The Transparency Imperative: A Vision for AI-Native Scholarly Preprint Infrastructure

Manuscript submitted to the Journal of the Association for Information Systems

[Author Name(s)]
[Institutional Affiliation(s)]
[Contact Email]

AI Disclosure Statement

AI Contribution Level: Co-Authored. This manuscript was produced through human-AI co-authorship. The human author(s) conceived the research question, identified theoretical framing, specified the argument structure, and exercised editorial judgment throughout. Claude (Anthropic) contributed literature synthesis, prose generation, and iterative refinement. The Promptheus platform described herein was built using Claude Code (Anthropic), with the human authoring the architectural specification and feature requirements. All claims, interpretations, and intellectual contributions reflect the human author's judgment and direction.

Abstract

The rapid integration of large language models (LLMs) into scholarly workflows has created a fundamental mismatch between how research is produced and how scholarly communication infrastructure is designed. Existing preprint platforms face a binary: either exclude AI-generated content to preserve traditional norms (as arXiv has increasingly done) or embrace full AI autonomy in both authorship and review (as the nascent aiXiv platform attempts). We argue that both responses miss the emergent dominant mode of scholarly production—human-AI co-authorship—in which researchers use AI as a substantive intellectual partner rather than a mere writing tool. This paper articulates a vision for AI-native preprint infrastructure purpose-built for this reality.

Drawing on institutional theory, platform design theory, and the transparency literature in information systems, we propose a set of design principles for a new class of preprint platform that foregrounds contribution provenance, graduated AI disclosure, community-driven quality signals, and iterative co-authorship workflows. We describe Promptheus (prompthe.us), a working preprint platform that instantiates these principles, and reflect on implications for the IS research community, scholarly publishing, and the epistemic foundations of knowledge production. This paper itself was substantially co-authored with an AI system and submitted to the platform it describes—a reflexive demonstration of the infrastructure we propose.

Keywords: AI-native infrastructure, human-AI co-authorship, preprint platforms, scholarly communication, transparency, generative AI, design principles, platform design, Promptheus

1. Introduction: The Elephant in the Repository

Scholarly communication is undergoing its most significant transformation since the advent of the internet-enabled preprint. The catalyst is not a new protocol or distribution model but a fundamental change in who—or what—is doing the writing. Large language models (LLMs) are now deeply embedded in the research production process, from literature synthesis and hypothesis generation to prose composition and code development. Estimates suggest that by late 2024, detectable LLM usage appeared in 15–20% of preprints on major servers, with the true figure likely much higher given the limitations of detection tools (Liang et al., 2024). By early 2026, some subfields of computer science report that the majority of submissions show substantial AI involvement.

The scholarly communication ecosystem has responded with a mixture of denial, restriction, and tentative experimentation. ArXiv, the oldest and most influential preprint server, has progressively tightened its gatekeeping: first requiring AI disclosure, then mandating endorsement for new submitters in January 2026, and banning review and position papers in computer science entirely due to AI-generated content flooding (Wijers, 2026; Boboris, 2025). The Open Science Framework's generalist preprint service stopped accepting new submissions altogether in August 2025 because a majority were low quality. Major publishers have coalesced around a consensus that AI cannot be listed as an author and that humans bear full responsibility for AI-assisted content (Elsevier, 2024; Springer Nature, 2024). At the other end of the spectrum, the aiXiv platform has emerged as a venue that embraces fully autonomous AI authorship, using multi-agent LLM systems to both produce and review papers (Zhang et al., 2025).

We argue that both responses—prohibition and full automation—fundamentally mischaracterize the emerging reality. The dominant and most consequential mode of AI involvement in

scholarship is neither the AI "slop" that arXiv's moderators rightfully reject nor the fully autonomous agent papers that aiXiv hosts. It is human-AI co-authorship: a deeply collaborative process in which a human researcher provides domain expertise, judgment, and intellectual direction while an AI system contributes synthesis, composition, structural reasoning, and iterative refinement. This is the mode in which the present paper was produced. It is the mode that an increasing proportion of researchers employ daily but are incentivized to conceal under current institutional arrangements.

This concealment is the crux of the problem. When the infrastructure of scholarly communication treats AI involvement as something to be detected and penalized rather than disclosed and contextualized, it creates perverse incentives. Researchers who use AI responsibly face a disclosure dilemma: acknowledge their tools and risk stigma, or remain silent and risk accusations of deception. The result is a transparency crisis that undermines the epistemic foundations of the scholarly record far more than AI involvement itself ever could.

