🤖 Resistance Is Futile: It's Time to Rethink Generative AI Usage Policies in the SBES Research Community

Nabor C. Mendonça
Programa de Pós-Graduação em Informática Aplicada
Universidade de Fortaleza
Fortaleza, CE, Brazil
nabor@unifor.br

Abstract

Generative artificial intelligence (AI) is reshaping both software engineering practice and the scholarly processes through which software engineering research is conceived, written, reviewed, and disseminated. These tools create real risks for scientific publishing, but they also expose the weakness of policies that treat AI-generated text as the main problem. In this paper, we use the 40-year milestone of SBES as an opportunity to discuss how software engineering researchers should regulate generative AI use in scholarly work. We argue that the SBES research community should move from prohibition-centered rules to accountability-centered governance. Using the evolution of SBES generative AI usage policies from 2023 to 2026 as a motivating case, we show how current formulations combine reasonable principles, such as human authorship, disclosure, and author responsibility, with fragile textual-origin restrictions, such as prohibiting articles or sections entirely produced by generative AI. We argue that these restrictions are difficult to define, hard to enforce fairly, and poorly prepared for a near future in which AI assistance becomes increasingly embedded in scholarly workflows and peer review at scale. As an alternative, we propose a layered model that classifies AI use according to its role and epistemic impact, ranging from low-risk mechanical assistance to epistemically significant support and unacceptable delegation of scientific judgment. We also discuss AI-assisted peer review as a harder case involving confidentiality, data governance, infrastructure access, scalability, and reviewer accountability. Our central claim is that AI resistance is not an effective policy: for a community grounded in process, automation, validation, and quality assurance, responsible AI use should be something the community can disclose, inspect, teach, and regulate.

Keywords

Generative AI, research integrity, scientific authorship, peer review, software engineering research

1 Introduction

The Brazilian Symposium on Software Engineering (SBES) has long served as a central venue for the Brazilian software engineering research community. Previous retrospective studies have analyzed the history, publication patterns, internationalization, and thematic evolution of SBES, documenting its role in consolidating software engineering research in Brazil [8, 13, 27]. Over four decades, this community has followed, discussed, and contributed to major transformations in the discipline, including empirical software engineering, software architecture, open-source ecosystems, DevOps, cloud-native systems, and AI-assisted software development.

Generative artificial intelligence (AI), particularly large language models (LLMs), has become a new object of software engineering research [10, 20] and a development tool [9]. It is also entering the research process itself, with recent systems supporting activities such as hypothesis generation, literature-based reasoning, experimental planning, data analysis, and manuscript production [15, 16, 28]. Generative AI is therefore reshaping the practices through which research is conceived, conducted, written, reviewed, and disseminated [5, 11, 29].

This trend can impact scientific research in two ways. First, AI can affect the quality of individual manuscripts by shaping how claims, evidence, and arguments are expressed. Second, the scale of scholarly communication is changing: in high-growth areas such as artificial intelligence, rapidly increasing submission volumes are placing unprecedented pressure on peer review [30]. Similar tensions have started to emerge in the software engineering community, where a recent survey reports high review load, uneven review quality, and initial disagreement over the role of LLMs in reviewing [4].

LLMs and related generative AI tools can produce fluent but incorrect text, fabricate references, obscure weak contributions behind polished prose, and challenge traditional assumptions about authorship, originality, peer review, and scientific accountability [1, 40]. Research communities are therefore right to define explicit policies regulating the use of generative AI in scientific publishing [12, 43]. The central policy-design problem is to formulate rules that are conceptually coherent, enforceable in practice, and aligned with the values they claim to protect [28].

In this paper, we use the 40-year milestone of SBES as an opportunity to examine a policy question that is likely to shape the community's immediate future: how should software engineering researchers govern generative AI use in writing, reviewing, and publishing scientific work? We argue that the SBES research community should move from its current prohibition-centered generative AI policies to accountability-centered governance that recognizes AI assistance as a potential component of sustainable scholarly work while establishing appropriate controls for its risks. We use the evolution of SBES generative AI usage policies from 2023 to 2026 as a situated case of how the community is responding to this challenge.

These policies include provisions that are broadly aligned with emerging publication norms, such as prohibiting generative AI tools from being listed as authors, requiring disclosure of some forms of AI assistance, and preserving human responsibility for submitted content. However, since 2024 they have also prohibited articles or sections entirely produced by generative AI tools. We argue that this textual-origin criterion is the clearest weakness of the SBES policies and provide limited preparation for a medium-term scenario in which generative AI becomes deeply embedded in scholarly workflows. As AI tools become more capable and submission volumes continue to grow, governance must address how to make AI assistance responsible, equitable, verifiable, and compatible with human scholarly judgment.

As a discipline, software engineering studies how tools, processes, automation, quality assurance mechanisms, traceability practices, and human oversight shape the production of complex artifacts. A research community grounded in these ideas should be especially well positioned to reason about generative AI as part of a scholarly production process. Policies that focus primarily on prohibiting AI-generated textual units risk treating generative AI as a source of textual contamination. Its use as a sociotechnical tool requires governance through process, transparency, validation, and accountability.

Our central claim is that AI resistance is not an effective policy. Generative AI will become part of scientific writing and reviewing regardless of prohibitions in calls for papers. Scholarly publishing will become more sustainable only through policies that govern AI assistance as part of routine scholarly practice. The challenge for the SBES research community is therefore to govern AI use in ways that preserve scientific integrity while acknowledging how scholarly work is changing.

This paper makes three contributions to the SBES 40 Years discussion on the perspectives and challenges of using artificial intelligence in software engineering. First, it provides a situated retrospective analysis of how SBES generative AI usage policies evolved from 2023 to 2026, showing how the community has moved from provisional restrictions toward a more enforceable regulatory stance. Second, it uses this evolution to identify a prospective challenge for AI-assisted software engineering research: textual-origin rules are difficult to define and enforce, and they provide limited guidance for a future in which AI becomes part of ordinary scholarly writing, artifact production, data analysis, and peer review. Third, it proposes an accountability-centered AI usage policy framework for the SBES research community, including a layered model that classifies AI use according to its role and epistemic impact, ranging from low-risk mechanical assistance to epistemically significant support and unacceptable delegation of scientific judgment.

This paper presents the author's position on AI governance and does not empirically validate specific governance mechanisms. Its objective is to articulate a coherent policy framework, grounded in recent literature and current publication policies, that can inform discussion and future refinement by the SBES community. The SBES 40-year milestone provides an appropriate occasion for this discussion because it concerns how the community chooses to govern one of the most significant methodological transformations likely to shape software engineering research over the coming decade.

2 Related Work

Research on generative AI in scholarly publications has expanded rapidly since the public release of large language model-based tools. This section discusses how the impact of generative AI on scholarly writing and peer review has been reported in the literature. It also situates this discussion within the broader policy landscape adopted by a representative sample of research venues and publication organizations

2.1 AI in Scholarly Writing and Peer Review

Recent work has examined how generative AI is reshaping scholarly writing, authorship, publication policy, and peer review. Empirical studies show that publishers and journals have rapidly introduced guidelines, but with substantial heterogeneity. Ganjavi *et al.* [12] analyzed the top 100 largest academic publishers and top 100 highly ranked journals, finding broad convergence against granting authorship to generative AI systems but considerable variation in allowed uses and disclosure requirements. Bhavsar *et al.* [3] similarly audited policies among scientific, technical, and medical publishers, reporting that only a minority had public policies at the time of analysis, although most policies that existed required disclosure and none openly allowed AI chatbot authorship. Yin *et al.* [43] extended this line of work to medical journals, considering both author and reviewer guidelines and showing that some recommendations, including use for data analysis and interpretation, remain underspecified.

