JOM Forum: Generative AI and Empirical Research Methods in Operations Management

Timofey Shalpegin (tshalpegin@mbsc.edu.sa), Prince Mohammed Bin Salman College of Business and Entrepreneurship, King Abdullah Economic City, Saudi Arabia
Tyson R. Browning (t.browning@tcu.edu), Neeley School of Business, Texas Christian University, Fort Worth, TX, USA
Ajay Kumar (akumar@em-lyon.com), EMLYON Business School, Lyon, France
Guangzhi Shang (gshang@business.fsu.edu), W. P. Carey School of Business, Arizona State University, Phoenix, Arizona, USA, College of Business, Florida State University, Tallahassee, Florida, USA

1. Introduction

Generative Artificial Intelligence (Gen-AI) is arguably the fastest-adopted technology in history (Mariani & Dwivedi, 2024). Like past transformative technologies—such as computers and the Internet—Gen-AI brings new opportunities and challenges to research. However, its distinctive features may result in an adoption pattern and impact that differ from those of earlier technologies. Anthony et al. (2023) offered a novel perspective on studying AI. Traditionally, technologies are viewed either as tools to improve performance or as mediums to enhance collaboration; however, AI can be seen as a counterpart or an agent interacting with human agents (c.f. Bendoly et al. 2024, Angelopoulos et al. 2023). Along these lines, the popular press has already labeled Gen-AI models as a "superhuman research assistant" in the research process (*The Economist*, 2023). With the formation of such hybrid teams in research, it is more critical than ever to define the roles of team members.

Gen-AI refers to a class of AI algorithms that can generate novel content—such as text, images, and code—based on the data upon which they have been trained, rather than merely analyzing existing information. Unlike traditional analytical tools, Gen-AI can actively contribute to the creative and decision-making processes. This broad category includes a variety of tools, such as image generators, code generation assistants, and large language models (LLMs), which specialize in natural language processing. As AI shifts from a tool to a research counterpart, clear ethical guidelines are essential.

Gen-AI is set to profoundly shape research in OM. This Forum discussion explores its potential impact by examining current editorial policies in OM journals, reviewing the present use of Gen-AI in OM research, speculating on future developments, and addressing ethical issues. By integrating these perspectives, it aims not to replace existing editorial policies or ethical guidelines but to initiate a dialogue within the OM community.

2. Editorial Policies on Gen-AI

The rapid adoption of Gen-AI has sparked debate about its impact on research integrity. Bockting et al. (2023) warned that Gen-AI is altering how scientists find information, conduct studies, and evaluate publications. To address these challenges, they introduced a set of "living guidelines" to define acceptable uses of Gen-AI. Editorial policies soon followed across disciplines, adapting research standards to these emerging challenges. OM journals have introduced comprehensive guidelines on Gen-AI (*JOM*, 2024), permitting its use in language improvement, literature classification, and data collection, coding, and analysis while restricting synthetic data generation and image manipulation. Table 1 summarizes the key points of this policy as well as Wiley's general policy on AI.

Table 1. Editorial Policies

	***JOM*, *POM*, *Decision Science*, *IJPE* Shared Policy (*JOM*, 2024)**	**Wiley (2025)**
Authorship	AI tools cannot be listed as authors	
Disclosure	Use of AI (except for language improvement) must be disclosed in the methods section, acknowledgments, and submission page. Justification is required.	Use of AI (except for language improvement) must be described in detail in the methods section or acknowledgment. The final decision on AI use is at the editor's discretion.
Language Improvement	Permitted, but authors must check for potential plagiarism in AI-assisted edits	Permitted for spelling, grammar, and general editing
Ideation	Not explicitly mentioned	
Literature Classification	Permitted, but with disclosure and verification requirements	Not explicitly mentioned
Data Collection	Permitted, but AI cannot generate synthetic data	Permitted, but AI cannot create, alter, or manipulate original research data and results
Coding Assistance	Permitted for processing large corpora and rich media, with verification requirements	Not explicitly mentioned
Data Analysis	Permitted if outputs are replicable and interpretable by humans. AI cannot be a black-box analysis tool.	Not explicitly mentioned, but authors are responsible for accuracy and verification.
Image Creation	Prohibited for creation or manipulation of images, figures, and other empirical data	Prohibited for creating, altering, or manipulating figures or images unless justified by editorial discretion
Review Process	Editors and reviewers cannot use AI tools to generate reports or upload manuscripts into AI tools	Editors and peer reviewers may use AI to improve written feedback but must declare it. Manuscripts must not be uploaded into AI tools.

Both policies are generally aligned and represent a solid foundation for researchers engaging Gen-AI tools in their work. Notably, they do not require the disclosure of prompts and Gen-AI's responses, unlike the *Science* family of journals (Thorp & Vinsen, 2023). Additionally, they do not explicitly address idea generation and exploration, unlike other policies (e.g., Taylor & Francis, 2025).

However, the shared journal policy (*JOM*, 2024) does refer to the "prohibition of AI as a source of generalized overviews, ideas and concepts." It is possible that this could be unintentionally misinterpreted so as to make some researchers reluctant to use Gen-AI during the ideation phase. Some further discussion is therefore warranted.