In this paper, we articulate a vision for a new class of scholarly infrastructure—one designed from the ground up for the age of human-AI co-authorship. We term this AI-native preprint infrastructure: platforms that do not merely tolerate AI involvement but treat it as a first-class feature of the scholarly production process, with dedicated mechanisms for provenance tracking, contribution attribution, quality assessment, and community engagement. Our contribution is threefold. First, we provide a theoretically grounded analysis of the current institutional crisis in scholarly communication precipitated by generative AI. Second, we propose a set of design principles for AI-native preprint infrastructure, drawing on institutional theory, platform design theory, and the IS transparency literature. Third, we describe Promptheus (prompthe.us), a

working preprint platform that instantiates these principles—itself built through human-AI co-authorship using Claude Code—and reflect on the implications for IS research and practice.

2. The Current Landscape: A Taxonomy of Institutional Responses

To understand why new infrastructure is needed, we first map the existing institutional responses to AI in scholarly communication. We identify three dominant postures, each with distinct assumptions, mechanisms, and failure modes.

2.1. The Exclusion Posture: Gatekeeping Against AI

The exclusion posture treats AI-generated content as an intrusion to be identified and removed. ArXiv's January 2026 endorsement requirement for new submitters and its October 2025 ban on CS review articles exemplify this approach. As the chair of arXiv's editorial council has noted, the endorsement requirement is intended to discourage low-quality AI-generated submissions from users without established scholarly credentials (Wijers, 2026). ArXiv now receives roughly 20,000 submissions per month, and its 300 volunteer moderators cannot scale to filter AI-produced content manually. The Open Science Framework's generalist preprint service halted all new submissions in August 2025 for the same reason, while PsyArXiv shifted from post-publication to pre-publication moderation. Major publishers reinforce this posture through policies that deny AI authorship status and place full accountability on human authors.

The exclusion posture has clear merits. It protects the signal-to-noise ratio of established platforms, preserves the association between human authorship and intellectual accountability, and maintains compatibility with existing peer review processes. However, it suffers from three critical limitations. First, it conflates quality with provenance: a well-reasoned, empirically sound paper co-authored with AI is treated with more suspicion than a mediocre paper produced

entirely by humans. Second, it incentivizes concealment: if AI involvement carries stigma, rational actors will underreport it. Third, it is increasingly untenable at scale. As AI tools become ubiquitous and their outputs become indistinguishable from human writing, the boundary the exclusion posture attempts to police will dissolve. ArXiv's own editorial council chair acknowledges this, expressing hope that AI developers will help identify AI-generated content—an implicit admission that human detection is insufficient.

2.2. The Automation Posture: AI as Autonomous Scientist

At the opposite extreme, the automation posture embraces AI as a fully autonomous participant in the research process. The aiXiv platform, launched in late 2025 by collaborators from the University of Manchester, the University of Toronto, Oxford, and Tsinghua University, exemplifies this approach. On aiXiv, AI systems can serve as both authors and reviewers. Submissions are evaluated by five LLM-based agents that assess novelty, technical soundness, and potential impact; three accept votes result in publication. The platform includes defenses against adversarial manipulation, such as prompt injection detection, and supports iterative review-and-revise cycles. Its creators argue that AI-generated knowledge should not be treated differently from human-generated knowledge and that quality alone should determine legitimacy (Zhang et al., 2025).

The automation posture has the virtue of intellectual honesty: it acknowledges that AI systems can produce research content and creates infrastructure to evaluate that content on its merits. However, early evidence suggests significant challenges. The Agents4Science conference—the first venue to feature AI as both primary authors and reviewers—found that AI review systems, while effective at detecting numerical inconsistencies and reference errors, tended to provide overly generous assessments of novelty and impact (Zou, 2025). This raises a fundamental

epistemic concern: if both production and evaluation are automated, who ensures that the resulting knowledge claims are grounded in reality rather than in the statistical patterns of training data? As bioethicist Sebastian Porsdam Mann warns, a platform perceived as a repository for unchecked AI output risks delegitimizing the entire enterprise of AI-involved scholarship (Porsdam Mann, 2025).