A second group of studies has proposed ethical principles and governance frameworks for responsible use of generative AI in scientific practice. Porsdam Mann *et al.* [26] propose three criteria for ethical LLM-assisted writing: human vetting and guaranteeing, substantial human contribution, and acknowledgement and transparency. Their proposal is particularly relevant to our argument because it avoids treating AI use as intrinsically illegitimate and instead emphasizes human responsibility and disclosure at a level sufficient to judge credibility, responsibility, and reproducibility. Helmy *et al.* [19] offer practical rules for careful use of generative AI in science, emphasizing both opportunities for research workflows and risks such as hallucinations, plagiarism, bias, overreliance, and loss of research skills. Watson *et al.* [41] provide a more theoretical account of generative AI and scientific authorship, arguing against simplistic distinctions between technology, individuals, and scholarly practice.

A separate line of work focuses on AI-assisted peer review. Li *et al.* [24] report that many top medical journals prohibit or limit AI use in peer review, with confidentiality as the dominant rationale. Within empirical software engineering, Bogner and Verdecchia [4] provide a complementary community-level view by surveying reviewers about review load, perceived review quality, current LLM use, and possible improvements to the review process. Their study is particularly relevant here because it shows a community divided between forbidding LLMs and exploring their responsible use, while expressing stronger support for sanctions against clearly unethical LLM use in reviews. Recent work outside software engineering complements this picture from two directions: Wei *et al.* [42] argue that AI-assisted review should become a research and infrastructure priority for scaling machine-learning peer review, whereas

Table 1: Illustrative mapping of generative AI usage policies in selected research venues and organizations.

Venue / organization	Main observed policy orientation	Preliminary classification
SAST / SBCARS / CBSoft-related calls	Adopt formulations broadly similar to SBES, combining disclosure requirements with restrictions on fully AI-produced textual units.	Disclosure plus textual-origin prohibition
BRACIS 2026	States that generative AI models do not satisfy authorship criteria; places full responsibility on human authors for any AI-assisted writing, including correctness and plagiarism.	Accountability-centered
FSE / ASE / MSR	Follow ACM-style guidance: generative AI tools may not be authors; use of generative AI to create content is permitted but must be disclosed.	Disclosure and accountability
ICSE / SANER	Refer to ACM/IEEE-style publication and authorship policies, emphasizing human authorship, disclosure, and author responsibility.	Policy by reference to ACM/IEEE
ICML	Allows use of generative AI tools, including LLMs, to assist in writing or research; excludes LLMs from authorship; assigns full responsibility to human authors.	Permissive with accountability
AAAI / AIES	Exclude generative AI systems from authorship and from being treated as citable sources; emphasize author responsibility for truthfulness, originality, and plagiarism.	Accountability plus authorship/source restriction
ACL / EMNLP / ARR	Require or encourage disclosure of generative assistance, often through structured submission forms or checklists; emphasize authorship and integrity.	Structured disclosure
ACM / IEEE / SBC	Converge on human authorship and responsibility, while differing in the level of detail required for disclosing AI-generated content.	Institutional disclosure and accountability

Gartenberg *et al.* [14] provide journal-level evidence that AI use, under existing publication incentives, may increase submission volume while degrading writing and review quality. At the same time, recent empirical work on AI reviewers suggests that AI systems can provide useful complementary scrutiny while remaining unsuitable as replacements for human reviewers. Kim *et al.* [22] evaluate review items from human and AI reviewers of Nature-family papers and show that AI reviewers can raise correct, significant, and well-evidenced criticisms, but also produce more incorrect items, show weaker calibration to subfield norms, and overlap with one another more than human reviewers do.

Collectively, these studies shift the trend from blanket permission or prohibition to the effects of specific uses on authorship, accountability, disclosure, verifiability, confidentiality, and editorial judgment [28]. This motivates our focus on publication policies as process artifacts that define how authors, reviewers, and committees are expected to act under conditions of increasing AI mediation.

2.2 AI Policies Across Research Venues

To complement the previous literature review, we also examined how generative AI usage is currently addressed in selected conference calls and publication policies in software engineering, artificial intelligence, and Brazilian computing venues. Our goal was not to provide an exhaustive survey, but to situate the discussion within a broader landscape of responses adopted by related research communities. For each policy, we considered six aspects: AI authorship, disclosure, the distinction between mechanical editing and substantive generation, textual-unit prohibitions, human responsibility, and treatment of AI use in research methods or peer review.

Table 1 summarizes the main patterns observed in this focused mapping. Across the analyzed venues, three points of convergence are visible. First, there is broad agreement that generative AI systems should not be listed as authors. Second, most policies assign responsibility to human authors for the correctness, originality, integrity, and scholarly value of the submitted work. Third, many venues require or encourage disclosure of relevant generative AI use, especially when AI tools contribute to the creation of textual, visual, computational, or analytical content. The main divergence concerns how policies treat AI-generated content as a textual artifact.

Several major software engineering venues that follow ACM-style guidance allow the use of generative AI tools to create content, provided that such use is disclosed. In this model, the central requirement is not that every sentence originate from a human author, but that authors disclose relevant uses and remain responsible for the work. Similarly, some AI and machine learning venues explicitly allow authors to use generative AI tools to assist in writing or research, while emphasizing full authorial responsibility and excluding LLMs from authorship. These policies are not permissive in the sense of ignoring risks; rather, they regulate AI use through disclosure and accountability.

Brazilian venues policies also show variation. The Code of Conduct of the Brazilian Computer Society (SBC), approved in 2024, adopts a disclosure-and-accountability formulation: authors must declare the use of generative AI tools for content generation, writing, or revision, identify the tools used, describe where they were employed, and remain responsible for the entire content [37]. The Brazilian Conference on Intelligent Systems (BRACIS) similarly emphasizes human responsibility and excludes generative AI tools from authorship. Some calls associated with the Brazilian Conference on Software: Theory and Practice (CBSoft), however, combine disclosure requirements with restrictions on fully AI-produced textual units. This variation is important because it shows that even within related institutional contexts, policies may differ in how they balance disclosure, accountability, and textual-origin restrictions.

This paper builds on the literature and policy landscape summarized above, but differs in scope and argument. Rather than surveying journal policies broadly or proposing a discipline-neutral guideline, we analyze generative AI policies in the SBES research community as a situated case. The next section therefore turns to the recent evolution of SBES policies.

Table 2: Evolution of selected generative AI policy features in recent SBES calls. N/I = not identified in the call.

