays of specific submission types (i.e., "*do not submit to us but to outlay XYZ, which we trust would be better suited.*"). Although opening and operating a separate "*BISE for AI-generated submissions*" could potentially redirect submissions from BISE, and hence relieve the congestion problem, such a separate track seems as of now premature and not a good fit with our scholarly identity.

These operational measures manage volume but do not resolve the deeper question that Stages 3 and 3* raise "What defines a'contribution ' when the unit of work becomes a system"?

5 Outlook: Academic Publishing in the Age of Cognitive Automation

When agentic pipelines lower the cost of generating plausible manuscripts, paper supply eventually outpaces audience attention. The community cannot rely on metrics shaped during a time when producing a single polished paper required substantial resources. Moreover, today's dominant metrics (i.e., publications, citations, placements) fail to reward activities essential for cumulative knowledge production. Cognitive automation widens this gap by making manuscript production more attractive than strengthening the robustness and reusability of findings. A deeper risk is the cultural shift that occurs when communities start to equate fluent output with understanding, and convenience crowds out core scholarly habits such as patience with complexity, tolerance for ambiguity, and willingness to challenge assumptions (Acemoglu et al. 2026; Gopal et al. 2025). Eventually, this shift results in a failure to engage with the work of academic colleagues. The preservation of knowledge and scientific expertise among human colleagues and younger generations should itself be considered a desirable outcome. Creating a machine that "knows" and performs all research while humans progressively know less and less is not a desirable direction.

Since academic publishing and knowledge production might become two different things (Cunningham 2026), journals and communities should broaden their definitions and understandings of what qualifies as a scientific contribution. Football clubs, for example, do not just track scored goals but also assists, defensive stops, and coaching infrastructure. Scholarly publishing historically over-recognizes the equivalent of goals (novel, significant, citable findings) and undervalues reproductions, robustness work, and reusable infrastructure. As the marginal cost of producing "goals" drops, verification work gains relative value. Journals can recognize verified reproductions, well-designed null findings and boundary conditions, evidence packages (data, code, documentation), robustness contributions (sensitivity analyses, alternative specifications), and reusable research infrastructure. Stage 3 pipelines have the potential to produce these verification outputs efficiently by default. When journals treat them as first-class contributions, agentic pipelines become part of the solution instead of a source of congestion.

The overarching epistemic question about "What defines a 'contribution'?" is non-trivial, and it touches on philosophical debates around agency, intentionality, and the attribution of creative or intellectual authorship, which would go beyond the scope of this editorial. Therefore, we will take up this philosophical dimension in one of the following editorials, where it will receive the dedicated discussion it demands.

6 Conclusion

BISE researchers have long studied how information systems change organizations and institutions. GenAI is now also changing the academic publication process. Optimistic scholars argue that agentic knowledge production lowers the cost of rigorous research by strengthening documentation, reproducibility, and verification. And certainly, all of us would be satisfied if AI solved pressing problems in various disciplines. However, pessimistic scholars also have a point when they say that the same tools generate spam and distort incentives when researchers and institutions optimize for volume, and that they could eventually eliminate human research in the future. Ultimately, institutional incentives will determine the outcomes. Journals must address congestion as an operational problem while addressing the underlying transition of automated knowledge production by responding with policies and

safeguards that preserve scientific values. The choices that researchers and journals make now will define academic research for the decades to come.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Abbasi A, Bichler M, Gopal R, Karahanna E, Sarker S (2026) ISR special issue: generative AI and new methods of inquiry in information systems research. https://pubsonline.informs.org/page/isre/calls-for-papers

Acemoglu D, Kong D, Ozdaglar A (2026) AI, human cognition and knowledge collapse. National Bureau of Economic Research Working Paper W34910. https://doi.org/10.3386/w34910

Agrawal A, McHale J, Oettl A (2026) AI in science. National Bureau of Economic Research Working Paper W34953. https://doi.org/10.3386/w34953