Girotra et al. (2023), for example, demonstrated how Gen-AI can generate ideas, comparing the quality of these ideas to those produced by human subjects and showing that Gen-AI can be beneficial due to its high productivity. Supporting evidence comes from neuroscience research, where LLMs have already been shown to predict study outcomes better than human experts (Luo et al., 2024). This advantage arises because Gen-AI can access and process more information about existing research than any individual human, enabling it to identify parallels between a proposed study and similar ones already conducted in other fields. This capability can lead to reasonably reliable predictions. If this finding generalizes to other disciplines, including OM, Gen-AI could assist in identifying promising research directions and heading off redundant studies.

Along these lines, several research groups in computer science are actively working on developing agents capable of autonomous scientific research (e.g., Jansen et al., 2024; Lu et al., 2024), which could enable Gen-AI not only to generate research ideas but also to test them. Interestingly, Gyory et al. (2022) found that AI agents tend to perform unexpectedly well as process managers in engineering projects, in some instances leading to superior performance by AI-managed teams. Taken together, do these findings point toward a future transition of Gen-AI from the role of research assistant (*The Economist*, 2023) to that of research advisor?

3. Gen-AI's Impact on Empirical Research

Statistical Analysis: Gen-AI tools are advancing statistical analysis and machine-learning applications. Rossi et al. (2024) noted that such tools are increasingly capable of self-correction and self-improvement, which enhances their reliability over time. This scalability has the potential to significantly boost research productivity and innovation in OM. Specifically, Gen-AI facilitates a wide range of data-related tasks, including data preprocessing, anomaly detection, and the interpretation of unstructured data, thereby streamlining complex analytical processes and reducing manual effort. These benefits align with findings from Jackson et al. (2024), who emphasized that AI-driven automation improves data organization and retrieval, particularly in empirical studies that require large-scale dataset structuring.

Cao et al. (2024) further illustrated the impact of AI in statistical modeling by demonstrating how machine-learning algorithms enhance predictive accuracy across various business applications. While predictive accuracy does not imply causal inference, applying AI models across different datasets may

offer new pathways for generating scientific insights. For example, supply chain risk-prediction models trained on disruption events in global logistics networks can be tested across countries or supplier tiers to reveal which risk factors (e.g., lead-time fluctuations, country risk indices, tariffs, supplier financial health, order fulfillment rates, and geopolitical instability) are recurring. In addition, comparing feature (risk factor) importance across datasets can help identify robust predictors that may serve as candidates for future causal-inference studies.

While Gen-AI facilitates statistical analysis, it also enables misuse, from data forgery to subtler forms of research misconduct. For instance, researchers might exploit it to find correlations before forming a theory, exacerbating the data-snooping bias (White, 2000) and increasing false positives. This issue is particularly concerning when AI is used for exploratory data analysis without proper statistical controls, potentially leading to overfitting and spurious relationships in empirical studies. Additionally, AI-driven anomaly detection, while valuable, requires careful validation. Kim & Kim (2023) found that deep learning models can outperform traditional statistical techniques in identifying patterns and outliers, but their interpretability remains limited, raising concerns about the transparency of AI-driven statistical processes.

Natural Language Processing: The capabilities of Gen-AI in natural language processing and image recognition have also proven valuable. For example, Niu et al. (2025) employed ChatGPT-4 for text-based measurement tasks in financial filings, showcasing how LLMs can augment traditional natural language processing techniques. Their findings align with Bail (2024), who highlighted the role of AI-assisted text analysis in systematically processing large volumes of unstructured data, such as corporate disclosures and social media content, to extract structured insights with greater efficiency. Similarly, Jackson et al. (2024) emphasized that Gen-AI improves data retrieval and structuring in supply chain management research by enhancing the classification of complex datasets. Along the same lines, De Kok (2025) highlighted that generative LLMs can match or even exceed traditional machine-learning methods in text classification, sentiment analysis, and risk assessment. Unlike conventional supervised approaches requiring extensive, labeled training data, generative LLMs enable efficient classifications, significantly reducing data processing time. However, ensuring construct validity is still essential, as AI-generated classifications must be benchmarked against human-labeled data to maintain methodological rigor.

Beyond text analysis, Gen-AI has been successfully applied to image recognition and content comparison. Zhou and Lee (2024) utilized Gen-AI models to measure content similarity between images, offering new tools for method-focused research. Additionally, Huang et al. (2024) leveraged Gen-AI to create reading materials for eye-tracking experiments, demonstrating its utility in designing experimental stimuli.

Reproducibility: Another key advantage of Gen-AI in empirical OM research is its role in enhancing reproducibility by improving experimental design and forecasting expected outcomes. Lippert et al. (2024) demonstrated that GPT-4 performed as well as a cohort of 119 human experts in forecasting effect sizes in a complex behavioral experiment, with a correlation of 0.89 between its predictions and realized outcomes. This suggests that Gen-AI may serve as a valuable piloting tool before data collection, potentially optimizing research design and hypothesis testing. Although this approach does not replace empirical validation, improved design rigor can contribute to more reproducible findings. However, the study also found that earlier models, such as GPT-3.5, performed significantly worse, highlighting the variability in AI model reliability and the need for careful validation of AI-generated research insights.