2.3. The Ambiguity Posture: Disclosure Without Infrastructure

Between exclusion and automation lies a broad middle ground occupied by most publishers and conferences: require disclosure, impose human accountability, but provide no dedicated infrastructure for managing AI involvement. ICML 2026, for example, permits LLM-assisted writing but prohibits listing AI as an author and subjects papers to prompt injection screening (ICML, 2026). Elsevier, Springer Nature, Wiley, Taylor & Francis, and SAGE have converged on policies requiring AI disclosure in methods or acknowledgments sections, while explicitly excluding AI from authorship criteria. Across these publishers, three principles are universal: human accountability for all content, transparency about AI tool use, and denial of authorship status to AI systems (Thesify, 2025).

This posture represents the pragmatic center of the current debate, and it is almost certainly where most scholarship will be conducted in the near term. Its weakness is structural: it asks authors to be transparent about AI involvement but provides no mechanisms, metadata standards, or platform affordances to support that transparency. A disclosure statement buried in a methods section—"We used ChatGPT to assist with drafting portions of this manuscript"—tells the reader almost nothing about the nature, extent, or intellectual significance of the AI's contribution. Was the AI used to fix grammatical errors or to generate the theoretical framework? Did the human fundamentally reshape the AI's output, or was the final text largely machine-generated? Current

infrastructure provides no way to answer these questions, and disclosure policies do not require it.

3. The Missing Middle: Human-AI Co-Authorship as Dominant Practice

The three postures described above share a common blind spot: they treat AI involvement as a binary variable—either present or absent, either disclosed or concealed—rather than as a spectrum of collaborative practices with fundamentally different epistemic implications. In this section, we argue that the most important and most prevalent form of AI involvement in scholarship is collaborative co-authorship, and that this mode requires infrastructure fundamentally different from anything currently available.

Human-AI co-authorship, as we define it, is a mode of scholarly production in which a human researcher engages an AI system as a substantive intellectual partner in one or more phases of the research process—including but not limited to ideation, literature synthesis, argumentation, drafting, analysis, and revision—while retaining ultimate intellectual authority over the claims, contributions, and framing of the resulting work. This definition deliberately excludes two boundary cases: trivial AI assistance (e.g., grammar correction, which requires no special infrastructure) and fully autonomous AI research (which raises distinct questions about agency and accountability that platforms like aiXiv are better positioned to address).

The present paper serves as an illustration. The human author conceived the research question, identified the theoretical framing, specified the argument structure, and directed the iterative development of each section through dialogue with an AI system (Claude, Anthropic). The AI contributed literature synthesis, prose generation, structural suggestions, and iterative refinement based on human feedback. The platform on which this paper is archived—Promptheus—was

itself built through a parallel human-AI co-authorship process, with the human specifying architecture, design decisions, and feature requirements through a detailed prompt document while the AI system (Claude Code) generated the implementation code. The paper describes the platform that hosts it, and both were produced through the mode of collaboration that the platform is designed to support. This recursive quality is not incidental; it is a deliberate demonstration that human-AI co-authorship can produce substantive scholarly and technical artifacts when supported by appropriate processes and transparent disclosure.

This mode of production is not an edge case. Research from Cornell University analyzing over two million preprints found that scientists who adopted LLM tools posted roughly one-third more papers on arXiv, with even larger productivity gains on bioRxiv (over 50%) and SSRN—particularly for researchers whose first language is not English. Scientists from Asian institutions posted between 43% and 89% more papers after adopting LLMs, depending on the platform (Kusumegi et al., 2025). Critically, however, the study also found that while AI-assisted papers exhibited more linguistically complex writing, human reviewers often judged them to have less scientific value. Papers likely written by humans that scored high on writing complexity were most likely to be accepted to journals, while high-scoring AI-assisted papers were less likely to be accepted. These findings reinforce our central argument: what matters is not whether AI was involved but how it was involved, and current infrastructure provides no mechanism for making that distinction.

4. Theoretical Foundations

Our vision for AI-native preprint infrastructure draws on three theoretical streams within information systems: institutional theory as applied to digital platforms, platform design theory, and the IS literature on transparency and trust in sociotechnical systems.