Brodeur A, Sung SY, Miguel E, Vilhuber L, De La Guardia FH (2025) Assessing reproducibility in economics using standardized crowd-sourced analysis. National Bureau of Economic Research Working Paper W33753. https://doi.org/10.3386/w33753

Brunnermeier M, Goldsmith-Pinkham P (2026) Claude code for applied economists. https://markusacademy.substack.com/

Brynjolfsson E, Li D, Raymond L (2025) Generative AI at work. Q J Econ 140(2):889–942. https://doi.org/10.1093/qje/qjae044

Chawla D, Singh (2026) Researchers who use hallucinated references to face arXiv ban. Nature News, 19 May 2026. https://www.nature.com/articles/d41586-026-01595-5

Chen Y, Benton J, Radhakrishnan A, Uesato J, Denison C, Schulman J, Somani A, Hase P, Wagner M, Roger F, Mikulik V, Bowman SR, Leike J, Kaplan J, Perez E (2025) Reasoning models don't always say what they think. arXiv, version 1. https://doi.org/10.48550/ARXIV.2505.05410

Cunningham S (2026) Claude code for causal inference. https://causalinf.substack.com/s/claude-code

Dell'Acqua F, McFowland E, Mollick E, Lifshitz H, Kellogg KC, Rajendran S, Krayer L, Candelon F, Lakhani KR (2026) Navigating the jagged technological frontier: field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organ Sci 37(2):403–423. https://doi.org/10.1287/orsc.2025.21838

Engel C, Ebel P, Leimeister JM (2022) Cognitive Automation Electron Mark 32(1):339–350. https://doi.org/10.1007/s12525-021-00519-7

Filimonovic D, Rutzer C, Wunsch C (2025) Can GenAI improve academic performance? Evidence from the social and behavioral sciences. SSRN. https://doi.org/10.2139/ssrn.5225360

Gartenberg C, Hasan S, Murray A, Pierce L (2026) More versus better: artificial intelligence, incentives, and the emerging crisis in peer review. Organ Sci orsc.2026.ed.v37.n3. https://doi.org/10.1287/orsc.2026.ed.v37.n3

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. Inf Syst Res 36(4):1949–1967. https://doi.org/10.1287/isre.2025.editorial.v36.n4

Hanneke B (2026) E2ER: end-to-end researcher, version v0.81 [computer software]. Zenodo. https://doi.org/10.5281/zenodo.20452754

Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P (2023) Survey of hallucination in natural language generation. ACM Comput Surv 55(12):1–38. https://doi.org/10.1145/3571730

Keuper J (2025) Prompt injection attacks on LLM generated reviews of scientific publications (arXiv:2509.10248). arXiv. https://doi.org/10.48550/arXiv.2509.10248

Kohler B, Zollikofer D, Einsiedler J, Hoyle A, Ash E (2026) Read the paper, write the code: agentic reproduction of social-science results. arXiv, version 1. https://doi.org/10.48550/ARXIV.2604.21965

Laban P, Schnabel T, Neville J (2026) LLMs corrupt your documents when you delegate (arXiv:2604.15597). arXiv. https://doi.org/10.48550/arXiv.2604.15597

Latona GR, Ribeiro MH, Davidson TR, Veselovsky V, West R (2024) The AI review lottery: widespread AI-assisted peer reviews boost paper scores and acceptance rates. arXiv. https://doi.org/10.48550/arXiv.2405.02150

Liang W, Zhang Y, Wu Z, Lepp H, Ji W, Zhao X, Cao H, Liu S, He S, Huang Z, Yang D, Potts C, Manning CD, Zou JY (2024) Mapping the increasing use of LLMs in scientific papers. arXiv. https://doi.org/10.48550/arXiv.2404.01268

Lopez-Lira A (2026) ZeroPaper: an autonomous research system. https://doi.org/10.5281/ZENODO.20127842