At the same time, concerns remain regarding the reliability of AI-generated documentation and its impact on long-term reproducibility. Bail (2024) warned that the probabilistic nature of AI-generated content introduces inconsistencies, particularly as models evolve over time, making exact replication of AI-assisted research difficult. Moreover, discrepancies in how different Gen-AI models handle the same tasks further complicate reproducibility. For example, Ziems et al. (2024) reported substantial variations in text classification performance across models. Overall, while using Gen-AI for replication studies might be premature, it could be beneficial in identifying potential targets for such studies (Davis et al., 2024).

4. Ethical Implications

Despite its potential, the integration of Gen-AI into decision-making processes raises significant ethical concerns that researchers must actively address in their work. A survey by Van Noorden and Perkel (2023) highlighted the research community's ethical concerns about AI, emphasizing the need for scholars to navigate these challenges responsibly. Pazzanese (2020) highlighted three ethical concerns AI presents to society: (1) privacy, (2) bias and discrimination, and (3) the role of human judgment. These issues are particularly relevant in OM research, where the use of Gen-AI intersects with sensitive data handling, decision-making, and research integrity.

4.1. Privacy

Gen-AI systems in OM research often need access to extensive datasets, including sensitive information about suppliers, customers, and employees. This raises privacy concerns, especially when dealing with personally identifiable or commercially sensitive data. Some organizations have restricted Gen-AI usage due to these concerns (Fui-Hoon Nah et al., 2023), but the full implications for research practices remain underexplored. Cohen et al. (2022) called for stricter AI-specific data privacy regulations, which implies the need for researchers to proactively implement rigorous data security

protocols.

To protect sensitive information, researchers using Gen-AI must ensure that their methodologies comply with institutional and editorial privacy policies. One common approach is using Gen-AI systems that do not store prompts and responses or contribute training data to external models. While this minimizes data exposure risks, it also limits the system's ability to learn from real-world research applications. Researchers should be aware of potential moral hazards, where AI providers may claim strong privacy measures while having incentives to relax them for model enhancement. Compliance requires careful vetting of AI tools and adhering to transparent data-handling practices, which should be explicitly addressed in research methodologies. While Hu et al. (2022) and Fainmesser et al. (2023) have examined data privacy challenges in AI systems, further research is needed to adapt their findings to the research context.

4.2. Bias and Discrimination

Gen-AI systems can introduce biases that affect research outcomes, originating from training data, data selection processes, and algorithmic design (IBM, 2023). For example, if training datasets over-represent certain regions or demographics, research findings may be skewed, leading to misleading conclusions. In OM, this can be manifested in biased supplier selection tools, demand forecasts, innovation models, or hiring recommendations. As Chen (2023) demonstrated, algorithmic bias in AI-enabled recruitment can reinforce discriminatory patterns.

The impact of Gen-AI biases on the OM research process remains largely unexplored. Chen et al. (2025) found that while Gen-AI exhibits human-like biases in preferences, it is less biased in calculation-based decisions. Applying this to research, the use of Gen-AI for language improvement, literature classification, and ideation may be particularly susceptible to bias. However, limited evidence from other fields suggests that Gen-AI can also help mitigate bias in existing datasets. For example, in biomedical applications, generative models have been used to address the underrepresentation of certain demographic groups in clinical datasets (van Breugel et al., 2024).

Overall, OM researchers should incorporate bias detection and mitigation strategies into their methodologies, such as bias audits, adversarial testing, and transparency in AI-generated content. Rather than relying on Gen-AI as an unquestioned research aid, empirical studies should systematically assess its reliability and limitations.

4.3. Human Judgment

The increasing reliance on Gen-AI in decision-making processes within OM research can lead to an overreliance on technology, potentially sidelining human intuition, creativity, and ethical considerations. A central challenge is determining the appropriate role of AI in research decision-

making. While Gen-AI can enhance productivity, its recommendations may lack contextual awareness and ethical sensitivity, which are essential in research design and interpretation. Researchers must be aware of the trade-offs between efficiency and depth of understanding when integrating Gen-AI into their workflows. For instance, if AI-generated literature reviews prioritize frequently cited works, they may reinforce existing paradigms rather than fostering novel insights. Similarly, Gen-AI's capacity for ideation may depend on how it is used: while it can generate research questions efficiently, its effectiveness in refining innovative concepts remains contested (Girotra et al., 2023).

Another key consideration is cognitive diversity in research decision-making. While Gen-AI can streamline research tasks, overreliance on AI-generated outputs risks homogenizing research perspectives and limiting creativity (Messeri & Crockett, 2024). However, the impact of Gen-AI on creativity and decision-making remains inconclusive, with studies yielding contrasting findings. Chen and Chan (2024) found that when Gen-AI played a leading role in content generation tasks, it diminished output quality for expert creators, whereas using Gen-AI for feedback was beneficial for non-experts. In contrast, Jia et al. (2024) showed that Gen-AI can enhance creativity by automating repetitive, initial tasks. Further complicating the picture, Boussioux et al. (2024) found that solutions developed exclusively by humans exhibited higher novelty, while AI-assisted solutions demonstrated superior overall quality. These mixed findings underscore the importance of defining clear boundaries for Gen-AI's role in OM research. Do they also suggest that doctoral students and junior scholars—those in the OM research community with perhaps the highest familiarity with Gen-AI—might also not be in the best position to evaluate its consequences for the research process?