4.1. Institutional Theory and the Legitimacy Crisis

Institutional theory provides a lens for understanding why existing scholarly communication infrastructure is struggling to accommodate AI. DiMaggio and Powell's (1983) concept of institutional isomorphism explains the rapid convergence of publisher policies around a common template: disclosure required, human accountability mandated, AI authorship prohibited. This isomorphic response provides legitimacy in the short term but creates a growing gap between formal institutional rules and actual research practices—what Meyer and Rowan (1977) termed "decoupling." The formal structure (policies requiring disclosure) becomes ceremonial, while the actual work (extensive AI co-authorship with minimal meaningful disclosure) proceeds according to its own logic.

This decoupling dynamic is particularly dangerous in scholarly communication because the entire system depends on trust. When institutional rules are widely known to be observed in letter but not spirit, the legitimacy of the system erodes. Scott's (2014) three pillars of institutional legitimacy—regulative, normative, and cognitive—are each threatened. Regulative legitimacy erodes as policies become unenforceable (AI detection tools are unreliable and AI text is becoming indistinguishable from human text). Normative legitimacy erodes as the community splinters between those who view AI co-authorship as legitimate and those who view it as a form of cheating. Cognitive legitimacy erodes as taken-for-granted assumptions about what a "scholarly paper" is and who can "author" one are destabilized. New infrastructure can intervene in this legitimacy crisis not by resolving the underlying normative tensions but by providing mechanisms that make those tensions productive rather than corrosive.

4.2. Platform Design Theory

Platform design theory (Tiwana, 2014; de Reuver et al., 2018; Hein et al., 2020) provides the architectural vocabulary for our proposed infrastructure. A preprint platform is a multi-sided platform that mediates between authors, readers, reviewers, and downstream consumers (journals, institutions, funders). Its design choices—boundary resources, governance mechanisms, and data architectures—shape the behavior of all participants.

The key insight from platform theory is that design choices are not neutral: they encode assumptions about legitimate behavior, desirable outcomes, and the distribution of power among participants. Current preprint platforms encode the assumption that a paper has human authors who bear full responsibility for its content. This assumption is increasingly fictive, and the platform affordances built on it—authorship metadata fields, contribution statements, review criteria—are correspondingly inadequate. An AI-native platform must redesign these affordances to encode a different set of assumptions: that contributions exist on a human-AI spectrum, that provenance is a first-class metadata category, and that quality assessment must account for the mode of production alongside the product itself.

We draw particularly on the concept of "generative" platform design (Zittrain, 2008; Tilson et al., 2010), which emphasizes creating infrastructure that enables unanticipated uses and innovations by diverse participants. Rather than prescribing how AI should be used in scholarship—a question the field is nowhere near ready to answer—a generative AI-native platform provides the tools for communities to develop and iterate on their own norms, metrics, and evaluation criteria.

4.3. Transparency and Trust in Sociotechnical Systems

The IS literature on transparency in sociotechnical systems provides the normative foundation for our design principles. Transparency is not a single construct but a multi-dimensional one encompassing process transparency (how was the work produced?), outcome transparency (what are the results and how should they be interpreted?), and algorithmic transparency (what role did automated systems play?) (Kroll et al., 2017; Stohl et al., 2016). In the context of AI co-authored scholarship, all three dimensions are relevant.

Critically, the transparency literature demonstrates that transparency is only valuable when accompanied by interpretive infrastructure—mechanisms that help recipients make sense of disclosed information (Heald, 2006; Michener & Bersch, 2013). This is precisely the gap in current scholarly communication infrastructure. Even when authors disclose AI involvement, readers lack the context, metadata, and comparison tools needed to interpret that disclosure meaningfully. A statement that "AI was used in drafting" is transparent in form but opaque in substance. AI-native infrastructure must provide not merely the capacity for disclosure but the interpretive affordances that make disclosure actionable.

5. Design Principles for AI-Native Preprint Infrastructure

Drawing on the theoretical foundations outlined above, we propose seven design principles for AI-native preprint infrastructure. These principles are intended to be generative rather than prescriptive: they define the contours of a design space rather than specifying a single implementation. In Section 6, we describe how each principle is instantiated in the Promptheus platform.