	SBES call			
Policy feature	**2023**	**2024**	**2025**	**2026**
AI tools cannot be authors	Yes	Yes	Yes	Yes
AI-assisted editing is allowed	Yes	Yes	Yes	Yes
AI-generated parts allowed with acknowledgment	N/I	Yes	Yes	Yes
AI-produced articles or sections prohibited	N/I	Yes	Yes	Yes
"Parts" distinguished from "sections"	No	Yes	Yes	Yes
Author responsibility emphasized	Implicit	Yes	Yes	Yes
Tool/use-location declaration required	N/I	Yes	Yes	Yes
Desk rejection for non-compliant AI use	N/I	N/I	N/I	Yes
Future-submission restriction possible	N/I	N/I	N/I	Yes

3 SBES Generative AI Policy Evolution

Recent SBES calls for papers offer a concrete case of how the Brazilian software engineering research community is reacting to AI-assisted scholarship. In addition to manuscript-preparation rules, they define what the community currently treats as legitimate scholarly work under AI mediation.

Table 2 summarizes how selected policy features have evolved in recent SBES calls for papers. In 2023, the SBES Research Track adopted a provisional policy inspired by Elsevier and ICML, prohibiting text or images entirely produced using AI-assisted technologies and prohibiting AI tools from being listed as authors, while allowing AI-assisted editing, image improvement, and research on the use of AI to support software engineering activities [32]. From 2024 onward, SBES calls explicitly reference the SBC Code of Conduct, whose formulation is centered on disclosure, human authorship, and author responsibility [37]. However, the SBES calls also adopt a more specific and restrictive formulation: they allow AI-generated parts of the content with explicit acknowledgment and AI-assisted editing of existing text without disclosure, but *prohibit articles or sections entirely produced by generative AI* [33–36]. The 2026 call further states that evidence of non-compliant generative AI use may lead to *rejection without review* and may *prevent authors from submitting to future SBES editions* [36]. This trajectory suggests that the SBES policy has moved from a provisional response to a repeated and increasingly enforceable regulatory stance, while adding a textual-origin restriction that goes beyond the disclosure-and-accountability logic of the SBC Code of Conduct.

Notably, SBES does not simply reject generative AI use. Its recent calls recognize that such use is heterogeneous: correcting grammar, improving readability, polishing images, drafting parts of the text, generating methodological claims, and fabricating results are not equivalent practices. In this respect, the policy is partly aligned with an accountability-centered view: AI tools cannot be authors, human authors remain responsible, and some forms of AI assistance are permitted when disclosed. The problematic step is the additional attempt to define acceptability by the textual boundary of the generated material.

The central ambiguity appears in the distinction between AI-generated "parts" of the content and AI-generated "sections". The policy permits the former, with acknowledgment, but prohibits the latter. This raises a practical and conceptual question: what counts as a "part" of the content? A sentence? A paragraph? A table? A figure caption? A subsection? If a short section contains a single paragraph, would that paragraph be an acceptable "part" of the content or an unacceptable AI-produced section? Conversely, if several AI-generated paragraphs are distributed across different sections, would this be more acceptable than a single AI-generated paragraph that happens to constitute a complete short section?

The difficulty is that sections are rhetorical and editorial units, not stable units of intellectual contribution. Their size and function vary widely across papers and genres. A section may contain a complete methodological argument, a brief transitional paragraph, a short statement of threats to validity, or a purely descriptive summary of an artifact. Regulating AI use by section boundaries therefore risks making the acceptability of a practice depend on document structure rather than on scientific substance.

This is why the recent SBES policies are revealing. Their strongest elements are those that connect directly to scientific integrity: human authorship, responsibility, disclosure, and validation. Their weakest element is the attempt to regulate the provenance of textual units. A paper assembled from unverified AI-generated text, fabricated references, invented results, or arguments that the authors themselves cannot explain should clearly be treated as a serious violation of scientific standards. But these violations are better characterized as failures of accountability, verification, transparency, and scholarly competence than as failures of textual origin.

SBES is right to worry about AI-assisted manuscripts. The issue is that the current policy combines a defensible accountability requirement with a fragile textual-origin rule and, more broadly, remains oriented toward controlling current forms of AI-assisted writing rather than preparing for increasingly AI-mediated scholarly workflows. Policies that emphasize human responsibility, disclosure, validation, and verifiability can be made precise, teachable, and enforceable. Policies that prohibit "AI-produced sections" are harder to define, harder to audit, and more likely to generate uncertainty about what authors may responsibly do. They also provide limited guidance for a near future in which AI assistance may be routinely involved in literature work, data analysis, artifact inspection, manuscript preparation, and review support.

4 Why Prohibition-Centered Policies Are Bad

The discussion above suggests that the most visible fragility of the SBES policy is its reliance on a prohibition against articles or sections entirely produced by generative AI. This section develops that critique, but also argues that the problem is broader. Prohibition-centered policies tend to mischaracterize the challenge in four ways: they confuse textual origin with intellectual authorship, rely on unstable units of regulation, create incentives for concealment, and fail to anticipate scholarly workflows in which AI assistance may become increasingly embedded in writing, reviewing, and managing scientific communication at scale.

4.1 Textual Origin Is Not Authorship

First, textual origin is a poor proxy for scientific legitimacy. This assumption is weak. Scientific authorship is much more than the mere the act of typing sentences. It involves formulating research

questions, designing methods, selecting evidence, interpreting results, positioning claims in relation to prior work, responding to criticism, and taking responsibility for the content of the paper. A generative AI system may produce text, but it cannot assume responsibility for these intellectual and social obligations.

This distinction is important because generative AI can participate in the production of text without being the author of the scientific contribution. An author may ask an AI tool to rewrite a paragraph for clarity, generate a first draft from detailed notes, suggest alternative organizations for an argument, or produce a summary that is later checked, corrected, and incorporated into the paper. In these cases, knowing whether the human authors understand, validate, and endorse the resulting claims is more relevant than identifying whether the initial wording came from a human or from a machine.

The reverse is also true. Human-produced text is not automatically epistemically legitimate simply because it was written without AI assistance. A paragraph written entirely by a human author may contain unsupported assertions, misleading interpretations, incorrect references, plagiarized content, or conclusions that do not follow from the evidence. If the goal of a publication policy is to protect scientific integrity, then we argue that the more important issue is the quality, originality, validity, and accountability of the scholarly work.

For this reason, textual-origin rules risk targeting a superficial property of the manuscript. They focus on how text came into being rather than on whether the text is scientifically defensible. A policy centered on human accountability is more directly connected to research integrity: it asks who is responsible for the claims, who can justify the evidence, who can correct errors, and who can respond to criticism. These questions are more meaningful than asking whether a section was produced by a particular sequence of keystrokes, prompts, edits, and revisions.

4.2 "Entirely Produced" Is Hard to Define

The second problem is the instability of the category "entirely produced". At first glance, the expression appears to define a strict boundary. In practice, it raises more questions than it answers. If an AI system generates a section from a detailed outline written by the authors, is the section entirely produced by AI? If the authors subsequently revise the text, add references, remove claims, and rewrite several sentences, at what point does the section cease to be AI-produced? If a section is written by humans but translated, reorganized, and stylistically improved by AI, is it still entirely human-produced?

The problem becomes sharper when the policy distinguishes between using AI to produce "parts of the content" and using AI to produce an entire section. The notion of a "part" is itself undefined. It may refer to a sentence, a paragraph, a table, a figure caption, a subsection, or a conceptual fragment of the argument. This makes the boundary between acceptable and unacceptable use depend on arbitrary textual segmentation. A one-paragraph section generated with AI assistance may be prohibited because it is a complete section, whereas a longer multi-paragraph section containing several AI-generated paragraphs may be acceptable if those paragraphs are treated as parts of a larger section. Such a distinction is difficult to justify in terms of scientific integrity.