5. Conclusion

Gen-AI is transforming OM research by automating tasks, analyzing data, and generating insights, yet ethical and methodological challenges require careful consideration. This article examines Gen-AI's impact on OM research, focusing on editorial policies that define its responsible use. These policies reflect the tension between leveraging Gen-AI's benefits and maintaining research integrity, particularly in areas such as ideation and reproducibility. Furthermore, the adoption of Gen-AI brings up significant ethical challenges. Issues such as data privacy, algorithmic bias, and the potential erosion of human judgment require immediate attention. Table 2 summarizes some discussion points raised —but not answered—in this paper, alongside some arguments in favor and against. These points are intended to initiate a broader discussion in the OM community.

Gen-AI presents both opportunities and challenges in OM research. Responsible adoption requires ethical awareness, rigorous standards, and commitment to research integrity. The journey ahead calls for a collective effort from researchers, editors, and publishers to shape the future of Gen-AI in the OM research community, ensuring its positive impact on academia, industry, and society at

large.

Table 2. Suggested Discussion Points

Discussion Point	Arguments in Favor (Encouraging Symbiotic, Co-Creative Use of AI)	Arguments Against (Addressing Limitations, Biases, and Ethical Concerns)
Should the policy address the use of Gen-AI for idea generation?	Gen-AI can serve as a co-creative partner, helping researchers refine ideas, identify gaps, and explore cross-disciplinary insights. Researchers can use AI iteratively, rather than relying on it as a passive generator of ideas.	The risk of bias in AI-generated ideas mirrors human biases but is harder to detect. Over-reliance could stifle creativity and critical thinking. AI may reinforce mainstream ideas while overlooking unconventional but valuable perspectives.
Must prompts and Gen-AI-produced responses be disclosed in the review process?	Documenting AI interactions enhances transparency, reproducibility, and accountability. Researchers can pre-empt false-positive accusations of AI misuse by maintaining logs of AI contributions.	Strict documentation policies could discourage legitimate AI use and create unnecessary administrative burdens. Privacy concerns arise if AI prompts or responses contain proprietary or confidential research content.
Should Gen-AI be permitted to conduct exploratory data analysis in OM research?	Gen-AI can help by identifying patterns, anomalies, and relationships in large datasets more efficiently than traditional methods. This could accelerate nascent theory development and reduce manual effort in data preprocessing.	Overreliance on AI may lead to data snooping bias, increasing false positives. AI-driven insights could also be misinterpreted if researchers do not critically evaluate the validity of the detected patterns.
Can Gen-AI improve research reproducibility in OM?	AI can enhance reproducibility by automating the documentation of data analysis steps, ensuring consistency in research protocols, and identifying potential sources of errors. It can also help identify which studies are more worthy of replication.	The probabilistic nature of AI outputs introduces variability, making exact replication challenging. Different AI models may yield inconsistent results on the same task, complicating verification efforts.

Commentary

Jason Thatcher (Jason.Thatcher@colorado.edu), University of Colorado

Every ten years, a new generation of tools appears that challenges how we conduct research.

In the 1980s, we witnessed the first wave of statistical tools arrive on desktop computers—and naysayers said this was the end of academic research because PhD students would no longer understand statistics or how to use these tools. Today, we use the descendants of these tools without a thought. In the 1990s, we witnessed the emergence of online databases and reference management tools—and naysayers said research quality would decline because PhD students would no longer know how to search the literature or read academic papers. Today, we see these tools used to craft rich and thoughtful papers. In the early 2000s, we witnessed the Internet unlock access to archival data and online survey panels—and naysayers said our understanding of problems would decline because PhD students would not have an intimate understanding of data sources. Today, we see such data used to solve a vast array of interesting and relevant problems. In the 2010s, we witnessed social media disrupt how we share research with academic and non-academic audiences—and naysayers said this would make it harder to distinguish high-quality from low-quality work, because PhD students could now promote their findings broadly. Today, even the most hardened old faculty post their findings to SSRN, LinkedIn, and other platforms.

Now, in the 2020s, we are witnessing Generative AI demonstrate "superhuman research capabilities"—and naysayers claim we will no longer need humans for research because PhD students cannot match the cognitive capabilities of large language models. Yet, I see PhD students using generative tools to enable more sophisticated and nuanced research. While acknowledging there are many questions to resolve about the use of Generative AI in academic research, the cat is out of the bag: we can no more turn away from it than we can from statistical software, online databases, internet-enabled data access, or social media dissemination of our work.

If the cat is out of the bag, what should academics do? How can we collectively navigate the challenges posed by Generative AI and emerge stronger as a broader academy and across our disciplines? First, we must acknowledge that Generative AI differs from prior innovations. Earlier technological advances primarily supported scholars in their quest for truth. For instance, statistical tools made quantitative analysis accessible, yet still required researchers to formulate questions, interpret data, and derive conclusions. In contrast, generative AI actively participates in the quest for truth, as it is capable of generating data, text, analysis, and even complete research papers. This difference raises critical questions: (a) Who should have primacy in the research process—humans, AI, or both? (b) Who is responsible when AI generates content? and (c) How do we assess originality

and plagiarism in AI-generated research? These latter two questions are especially important since generative AI often relies on data of unclear origins and quality.