Lu C, Lu C, Lange RT, Yamada Y, Hu S, Foerster J, Ha D, Clune J (2026) Towards end-to-end automation of AI research. Nature 651(8107):914–919. https://doi.org/10.1038/s41586-026-10265-5

Lund B, Lamba M, Oh SH (2024) The impact of AI on academic research and publishing. arXiv, version 1. https://doi.org/10.48550/ARXIV.2406.06009

Matton K, Ness RO, Guttag J, Kıcıman E (2025) Walk the talk? Measuring the faithfulness of large language model explanations. arXiv, version 2. https://doi.org/10.48550/ARXIV.2504.14150

Mitchell M, Krakauer DC (2023) The debate over understanding in AI's large language models. Proc Natl Acad Sci USA 120(13):e2215907120. https://doi.org/10.1073/pnas.2215907120

Naddaf M (2025) AI is transforming peer review — and many scientists are worried. Nature. https://doi.org/10.1038/d41586-025-00894-7

Novy-Marx R, Velikov M (2026) Artificial intelligence–powered (finance) scholarship. J Econ Lit 64(1):5–37. https://doi.org/10.1257/jel.20251821

Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. Science 381(6654):187–192. https://doi.org/10.1126/science.adh2586

Park JS, O'Brien JC, Cai CJ, Morris MR, Liang P, Bernstein MS (2023) Generative agents: interactive simulacra of human behavior. arXiv, version 2. https://doi.org/10.48550/ARXIV.2304.03442

Peter S, Riemer K, West JD (2025) The benefits and dangers of anthropomorphic conversational agents. Proc Natl Acad Sci USA 122(22):e2415898122. https://doi.org/10.1073/pnas.2415898122

Riemer K, Peter S (2024) Conceptualizing generative AI as style engines: application archetypes and implications. Int J Inf Manag 79:102824. https://doi.org/10.1016/j.ijinfomgt.2024.102824

Sarker S, Chatterjee S, Xiao X, Elbanna A (2019) The sociotechnical axis of cohesion for the IS discipline: its historical legacy and its continued relevance. MIS Q 43(3):695–719. https://doi.org/10.25300/MISQ/2019/13747

Schick T, Dwivedi-Yu J, Dessì R, Raileanu R, Lomeli M, Zettlemoyer L, Cancedda N, Scialom T (2023) Toolformer: language models can teach themselves to use tools. arXiv, version 1. https://doi.org/10.48550/ARXIV.2302.04761

Song Y, Song Y, Pfister T, Yoon J (2026) PaperOrchestra: a multi-agent framework for automated AI research paper writing. arXiv, version 1. https://doi.org/10.48550/ARXIV.2604.05018

Van Dijcke D (2026) Coarse [computer software]. https://github.com/Davidvandijcke/coarse

Wu J, Zhu J, Liu Y, Xu M, Jin Y (2025) Agentic reasoning: a streamlined framework for enhancing LLM reasoning with agentic tools. arXiv, version 2. https://doi.org/10.48550/ARXIV.2502.04644

Yamada Y, Lange RT, Lu C, Hu S, Lu C, Foerster J, Clune J, Ha D (2025) The AI scientist-v2: workshop-level automated scientific discovery via agentic tree search. arXiv, version 1. https://doi.org/10.48550/ARXIV.2504.08066

Yanagizawa-Drott D (2026) Automated peer evaluator (APE) [computer software]. Social Catalyst Lab. https://ape.socialcatalystlab.org/

Yang R, Li Y, Li S (2026) ARIS: autonomous research via adversarial multi-agent collaboration. arXiv, version 1. https://doi.org/10.48550/ARXIV.2605.03042

Yoo Y (2024) Evolving epistemic infrastructure: the role of scientific journals in the age of generative AI. J Assoc Inf Syst 25(1):137–144. https://doi.org/10.17705/1jais.00870