Sections are rhetorical and editorial units, not stable units of intellectual contribution. Their length and function vary across papers, venues, and writing styles. One section may contain a complete methodological contribution; another may contain a short transitional paragraph; another may summarize related work; another may only introduce the structure of the paper. Regulating AI use at the level of section boundaries therefore risks making the acceptability of a practice depend on document organization rather than on the epistemic role of the content.

A more robust policy would avoid treating "entirely produced" as the key category. It would instead ask whether the use of generative AI had epistemic significance. Did it contribute to claims, evidence, analysis, interpretation, literature synthesis, or methodological decisions? If so, disclosure and traceability become important. Did it merely improve grammar or readability? If so, the relevance to scientific integrity is lower. This distinction is not perfect, but it maps more closely to the reasons why the community cares about AI use in the first place.

4.3 Prohibition Encourages Concealment

Third, broad or ambiguous prohibitions incentivize concealment over transparency. Authors may become less willing to disclose AI use, even when that use is legitimate or harmless. A policy that treats certain forms of AI-generated text as inherently disqualifying creates uncertainty: authors who used AI in a substantial but responsible way may worry that disclosure will be interpreted as evidence of misconduct. The result may be less transparency, not more.

This is especially problematic because the use of generative AI is difficult to verify externally. Current detectors of AI-generated text are unreliable in practical settings, and their performance can degrade substantially when text is paraphrased, revised, translated, edited, or otherwise transformed [23, 31]. Stylistic suspicion is also a poor basis for enforcement: polished prose, generic phrasing, or formulaic structure may result from many causes, including language editing, collaborative revision, disciplinary convention, cautious writing, or the writing patterns of less experienced authors. Prior work has also shown that AI-text detectors may produce unfair false positives for particular groups of writers, raising concerns about using such tools as evidence of misconduct [25]. A policy that depends on detecting whether a section was AI-produced risks encouraging subjective and uneven enforcement.

The concealment problem is particularly acute for inexperienced scientific writers, including undergraduate students, master's students, and early-stage doctoral researchers. In the SBES context, this issue is not primarily about non-native English writing, since the conference is Brazilian and accepts submissions in Portuguese. Rather, it concerns the uneven distribution of scientific writing expertise. Early-career authors may use generative AI to improve organization, clarity, argument flow, terminology, or adherence to academic style, not necessarily to outsource intellectual work. Yet these are precisely the authors who may have the greatest difficulty interpreting vague boundaries such as "parts of the content" or "entire sections". If policies are ambiguous or punitive, they

may discourage legitimate learning-oriented uses of AI, create fear around disclosure, and disproportionately affect authors who are still developing scholarly writing skills.

More fundamentally, prohibition-centered policies often frame the risk too narrowly. The central danger is that generative AI may introduce or amplify epistemic failures: fabricated references, unsupported claims, shallow literature synthesis, invalid comparisons, incorrect code, inappropriate statistical analyses, misleading interpretations, and plausible but false explanations. These failures can occur even when AI use is limited to small fragments of text, while a longer AI-assisted draft may be scientifically acceptable if the authors carefully verify every claim, check references, and ensure that the argument accurately reflects the evidence.

A policy designed around epistemic risk would therefore ask different questions: Did AI contribute to claims, evidence, analysis, code, or interpretation? Were outputs checked against primary sources? Were scripts or validation procedures preserved when the use affected the research process? These questions are more demanding than asking whether a section was AI-generated, but they are also more closely aligned with the goals of scientific publishing.

A policy centered on accountability should also distinguish noncompliance from misconduct. Not every failure to disclose AI assistance has the same severity, and not every AI-assisted passage indicates lack of scholarly responsibility. Strong sanctions are more defensible when there is verifiable evidence that authors failed to check AI-generated outputs with scientific consequences, such as hallucinated references, residual model instructions, fabricated data, or claims that the authors cannot substantiate.

Recent debates around arXiv's decision to sanction submissions containing clear evidence of unchecked LLM generation illustrate this distinction. Under the new practice, submissions with hallucinated references or other incontrovertible signs of unverified AI-generated content may lead to a one-year ban, followed by additional restrictions on future submissions [6, 18]. The measure has been welcomed by some research-integrity observers as a necessary response to low-quality AI-generated submissions, but also criticized as difficult to enforce fairly and at scale [17]. The lesson for SBES is that enforcement should be evidence-based, proportional, and appealable: strict where there is clear scholarly misconduct, but not punitive toward responsible and disclosed AI assistance.

4.4 Prohibition Does Not Scale

A final limitation is that prohibition-centered policies are poorly aligned with the likely evolution of scholarly work. They are often written as if generative AI were a bounded writing aid whose main risk is the production of suspicious textual fragments. This may describe part of the current problem, but it does not describe the emerging trajectory. Generative AI tools are increasingly integrated into literature search, summarization, coding, data analysis, artifact inspection, writing environments, and review workflows [15, 16, 28]. At the same time, rapidly growing submission volumes are placing pressure on peer review systems that already depend on scarce and unevenly distributed human expertise [30].

This does not mean that AI use should be encouraged without limits. Rather, policies should be future-facing: they should define acceptable roles for AI assistance, specify validation obligations, provide disclosure mechanisms, govern data and confidentiality conditions, and preserve human responsibility for scientific judgment. A list of prohibitions may look safe today, but it gives authors and reviewers little guidance once AI becomes part of ordinary writing, analysis, and review infrastructure.

5 Toward Accountability-Centered Policies

Rejecting textual-origin prohibitions does not mean abandoning regulation. It means regulating AI according to the role it plays in producing, validating, and assessing scientific knowledge. We therefore propose that the SBES research community move toward accountability-centered generative AI policies. Such policies should ask what the AI system was used for, how its outputs affected the scholarly contribution, how the authors validated those outputs, what level of disclosure or verifiability is appropriate, and how responsible AI assistance can be integrated into scalable scholarly workflows.

5.1 Software Engineering as a Policy Lens

The software engineering community should be especially skeptical of rules focused only on textual provenance. The field has long studied tool-mediated work, process design, automation, quality assurance, traceability, verification, validation, and human oversight [2, 7, 44]. Publication policies are themselves process artifacts: they define roles, responsibilities, acceptable practices, reporting obligations, and enforcement mechanisms.

Software engineering also provides a useful lesson about automation. Static analysis [39], code review [2], and continuous integration [44] support human work without eliminating the need for interpretation, validation, and accountability. Similarly, AI-assisted scholar work should be evaluated by asking whether claims are traceable to evidence, references and analytical outputs were checked, and the authors can defend the contribution. A mature policy should therefore distinguish low-risk assistance, uses requiring disclosure or supporting artifacts, and unacceptable delegation of scientific responsibility.

5.2 Principles for Accountability-Centered Governance

Building on the policy patterns and concerns discussed in the previous sections, we propose four principles for accountability-centered generative AI governance in the production of software engineering research papers.