Second, we must recognize that many questions posed by AI-skeptics today echo past concerns. Returning to statistical software, critics once argued that students' foundational understanding of statistical principles would erode, creating tool-dependent scholars who lacked conceptual depth. Yet, academic culture adapted and survived. Advanced PhD seminars shifted from teaching statistical mechanics to emphasizing conceptual understanding and application to real problems, producing scholars equipped to handle contemporary business and societal challenges. Similarly, evaluation shifted from testing technical proficiency to assessing students' abilities to explain, interpret, and justify their methodological choices. I hear similar skepticism today; in fact, I shared such concerns when generative AI first emerged. Yet, I now see that the sky is not falling. Instead, we will adapt—shifting our teaching toward problematization, conceptualization, and exploration. By doing so, we will prepare scholars to leverage advanced generative AI tools to enhance scholarly inquiry.

Third, and most importantly, just as statistical software, online databases, internet-enabled data, and social media positively transformed academia, we should embrace generative AI as a powerful opportunity for enhancing scholarship. When I look to the future, I envision researchers leveraging generative AI to explore literature, identify overlooked research gaps, or apply statistical methods to clarify complex challenges. For example, in a single seminar session, a professor could use generative AI to quickly simulate large datasets, visualize intricate relationships, and collaboratively explore interactions between variables in regression models, providing clearer insights, more robust conclusions, and a richer understanding of research processes. Such approaches could enable scholars to address previously inaccessible problems, both in the classroom and beyond, facilitating connections, conversations, and collaborations across disciplines—from data science and economics to history and sociology—on a scale previously unimaginable.

As noted by Shalpegin et al., the path forward will indeed be challenging. Many questions remain regarding authorship, ethics, data transparency, accountability, training, and equitable access. Of particular importance, we must ensure no scholar or institution, regardless of stature, is left behind in accessing these tools—there is no justification for exclusion in this new era. Yet, I know in my bones that if we thoughtfully and ethically integrate generative AI into research processes, we will elevate the quality, relevance, and accessibility of scholarly research. I've personally witnessed how generative AI transformed my own approach to writing and data analysis, revealing challenges and opportunities I hadn't imagined before. Rather than replacing human scholarship, generative AI will enrich it, empowering scholars to explore ideas more boldly, critically, and innovatively than ever before.

The cat is indeed out of the bag. Rather than attempting to put it back, let's thoughtfully embrace

generative AI, guiding its integration ethically to empower scholars, enrich our research, and fundamentally elevate the impact of academic inquiry.

Commentary

Jan C. Fransoo (jan.fransoo@tilburguniversity.edu), School of Economics and Management, Tilburg University

As Generative AI (GenAI) has been developing rapidly, researchers have started to experiment with its use in their research. Since many schools measure the productivity of research as the number of papers published per scholar (potentially weighted by some journal quality metric or limited to a specific number of journals of perceived higher quality), researchers who are targeting productivity-based incentives, such as obtaining degrees, tenure, or promotion, may seek application of these methods to their benefit. History has demonstrated that not all researchers may deploy methods ethically; universities and scholarly communities have increasingly put guardrails in place to avoid that false knowledge is generated (such as through forging of data) or knowledge is generated using false methods (such as plagiarism). Guardrails that have been put in place include now widely deployed practices such as pre-registration of experiments, publication of replication files following journal acceptance, or fully open science research traces in which each stage of the research process is time-stamped and logged in an Open Science database like https://osf.io/ (Foster and Deardorff, 2017). Further, also a system of punitive measures has been put in place, such as universities that may decide to reprimand faculty, or withdraw degrees.

As GenAI emerges and becomes more widely applied, the current guardrails system may need to be further expanded. I feel that the general perspective of Open Science, where researchers verifiably log their research process may provide a sound basis to do so; also prompts can easily be logged. This will further bring the added benefit of journal papers becoming much shorter as they can focus on the main theory and results sections while referring the reader to their timestamped log for everything regarding to data, methods, and analysis (which GenAI will be able to conveniently summarize). At last, we may also have our papers in length and accessibility similar to those which journals like *Nature* and *Science* have been able to do. Therefore, the more interesting question is how the research process itself will change and become more productive. In this case, I define "more productive" not in the sense of more research results or research papers per researcher, but rather more knowledge created per researcher. For instance, in biology, substantially more novel proteins have been sequenced in just the past year than ever before (Service, 2025) with potentially significant health benefits. Before this, such research might have been too expensive or time consuming as it would require extensive experimentation.

Empirical Operations Management is an applied scientific discipline in which the management of actual operations is studied. Given the engineering origins of our discipline, we are also generally interested in improving the outcome of those operations. As our research methods have become more data-intense over the past few decades, we have seen a bias in research emerging on sectors where ample data are available. For instance, online retail tends to be studied much more than physical retail, and high-tech manufacturing in developed countries tends to be much more studied than more manual agricultural production in developing countries. The availability of data drives our research arguably more than the value that certain research may create in developing theory or empirical understanding. With GenAI being more supportive of handling unstructured data and making the research process more efficient by assisting with data processing and analysis, I would argue that this offers a unique opportunity to drive research more towards novel questions that have not been studied and that have little relationship with prior work. That is, a new research idea is no longer primarily motivated by the proverbial gap in the literature, by the quote in the Wall Street Journal, or the serendipitous interview with the COO, but rather by in-depth engagement with actual operations.

This requires researchers to spend much more time identifying and conceptualizing novel research questions, and potentially making the work that is currently seen as "actual research work" (theorizing and analysis) more agentic to the researcher. Research that focuses primarily on readily accessible data could still be relevant, but the contribution bar can be substantially lifted, since the effort to perform the study will be considerably reduced. As conceptualizations and the "actual research work" becomes more disconnected, there is a risk that construct validity weakens. Hence, researchers need to explicitly address construct validity in a much more explicit way than is currently done. As many of the GenAI tools are opaque in terms of their algorithms, novel conceptions on how to argue for construct validity may need to be developed. In addition, researchers will be able to spend more of their time on the interpretation of the findings, relying on more abstract intuitions rather than more procedural reasoning.

It will require many of us to learn new skills and may cause drastic changes in PhD programs. Required skills have increasingly become method-dense, such as advanced econometrics or extensive theorizing. Such skills may be taken over by GenAI tools. This may be a good development if we then manage to get out into the real world that has not yet been captured by data, if we are able to truly develop novel conceptualizations, and can think more deeply about intuitive and conceptual interpretations of findings.

This development will aggravate issues of equal access. Many science disciplines have become increasingly concentrated at a limited number of well-funded schools in rich countries as they can

afford the expensive equipment. As top-notch GenAI tooling is becoming increasingly expensive (Wiggers, 2025), such research tools may become only accessible to a few institutions.

Commentary

Matthias Holweg (Matthias.Holweg@sbs.ox.ac.uk) and Benn Lawson (Benn.Lawson@sbs.ox.ac.uk), Saïd Business School, University of Oxford

Shalpegin et al. (2025)'s review of the use of generative AI (GenAI) in academic research identifies a set of issues of immediate concern, such as plagiarism and the falsification of data. These concerns – while valid – are not new, albeit greatly amplified by the widespread adoption of GenAI tools. This commentary argues that the wider implications of GenAI integration into academic research warrant greater consideration. Here, it is useful to consider Amara's Law that people tend to overestimate the impact of a technology in the short term while underestimating its long-term effects (Amara, 1974). In this sense, many of the current debates on the influence of GenAI are focal in nature focused on ethical failures in the production of individual research studies. Instead, the broader, long-term impacts may lie in the threat to the advancement of knowledge – a challenge that has not yet been fully addressed in scholarly discourse. To illustrate this threat, we outline three categories of failure modes – epistemic, methodological, and systemic – that, in our view, merit attention.

Epistemic Failure Modes

Multiple epistemic failure modes exist which can compromise the reliability and validity of knowledge. For one, the hallucination or fabrication of information, data and references is an artefact of any transformer-based architecture (Vaswani et al., 2017). This pollutes academic literature with misinformation, making it increasingly difficult to distinguish real findings from convincing AI-generated fabrications (Messeri and Crockett, 2024). Second, since GenAI models are trained on statistical patterns and lack causal reasoning, AI-generated research may mistakenly infer spurious causation from observed correlations, leading to misleading conclusions. AI systems are fundamentally unable to execute higher level causal reasoning, as defined by Judea Pearl's ladder of causation (Pearl, 2009; Pearl and Mackenzie, 2018). A further issue is model collapse, or degenerative feedback loops (Gibney, 2024; Shumailov et al., 2024). If AI-generated research enters training datasets, future AI models may rely on AI-generated content rather than original human research, leading to a self-referential system in which academic knowledge loses novelty and accuracy.

Methodological Failure Modes

Methodological failure modes affect the integrity of research methodologies. Because generative AI produces outputs stochastically, identical inputs may not always yield identical results. Simple alterations like changing the order of training data (Vinyals et al., 2015), using different models or even different versions of the same models (Mirzadeh et al., 2024), as well as adding redundant words in the prompt (Mirzadeh et al., 2024), all alter the output and compromise the reproducibility of research – potentially to the point of prompt instability (Stewart et al., 2024). This is particularly problematic for methodologies that require consistent and verifiable outcomes. This variability presents fundamental challenges for reproducibility, an essential pillar of academic research. Addressing this challenge will require new protocols, such as standardized prompt documentation and mechanisms to archive generated outputs. Merely submitting prompts, as some have suggested, is insufficient given the stochastic nature of GenAI output.

Furthermore, incomplete or non-transparent documentation of GenAI use poses a challenge. Researchers may fail to disclose the role GenAI played in their research process, preventing reviewers and readers from properly assessing the study's rigor or potential biases. Finally, while GenAI can summarize prior work efficiently, its focus on the "centre of the distribution" increases the likelihood that it will omit critical or contradictory studies. Using GenAI for literature reviews thus introduces the risk of biased or incomplete assessments, leading to an academic discourse that is built on an incomplete or misrepresented foundation (An et al., 2025).