Principle 1: Human accountability over textual origin. Authors should be responsible for every claim, reference, argument, method, result, figure, table, and conclusion in the paper, regardless of how the corresponding text was produced. The focus moves from where the text came from to who stands behind the claim. AI-assisted drafting is not inherently illegitimate; what would be illegitimate is submitting text that the authors did not understand, verify, or endorse. Similarly, manually written text should not be assumed to be legitimate solely because of its human origin. The relevant question should be whether the authors can explain and defend the content as part of their contribution. This principle is consistent with policies and guidelines that reject AI authorship while assigning responsibility for AI-assisted content to human

authors [3, 12, 26, 37], and with the broader research-integrity principle that scientific validity depends on the quality and verifiability of scholarly claims rather than on their mode of production.

Principle 2: Proportional disclosure. Not every use of generative AI requires the same level of reporting. A policy that requires detailed disclosure for every grammar correction or stylistic suggestion would be burdensome and likely ignored. Conversely, a policy that treats substantial AI-assisted synthesis, analysis, or argument generation as equivalent to spell checking would be too weak. Disclosure requirements should therefore be proportional to the epistemic impact of the use, reflecting the broader movement toward disclosure while avoiding the assumption that all AI assistance has the same relevance for research integrity [3, 12, 43]. Low-impact uses, such as grammar correction, spelling, formatting, or minor phrasing improvements, may not require explicit disclosure. Medium-impact uses, such as substantial rewriting, translation, restructuring of paragraphs, or generation of examples, may require a general declaration. High-impact uses, such as literature synthesis, data analysis, code generation, qualitative coding, methodological critique, or interpretation of results, should require more specific disclosure.

Principle 3: Verifiability for epistemically significant uses. When generative AI is used in ways that may affect the evidence, analysis, implementation, interpretation, or claims of a paper, authors should provide enough information for readers and reviewers to verify the resulting scholarly content. This principle responds to well-documented risks of hallucinated references, fabricated citations, and unverified AI outputs, while treating transparency as verifiability of scholarly outcomes rather than exhaustive logging of AI interactions [1, 19, 26, 40]. The point is not to archive every AI prompt. Such interactions are often exploratory, non-linear, and impractical to document exhaustively. The point is to make AI-affected scholarly outputs checkable through conventional artifacts and procedures: checked references, available scripts, documented analysis decisions, reproducible results, inspected code, or explicit human adjudication.

For example, if AI is used to help write an analysis script, the important artifact is not the prompt that produced an initial version of the script, but the final script, its validation, and its consistency with the reported method. If AI is used to support qualitative coding, the relevant evidence is the coding protocol, human review process, disagreement resolution, and examples of coded data. If AI is used to summarize prior work, the authors should verify the summary against primary sources and cite those sources directly. In all cases, the policy should require verifiability of the scholarly outcome, not exhaustive reconstruction of the AI interaction.

Principle 4: No delegation of scientific judgment. Generative AI may assist with drafting, checking, summarizing, reorganizing, or critiquing, but it should not replace human judgment about contribution, validity, interpretation, novelty, or significance [22, 24, 26]. Authors may use AI to identify possible threats to validity, but they must decide which threats are real and how they affect the study. They may ask AI to suggest related work, but they must verify relevance and accuracy. They may use AI to generate code or analysis scripts, but they must inspect, test, and validate those artifacts. AI outputs become part of scientific practice only after human evaluation.

5.3 A Layered Model of Generative AI Use

To make these principles operational, we propose the layered model of generative AI use in scientific writing shown in Figure 1. The model classifies AI assistance according to its role in the scholarly process and its potential epistemic impact. Its central axis is the extent to which the AI-assisted output can affect the paper's claims, evidence, methods, analysis, artifacts, or interpretation. As epistemic impact increases, governance requirements also increase: from no mandatory disclosure for low-impact assistance, to disclosure and verifiability for uses that affect the scientific contribution, to prohibition or sanctions when AI is used to fabricate, plagiarize, or delegate scientific judgment.

At Level 0, AI is used for mechanical assistance, such as spelling correction, grammar correction, formatting, or minor phrasing improvements. These uses normally do not affect meaning, evidence, or interpretation and therefore should not require mandatory disclosure. Level 1 covers language and presentation support, such as translation, paragraph-level rewriting, terminology improvement, or visual presentation. These uses may substantially affect readability and rhetorical form, but they do not necessarily affect the scientific contribution; a general disclosure may be appropriate when their use is substantial. Level 2 covers AI-assisted drafting and structuring, such as drafting passages from author-provided notes, generating examples, suggesting section structure, or preparing initial summaries for author revision. These uses are more substantial because they shape parts of the manuscript, but they remain acceptable when authors select, revise, validate, and take responsibility for the resulting text.

Level 3 covers epistemically significant assistance. This includes uses such as literature synthesis, code generation, statistical analysis, qualitative coding, data extraction, methodological critique, interpretation support, or artifact inspection. These uses may directly affect the evidence, analysis, or claims of the paper, and therefore require stronger governance: detailed disclosure and verifiability through appropriate scholarly artifacts or validation procedures. Finally, Level 4 covers unacceptable delegation or misconduct, including fabricated references, invented results, unverified AI-generated claims, plagiarism, submission of content authors cannot explain, or delegation of scientific judgment to AI. These cases should be prohibited and may justify rejection or sanctions, depending on severity.

The layered model should be viewed as an initial conceptual framework rather than a definitive taxonomy. Its primary purpose is to provide a vocabulary for discussing governance choices. Future experience and community feedback may motivate refinement of the proposed levels and corresponding governance recommendations.

By replacing a binary allowed/prohibited distinction with levels of use, the model gives SBES authors, reviewers, and chairs not only a conceptual basis but also a teachable vocabulary for discussing and governing AI assistance in scholarly production. This paper itself illustrates the need for such a distinction: it was prepared for a venue governed by the SBES policy being analyzed, while its acknowledgments section demonstrates how generative AI use can be classified and disclosed under the new layered governance model proposed here.

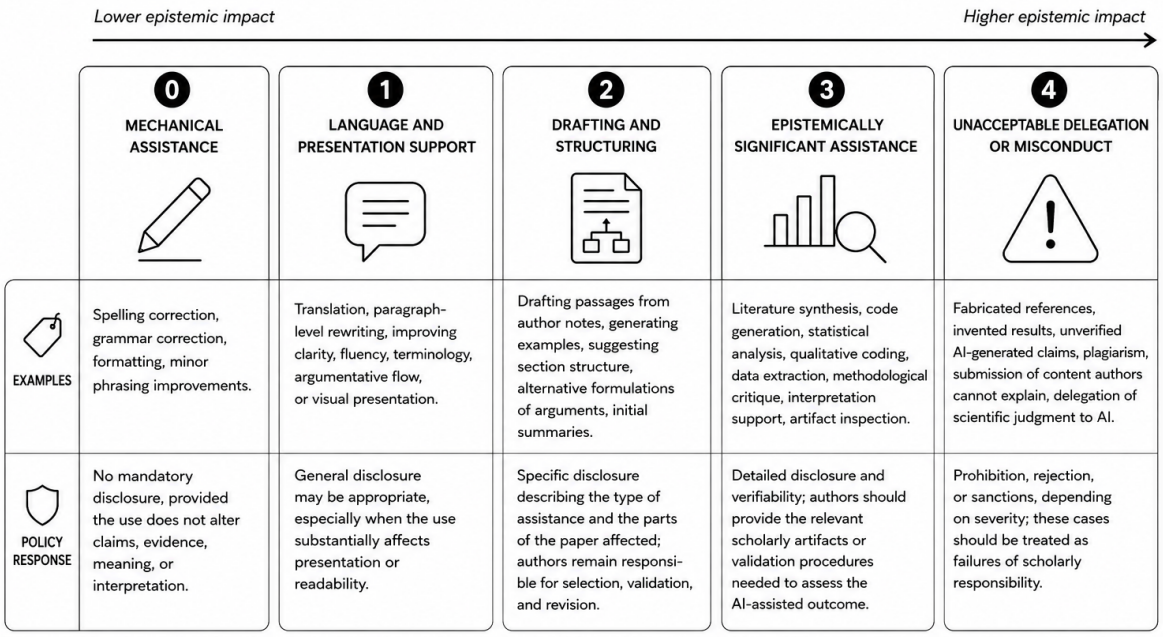

Figure 1: Layered model for governing generative AI use according to epistemic impact and governance requirements.