Systemic Failure Modes

Systemic failure modes change the nature of academic research in unintended ways. One of the most significant, yet often overlooked, consequences of GenAI in academia is the potential for overproduction of research output. By streamlining processes such as literature reviews, hypothesis generation, data analysis, and manuscript writing, GenAI drastically lowers the barriers to publication (Khalifa and Albadawy, 2024). ChatGPT's own estimates suggest AI tools could reduce the effort required for academic research by 30–60%, effectively doubling output. While increased efficiency is beneficial, the overproduction of academic papers risks overwhelming peer review systems and diluting the overall quality of published research. GenAI's ability to facilitate the rapid generation of superficially convincing, but low-quality submissions, requires the development of new mechanisms to filter, evaluate, and prioritize truly valuable contributions. A related concern is the erosion of peer review quality. Gen AI-generated papers may be not only difficult for reviewers to evaluate, but the potential use of AI-generated reviews could weaken necessary critical analysis skills of scholars (Conroy, 2023). If the academic vetting process becomes compromised, trust in published research may diminish.

Conclusion

GenAI tools are now ubiquitous and attempts to police or prohibit their use are likely to be futile in the long run. Instead, as a field we should embrace these tools while maintaining clarity on their limitations. This may mean that the nature and format of academic journal submissions will evolve. Transparency should be the guiding principle, in accordance with the well-known adage that "sunlight is the best disinfectant." The format of submissions is likely to change too: Traditional, elaborate literature review sections may become obsolete and replaced by detailed declarations of technology use and access to underlying data for verification and reproduction. Furthermore, research conducted with AI assistance can only be properly assessed by human reviewers who understand its capabilities and limitations. We need protocols and guidelines for reviewers to help them in this endeavour.

Editorial policies must confront the broader threat that GenAI-aided or GenAI-generated research poses to the theoretical foundations of any academic field. GenAI systems are fundamentally incapable of developing and testing contrarian theories, which are essential for the systematic advancement of knowledge (Felin and Holweg, 2024). As AI-generated research increasingly becomes part of future training datasets, we risk theoretical stagnation. Without new and diverse inputs, GenAI systems may continuously regenerate variations of the same concepts, making it harder for academic fields to adopt truly novel thinking. The widespread use of AI in research thus presents a real and pressing danger to the ultimate goal of academic endeavours: The creation of genuinely new knowledge. Rather than allowing AI to define the boundaries of scholarship and potentially create "monocultures of knowing" (Messeri and Crockett, 2024), we must ensure that it serves as a tool for fostering true innovation that advances our knowledge.

References

Amara, R. (1974). The futures field: Functions, forms, and critical issues. Futures 6, 289-301

An, Y., Zhao, X., Yu, T., Tang, M., Wang, J. (2025). Systematic Outliers in Large Language Models. arXiv preprint arXiv:2502.06415

Angelopoulos, S., Bendoly, E., Fransoo, J., Hoberg, K., Ou, C., Tenhiälä, A. (2023). Digital transformation in operations management: Fundamental change through agency reversal. Journal of Operations Management, 69(6), 876-889.

Anthony, C., Bechky, B. A., & Fayard, A. L. (2023). "Collaborating" with AI: Taking a system view to explore the future of work. Organization Science, 34(5), 1672-1694.

Bail, C. A. (2024). Can Generative AI improve social science?. Proceedings of the National Academy of Sciences, 121(21), e2314021121.

Bendoly, E. Chandrasekaran, A., do Rego Ferreira Lima, M., Handfield, R., Khajavi, S., Roscoe, S. (2024). The role of generative design and additive manufacturing capabilities in developing human–AI symbiosis: Evidence from multiple case studies. Decision Sciences Journal, 55(4), 325-345.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). The crowdless future? Generative AI and creative problem-solving. Organization Science, 35(5), 1589-1607.

Bockting, C. L., Van Dis, E. A., Van Rooij, R., Zuidema, W., & Bollen, J. (2023). Living guidelines for generative AI—why scientists must oversee its use. Nature, 622(7984), 693-696.

Cao, Z., Li, M., & Pavlou, P. A. (2024). AI in business research. Decision sciences, 55(6), 518-532.

Chen, Y., Kirshner, S., Ovchinnikov, A., Andiappan, M., Jenkin, T. (2025). A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do?. Manufacturing & Service Operations Management. Forthcoming.

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanities and Social Sciences Communications, 10(1), 1-12.

Chen, Z., & Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. Management Science, *70*(12), 9101-9117.

Cohen, M., & Tang, C. (2024). The Role of AI in Developing Resilient Supply Chains. Georgetown Journal of International Affairs, Walsh School of Foreign Service.

Conroy, G. (2023). How ChatGPT and other AI tools could disrupt scientific publishing. Nature 622, 234-236

Davis, A. M., Mankad, S., Corbett, C. J., & Katok, E. (2024). OM Forum—The best of both worlds: Machine learning and behavioral science in operations management. Manufacturing & Service Operations Management, 26(5), 1605-1621.

De Kok, T. (2025). ChatGPT for Textual Analysis? How to use Generative LLMs in Accounting Research. Management Science. Forthcoming.

Fainmesser, I. P., Galeotti, A., & Momot, R. (2023). Digital privacy. Management Science, 69(6), 3157-3173.

Felin, T., Holweg, M. (2024). Theory is all you need: AI, human cognition, and causal reasoning. Strategy Science 9, 346-371

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277-304.

Foster, E. D. and Deardorff, A. (2017). Open science framework (OSF). Journal of the Medical Library Association : JMLA, 105(2):203–206.

Gibney, E. (2024). AI models fed AI-generated data quickly spew nonsense. Nature 632, 18-19

Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. (2023). Ideas are dimes a dozen: Large language models for idea generation in innovation. Available at SSRN 4526071.

Gyory, J. T., Soria Zurita, N. F., Martin, J., Balon, C., McComb, C., Kotovsky, K., & Cagan, J. (2022). Human versus artificial intelligence. Journal of Mechanical Design, 144(2), 021405.

Hu, M., Momot, R., & Wang, J. (2022). Privacy management in service systems. Manufacturing & Service

Operations Management, 24(5), 2761-2779

Huang, Z., Jiang, J., Xu, J., & Zhou, Y. (2024). Exploring the Use of Chatgpt for Generating Chinese Reading Materials in Eye-Tracking. Available at SSRN 4805567.

IBM. (2023). Shedding light on AI bias with real world examples. IBM Data and AI Team. https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/accessed on Sep 15, 2024.

Jackson, I., Ivanov, D., Dolgui, A., & Namdar, J. (2024). Generative artificial intelligence in supply chain and operations management. International Journal of Production Research, 1-26.

Jansen, P., Côté, M. A., Khot, T., Bransom, E., Mishra, B. D., Majumder, B. P., ... & Clark, P. (2024). DISCOVERYWORLD: A Virtual Environment for Developing and Evaluating Automated Scientific Discovery Agents. arXiv preprint arXiv:2406.06769.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. Academy of Management Journal, 67(1), 5-32.

JOM. (2024). Shared Editorial Policies on AI. URL https://www.jom-hub.com/ethical-guidelines accessed on Jan 6, 2025.

Khalifa, M., Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. Computer Methods and Programs in Biomedicine Update, 100145

Kim, H., & Kim, H. (2023). Deep embedding kernel mixture networks for conditional anomaly detection in high-dimensional data. International Journal of Production Research, 61(4), 1101-1113.

Lippert, S., Dreber, A., Johannesson, M., Tierney, W., Cyrus-Lai, W., Uhlmann, E. L., ... & Pfeiffer, T. (2024). Can large language models help predict results from a complex behavioural science study?. Royal Society Open Science, 11(9), 240682.

Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., & Ha, D. (2024). The AI scientist: Towards fully automated open-ended scientific discovery. arXiv preprint arXiv:2408.06292.

Luo, X., Rechardt, A., Sun, G., Nejad, K. K., Yáñez, F., Yilmaz, B., ... & Love, B. C. (2024). Large language models surpass human experts in predicting neuroscience results. Nature Human Behaviour, 1-11.

Mariani, M., & Dwivedi, Y. K. (2024). Generative artificial intelligence in innovation management: A preview of future research developments. Journal of Business Research, 175, 114542.

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. Nature, 627(8002), 49-58.

Mirzadeh, I., Alizadeh, K., Shahrokhi, H., Tuzel, O., Bengio, S., Farajtabar, M. (2024). Gsm-symbolic: Understanding the limitations of mathematical reasoning in large language models. arXiv preprint arXiv:2410.05229

Niu, Y., Wu, J., Jiang, S., & Jiang, Z. (2025). The bullwhip effect in servitized manufacturers. Management Science.

Pazzanese, C. (2020). Great promise but potential for peril. The Harvard Gazette.

Pearl, J. (2009). Causality. Cambridge University Press

Pearl, J., Mackenzie, D. (2018). The book of why: the new science of cause and effect. Basic Books

Rossi, S., Rossi, M., Mukkamala, R., Thatcher, J., & Dwivedi, Y. (2024). Augmenting research methods with foundation models and generative AI. International Journal of Information Management, 102749.

Service, R. F. (2025). The AI revolution comes to protein sequencing. Science.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., Gal, Y. (2024). AI models collapse when trained on recursively generated data. Nature 631, 755-759

Stewart, I., Horawalavithana, S., Kennedy, B., Munikoti, S., Pazdernik, K. (2024). Surprisingly Fragile: assessing and addressing prompt instability in multimodal foundation models. arXiv preprint arXiv:2408.14595

Taylor & Francis. (2025). AI Policy. URL https://taylorandfrancis.com/our-policies/ai-policy/ accessed on Jan 10, 2025.

The Economist. (2023, September 13). How scientists are using artificial intelligence. URL https://www.economist.com/science-and-technology/2023/09/13/how-scientists-are-using-artificial-intelligence accessed on Jan 24, 2025.

Thorp, H. H., & Vinsen, V. (2023). Change to policy on the use of generative AI and large language models. *Science*.

van Breugel, B., Liu, T., Oglic, D., & van der Schaar, M. (2024). Synthetic data in biomedicine via generative artificial intelligence. Nature Reviews Bioengineering, 1-14.

Van Noorden, R., & Perkel, J. M. (2023). AI and science: what 1,600 researchers think. Nature, 621(7980), 672-675.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems 30

Vinyals, O., Bengio, S., Kudlur, M. (2015). Order matters: Sequence to sequence for sets. arXiv preprint arXiv:1511.06391

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097-1126.

Wiggers, K. (2025). Openai reportedly plans to charge up to $20,000 a month for specialized ai 'agents'.

Wiley. (2025). Best Practice Guidelines on Research Integrity and Publishing Ethics. URL https://authorservices.wiley.com/ethics-guidelines/index.html accessed on Jan 14, 2025.

Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. PNAS nexus, 3(3), 1-8.

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science?. Computational Linguistics, 50(1), 237-291.