6 AI-Assisted Peer Review as the Hard Case

Peer review is where the scalability problem discussed above becomes hardest to ignore. The discussion so far has focused mainly on the use of generative AI by authors in the production of research papers. Peer review raises a related but more difficult problem. In authorship, the main policy challenge is to ensure that human authors remain responsible for the claims, evidence, methods, analyses, and conclusions of the submitted work. In peer review, the challenge is broader: AI assistance may affect the evaluation of confidential manuscripts, the protection of unpublished ideas, the independence of reviewers, and the legitimacy of editorial decisions.

This makes peer review the hard case for generative AI governance. It is also the case in which prohibition-centered policies are likely to be least effective. Reviewing is largely private, performed under time pressure, and difficult to audit. A reviewer may use generative AI to summarize a paper, organize notes, improve the wording of a review, identify possible weaknesses, inspect artifacts, compare claims with cited work, or draft an initial critique. Some of these uses may be harmless or even beneficial; others may compromise confidentiality, introduce errors, or amount to delegating scientific judgment. A policy that simply prohibits AI use in reviewing may therefore create a false sense of control while pushing actual use into invisibility.

6.1 AI Reviewers

AI-assisted peer review should not be dismissed as inherently illegitimate, especially as the scale of scholarly publishing makes purely human, unaided review increasingly difficult to sustain. Major AI conferences already operate at a scale that requires tens of thousands of reviewers and hundreds or thousands of area chairs. NeurIPS, for example, reported that main-track submissions grew from 9,467 in 2020 to 21,575 in 2025, requiring more than 20,000 reviewers and over 1,600 area chairs [30]. ICLR 2025 similarly reported 11,603 submissions, 18,325 reviewers, and 823 area chairs, with thousands more submissions and reviewers than in the previous year [21]. Such growth strains reviewer assignment, expertise matching, calibration, and quality control. Although recent work has framed AI-assisted peer review as a possible response to this scalability problem, they warn that AI may also reinforce a "more rather than better" output when publication incentives reward volume over rigor [14, 42].

In this context, AI assistance is both a source of risk and a possible component of review infrastructure. Recent empirical evidence suggests that AI reviewers can produce useful criticisms under certain conditions. In a large-scale expert annotation study, domain scientists evaluated individual review items produced by human reviewers and frontier-LLM reviewers for Nature-family papers [22]. The study found that AI reviewers could raise correct, significant, and well-evidenced criticisms, including issues not raised by human reviewers. It also found that AI reviewers were particularly useful in labor-intensive forms of scrutiny, such as inspecting source code, checking internal consistency, and identifying methodological or statistical concerns. Similarly, a randomized study at ICLR 2025 evaluated LLM-generated feedback on more than 20,000 reviews and found that a substantial fraction of reviewers incorporated suggestions, leading to more informative reviews as judged by blinded evaluators [38].

These findings shift the question from whether AI can generate review-like text to whether it can support specific review tasks without taking over the reviewer's judgment. AI tools may help reviewers identify inconsistencies, inspect artifacts, flag mismatches between claims and evidence, improve the clarity and actionability

of comments, or help chairs detect review-quality problems. Such assistance may become increasingly important for maintaining review quality under growing submission pressure. However, these benefits do not justify delegating editorial judgment to AI systems.

The same evidence also shows why AI reviewers should not replace human reviewers. AI-generated criticisms are not uniformly reliable: they may contain incorrect items, miss information explicitly present in the manuscript or supplementary material, or miscalibrate the severity of a concern because they lack field-specific knowledge about what is standard or expected in a subcommunity [22]. A review comment that is fluent and detailed but factually wrong can mislead authors, burden the review process, and distort editorial judgment.

Another limitation concerns diversity of perspective. Peer review is valuable not only because individual reviewers may be correct, but because different reviewers bring different expertise, assumptions, and priorities to the same manuscript. AI reviewers appear to overlap with one another substantially more than human reviewers do, suggesting that a panel of AI reviewers would likely provide less diversity than a panel of human reviewers [22]. Even if a single AI reviewer can be useful, replacing human review panels with multiple AI reviewers would risk narrowing the evaluative perspective.

The implication is that AI reviewers are not substitutes for human reviewers and should be treated as support tools. They may support consistency checking, artifact inspection, methodological scrutiny, review organization, and comment improvement. But they cannot assume sole responsibility for evaluating novelty, contribution, methodological adequacy, significance, or acceptability. These remain human scholarly judgments. If AI assistance becomes necessary to support peer review at scale, then the relevant policy problem should be how venues can provide bounded, equitable, confidentiality-preserving, and accountable forms of AI support.

6.2 AI-Assisted Review Confidentiality

AI-assisted peer review also introduces a confidentiality problem that is less central in author-side writing assistance. Submitted manuscripts often contain unpublished ideas, data, artifacts, and results. However, confidentiality should not be reduced to a simplistic distinction between "local processing" and "third-party AI". In practice, peer review already relies on third-party infrastructures: reviewers receive manuscripts through web-based submission systems, store PDFs in institutional or commercial cloud services, synchronize files across devices, use online editors, and exchange review-related messages through email providers. Some of these services now integrate generative AI features directly into storage and productivity environments, making the boundary between storing a confidential manuscript and processing it with AI increasingly unstable.

What matters is the guarantees governing the processing of the manuscript. Does the service preserve confidentiality? Is the content used for model training or product improvement? Can it be accessed by human reviewers employed by the provider? What are the retention policies? Is the tool used under an institutional agreement or a personal account? Are data isolated from other users and organizations? A reviewer uploading a manuscript to an unapproved public chatbot may clearly violate confidentiality. But a tool used under confidentiality-preserving conditions, such as no-training commitments, limited retention, restricted human access, and organizational data isolation, is a different case.

Current policies often leave this distinction underdeveloped. Some publishers broadly instruct reviewers not to upload manuscripts into generative AI tools, while others qualify the restriction in terms of third-party systems that do not promise confidentiality. These formulations acknowledge the risk, but they often do not clearly distinguish public chatbots, enterprise services, institutionally governed tools, and local or venue-approved models. As a result, confidentiality can become a proxy argument for prohibiting AI-assisted review altogether, rather than a criterion for specifying acceptable data-governance conditions.

This distinction also has an equity dimension. If acceptable AI-assisted review is implicitly limited to local models or private institutional deployments, then access to AI support becomes dependent on the reviewer's infrastructure. Researchers at well-resourced institutions may have access to secure enterprise tools or local computing resources, whereas others may only have access to public third-party services. A policy that simply prohibits external AI tools without offering approved alternatives may therefore restrict useful assistance unevenly and push less-resourced reviewers either to forgo AI support or to use it invisibly. Confidentiality-preserving AI review should therefore be treated as a venue-level governance problem, instead of an individual reviewer's infrastructure problem.

Beyond confidentiality, reliance on externally hosted AI services also raises broader questions concerning intellectual property, long-term dependence on commercial providers, and equitable access to AI capabilities. These issues extend beyond SBES itself, but they reinforce the importance of venue policies that clearly specify acceptable data-governance conditions while remaining compatible with different technological infrastructures.

6.3 Making AI-Assisted Review Governable

A workable AI-assisted review policy should separate clear violations from bounded assistance. Asking an AI system to produce or justify an accept/reject recommendation should be prohibited, because it delegates scientific judgment. Uploading a confidential manuscript to a public AI service with unclear retention, training, or confidentiality practices may also breach the reviewer's obligations. By contrast, using AI to improve the wording or organization of reviewer-written comments may be acceptable when no confidential manuscript content is exposed. Between these extremes, AI assistance may be acceptable for support tasks such as consistency checking, artifact inspection, literature cross-checking, or generating questions for closer human examination, provided that the reviewer verifies the output and remains responsible for the final review [4].

For SBES, this means that AI-assisted peer review should be governed as a process-design problem, but not necessarily as a centralized tool-approval problem. The community faces the usual pressures of conference reviewing: limited reviewer availability, variable review quality, and the difficulty of assessing increasingly complex empirical artifacts. AI assistance could plausibly help with some of these problems. However, a policy should focus less on

whether a tool is formally approved by the venue and more on the obligations under which the reviewer uses it: preserving confidentiality, avoiding data uses incompatible with the review process, verifying AI-assisted outputs, and not delegating evaluative judgment. Reviewers already rely on third-party infrastructures for email, storage, synchronization, editing, and backup. Treating AI services as categorically different from all other third-party services is difficult to justify unless the policy specifies the relevant difference in terms of data retention, model training, human access, confidentiality guarantees, or institutional agreements.

The role of the venue should therefore be to define boundaries and responsibilities rather than to pretend that all review-related data flows can be centrally controlled. If the venue can provide or recommend secure infrastructure, that would be valuable, especially for equity reasons. But in the absence of such infrastructure, reviewers should not be left with a simplistic rule that prohibits AI assistance while tolerating comparable third-party processing in ordinary review workflows.

6.4 AI Impact on Reviewer Recognition

AI-assisted reviewing also invites reflection on how conferences recognize reviewing work. Venues such as SBES often award "distinguished reviewer" honors to program committee members whose reviews are considered especially valuable. Even before the rise of generative AI, such awards were not uncontroversial. A scientific conference exists primarily to select, discuss, and disseminate research contributions from authors. Publicly rewarding reviewers may help signal that reviewing is valued, but it can also turn an internal service role into an object of institutional self-congratulation.

Generative AI makes this tension sharper. If AI tools can improve the fluency, structure, and apparent completeness of review text, then polished prose becomes an even weaker proxy for reviewer quality. Program chairs do not directly observe the reviewer's reasoning process; they observe review text, scores, confidence, timeliness, and participation in discussion. As AI assistance becomes more common, the meaning of "distinguished reviewer" awards becomes increasingly ambiguous: are they recognizing human expertise, diligence, responsible tool use, quality of written feedback, or contribution to the committee process? This does not mean that conferences should ignore reviewing work, but it does suggest that individual reviewer awards should not be treated as an unquestioned tradition. If their meaning cannot be defined in a fair and transparent way, discontinuing them may be a defensible option.

7 Conclusion

Over the past four decades, the SBES community has successfully navigated major disciplinary transformations. Generative AI is different because it affects both what software engineering studies and how software engineering papers are produced. Although the SBES community is right to define explicit AI usage policies, rules prohibiting AI-produced textual units rely on fragile distinctions between editing and generation. Such restrictions are difficult to enforce and only loosely connected to actual research integrity. A more robust path forward governs AI through human responsibility: authors must stand behind the claims, disclose AI use when it matters, and make AI-affected outputs checkable. The layered model proposed in this paper operationalizes this shift. By distinguishing AI usage levels from mechanical assistance to unacceptable delegation, and by linking them to specific governance requirements, we move from binary prohibitions to informed accountability, aligning our policies with the very intellectual tradition of software engineering.

Naturally, the proposals presented here are not the only possible accountability-centered governance model. Different research communities may adopt different balances between disclosure, verification, confidentiality, and acceptable AI use. Likewise, the layered model proposed here should be understood as a starting point for discussion rather than as a finalized policy specification.

The same reasoning applies to peer review, but with additional constraints. AI-assisted reviewing tasks may be allowed and even incentivized, especially under growing submission pressure, but confidentiality, infrastructure access, and non-delegation of evaluative judgment must be strictly governed. Bringing this AI-assisted policy framework into reality is a collective effort that will require the development of disclosure templates, reviewer guidance, and venue-specific rules for confidentiality, verification, and accountability. We therefore invite the SBES community to refine, challenge, and extend these ideas through future policy discussions and empirical studies on AI-assisted scholarly practice.

Artifact Availability

The analysis reported in this paper is based on publicly available calls for papers, publication policies, and related literature. Therefore, no additional artifacts are associated with this work.

Acknowledgements

Generative AI tools were used during the preparation of this manuscript for language revision, restructuring of selected passages, drafting from author-provided ideas, and support in summarizing and organizing related work. In the terms of the layered model proposed in this paper, these uses correspond to Levels 1, 2, and limited Level 3 assistance. The author(s) revised and validated all arguments, claims, interpretations, references, and textual and visual artifacts, and remain solely responsible for the final manuscript.

References

[1] Hussam Alkaissi and Samy I. McFarlane. 2023. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus* 15, 2 (2023), e35179.
[2] Alberto Bacchelli and Christian Bird. 2013. Expectations, Outcomes, and Challenges of Modern Code Review. In *Proc. of the 35th International Conference on Software Engineering (ICSE)*. IEEE, 712–721.
[3] Daivat Bhavsar et al. 2025. Policies on Artificial Intelligence Chatbots among Academic Publishers: A Cross-Sectional Audit. *Research Integrity and Peer Review* 10, 1 (2025), 1.
[4] Justus Bogner and Roberto Verdecchia. 2026. The State of Peer Review in Empirical Software Engineering: A Community Survey on Review Load, Quality, and GenAI Use. *SIGSOFT Software Engineering Notes* (2026).
[5] Yihan Cao et al. 2025. A Survey of AI-Generated Content (AIGC). *Comput. Surveys* 57, 5 (2025), 1–38.
[6] Dalmeet Singh Chawla. 2026. Researchers who use hallucinated references to face arXiv ban. *Nature* (2026). https://www.nature.com/articles/d41586-026-01595-5
[7] Jane Cleland-Huang et al. 2014. Software Traceability: Trends and Future Directions. In *Proc. of the on Future of Software Engineering (FOSE)*. ACM, 55–69.
[8] Paulo Anselmo da Mota Silveira Neto et al. 2013. 25 Years of Software Engineering in Brazil: Beyond an Insider's View. *Journal of Systems and Software* 86, 4 (2013), 872–889.
[9] Matteo Esposito et al. 2026. Generative AI for software architecture. Applications, challenges, and future directions. *Journal of Systems and Software* 231 (2026),

112607.
[10] Angela Fan et al. 2023. Large Language Models for Software Engineering: Survey and Open Problems. In *2023 IEEE/ACM International Conference on Software Engineering: Future of Software Engineering (ICSE-FoSE)*. IEEE, 31–53.
[11] Stefan Feuerriegel et al. 2024. Generative AI. *Business & Information Systems Engineering* 66 (2024), 111–126.
[12] Cyrus Ganjavi et al. 2024. Publishers' and Journals' Instructions to Authors on Use of Generative Artificial Intelligence in Academic and Scientific Publishing: Bibliometric Analysis. *BMJ* 384 (2024), e077192.
[13] Alessandro Garcia. 2013. Software Engineering in Brazil: Retrospective and prospective views. *Journal of Systems and Software* 86, 4 (2013), 869–871.
[14] Claudine Gartenberg et al. 2026. More Versus Better: Artificial Intelligence, Incentives, and the Emerging Crisis in Peer Review. *Organization Science* 37, 3 (2026), 795–812.
[15] Ali Essam Ghareeb et al. 2026. A Multi-Agent System for Automating Scientific Discovery. *Nature* (2026).
[16] Juraj Gottweis et al. 2026. Accelerating Scientific Discovery with Co-Scientist. *Nature* (2026).
[17] Jack Grove. 2026. Ban for authors submitting AI content 'welcome but unenforceable'. Times Higher Education. https://www.timeshighereducation.com/news/ban-authors-submitting-ai-content-welcome-unenforceable Accessed: 2026-05-29.
[18] Anthony Ha. 2026. Research repository ArXiv will ban authors for a year if they let AI do all the work. TechCrunch. https://techcrunch.com/2026/05/16/research-repository-arxiv-will-ban-authors-for-a-year-if-they-let-ai-do-all-the-work/ Accessed: 2026-05-29.
[19] Mohamed Helmy et al. 2025. Ten Simple Rules for Optimal and Careful Use of Generative AI in Science. *PLOS Computational Biology* 21, 10 (2025), e1013588.
[20] Xinyi Hou et al. 2023. Large Language Models for Software Engineering: A Systematic Literature Review. *ACM Transactions on Software Engineering and Methodology* (2023).
[21] International Conference on Learning Representations. 2025. ICLR 2025 Fact Sheet. https://media.iclr.cc/Conferences/ICLR2025/ICLR2025_Fact_Sheet.pdf Accessed: 2026-05-28.
[22] Seungone Kim et al. 2026. On the limits and opportunities of AI reviewers: Reviewing the reviews of Nature-family papers with 45 expert scientists. *arXiv preprint arXiv:2605.20668* (2026).
[23] Kalpesh Krishna et al. 2023. Paraphrasing Evades Detectors of AI-Generated Text, but Retrieval Is an Effective Defense. In *Advances in Neural Information Processing Systems*, Vol. 36.
[24] Zhi-Qiang Li et al. 2024. Use of Artificial Intelligence in Peer Review Among Top 100 Medical Journals. *JAMA Network Open* 7, 12 (2024), e2448609.
[25] Weixin Liang et al. 2023. GPT Detectors Are Biased Against Non-Native English Writers. *Patterns* 4, 7 (2023).
[26] Sebastian Porsdam Mann et al. 2024. Guidelines for Ethical Use and Acknowledgement of Large Language Models in Academic Writing. *Nature Machine Intelligence* (2024).
[27] Nabor C. Mendonça et al. 2022. A Decade of Internationalization of the Brazilian Symposium on Software Engineering: The Good, the Bad, and the Ugly. In *Proc. of the 36th Brazilian Symposium on Software Engineering (SBES)*. ACM, 1–10.
[28] Miryam Naddaf. 2025. How Are Researchers Using AI? Survey Reveals Pros and Cons for Science. *Nature* (2025). https://www.nature.com/articles/d41586-025-00343-5
[29] Humza Naveed et al. 2023. A Comprehensive Overview of Large Language Models. *arXiv preprint arXiv:2307.06435* (2023).
[30] NeurIPS 2025 Program Committee Chairs. 2025. Reflections on the 2025 Review Process from the Program Committee Chairs. https://blog.neurips.cc/2025/09/30/reflections-on-the-2025-review-process-from-the-program-committee-chairs/ Accessed: 2026-05-28.
[31] Vinu Sankar Sadasivan et al. 2023. Can AI-Generated Text Be Reliably Detected? *arXiv preprint arXiv:2303.11156* (2023).
[32] SBES. 2023. SBES 2023: Call for Papers – Research Track. https://cbsoft2023.ufms.br/en-US/sbes/pesquisa Accessed: 2026-05-25.
[33] SBES. 2024. SBES 2024: Call for Papers - Research Track. https://cbsoft.sbc.org.br/2024/sbes/pesquisa/?lang=en Accessed: 2026-05-25.
[34] SBES. 2025. SBES 2025: Call for Papers – Research Track. https://cbsoft.sbc.org.br/2025/sbes/pesquisa/?lang=en Accessed: 2026-05-25.
[35] SBES. 2025. SBES 2025: Chamada de Trabalhos – Trilha de Ferramentas. https://cbsoft.sbc.org.br/2025/sbes/ferramentas/?lang=pt Accessed: 2026-05-25.
[36] SBES. 2026. SBES 2026: Call for Papers – Research Track. https://cbsoft.sbc.org.br/2026/en/symposiums/sbes/pesquisa/call/ Accessed: 2026-05-25.
[37] Sociedade Brasileira de Computação. 2024. Code of Conduct for Authors in Brazilian Computing Society's Publications. https://sol.sbc.org.br/index.php/indice/conduta Accessed: 2026-05-25.
[38] Nitya Thakkar et al. 2025. Can LLM Feedback Enhance Review Quality? A Randomized Study of 20K Reviews at ICLR 2025. *arXiv preprint arXiv:2504.09737* (2025).
[39] Carmine Vassallo et al. 2018. Context Is King: The Developer Perspective on the Usage of Static Analysis Tools. In *Proc. of the 25th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 38–49.
[40] William H. Walters and Esther Isabelle Wilder. 2023. Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT. *Scientific Reports* 13 (2023), 14045. doi:10.1038/s41598-023-41032-5
[41] Steven Watson, Erik Brezovec, and Jonathan Romic. 2025. The Role of Generative AI in Academic and Scientific Authorship: An Autopoietic Perspective. *AI & Society* 40 (2025), 3225–3235.
[42] Qiyao Wei et al. 2025. The AI Imperative: Scaling High-Quality Peer Review in Machine Learning. *arXiv preprint arXiv:2506.08134* (2025).
[43] Shuhui Yin et al. 2025. Generative Artificial Intelligence (GAI) Usage Guidelines for Scholarly Publishing: A Cross-Sectional Study of Medical Journals. *BMC Medicine* 23 (2025), 77.
[44] Yangyang Zhao et al. 2017. The Impact of Continuous Integration on Other Software Development Practices: A Large-Scale Empirical Study. In *Proc. of the 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 60–71.