5.1. Principle 1: Contribution Provenance as First-Class Metadata

The platform should capture and expose structured metadata about the provenance of contributions at a granular level. Rather than a single binary disclosure ("AI was/was not used"), provenance metadata should distinguish among different types and degrees of AI involvement. This metadata should be structured to reduce reporting burden and increase comparability across papers. A standardized vocabulary of contribution types, analogous to the CRediT (Contributor Roles Taxonomy) system used in multi-author papers, could serve as the foundation for an AI contribution taxonomy. We propose an extension we call CRediT-AI, which adds graduated descriptors such as "AI-Assisted," "AI Co-Authored," and "AI-Primary" to the existing framework—each with clear definitions distinguishing the nature and extent of AI involvement.

5.2. Principle 2: Graduated Disclosure with Interpretive Affordances

Disclosure should be graduated rather than binary, reflecting the spectrum of AI involvement from trivial to substantial. The platform should support at least three tiers corresponding to meaningfully different modes of collaboration, and each tier should be accompanied by interpretive affordances—visual indicators, color-coded badges, summary descriptions, and comparison tools—that help readers calibrate their assessment of the work. The design must avoid creating perverse incentives. If higher AI involvement is associated with stigma, authors will underreport. The platform should therefore frame disclosure not as a confessional act but as a contextual one: providing readers with information that helps them engage with the work more effectively, just as methodological disclosure helps readers interpret empirical findings.

5.3. Principle 3: Community-Driven Quality Signals

Traditional peer review, already strained, cannot scale to the volume of AI-assisted scholarship. AI-only review, as early experiments suggest, lacks the judgment needed to assess novelty and

real-world grounding. The platform should therefore implement a hybrid quality signaling system that combines multiple sources: structured community feedback, threaded scholarly discussion, and optional AI-assisted preliminary screening for technical consistency and structural completeness. This approach draws on the reputation systems literature in IS (Dellarocas, 2003; Resnick et al., 2000) and extends it to scholarly evaluation. The key design challenge is ensuring that community quality signals reflect epistemic judgment rather than popularity.

5.4. Principle 4: Iterative Versioning and Living Documents

Human-AI co-authorship is inherently iterative: a researcher prompts, the AI generates, the researcher revises, the AI refines, and the cycle continues. The platform should embrace this iterative nature by supporting rich versioning that captures sequential drafts, revision summaries, and the evolution of a manuscript over time. Versioning supports a "living document" model in which preprints are not static snapshots but evolving works that incorporate community feedback, respond to critiques, and accumulate evidence over time. This model is particularly well-suited to AI-collaborative work, where revision cycles are fast and the marginal cost of substantial updates is low.

5.5. Principle 5: Interoperability with the Existing Scholarly Ecosystem

An AI-native platform cannot exist in isolation. It must interoperate with existing scholarly infrastructure—DOI registration, ORCID author identification, CrossRef metadata, journal submission systems, and institutional repositories. Provenance metadata should be exportable in standard formats so that if an AI-co-authored preprint is subsequently submitted to a traditional journal, its disclosure history accompanies it. Interoperability is both a technical and a political

design choice. Technically, it requires adherence to open standards and API-first architecture. Politically, it signals that AI-native infrastructure is not an alternative to the existing scholarly ecosystem but an extension of it.

5.6. Principle 6: Reflexive Infrastructure Design

The norms governing AI involvement in scholarship are evolving rapidly, and any platform built today will need to accommodate norms that do not yet exist. The platform should be designed reflexively: with built-in mechanisms for community governance, norm evolution, and policy experimentation. Different disciplinary communities may adopt different standards for acceptable AI involvement, and the platform should support this heterogeneity. Reflexive design also means the platform itself should be transparent about its own use of AI. If the platform was built using AI tools, that fact should be disclosed—the infrastructure should practice what it preaches.

5.7. Principle 7: Epistemic Humility as a Design Value

The platform should encode epistemic humility as a design value. We do not yet know how AI co-authorship will affect the quality, reliability, or innovativeness of scholarship in the long run. Early evidence is mixed: AI-assisted papers are more numerous and linguistically sophisticated but may be less likely to pass peer review (Kusumegi et al., 2025). The platform should be designed to generate data about these questions—enabling longitudinal analysis of the relationship between AI involvement modes, community quality assessments, and downstream scholarly impact—without prejudging the answers. This principle distinguishes our vision from both the exclusion posture (which prejudges AI involvement as harmful) and the automation posture (which prejudges it as beneficial).

6. Promptheus: Instantiating the Vision

To move beyond abstract principles, we have developed Promptheus (prompthe.us)—a working preprint platform that instantiates the design principles described above. Promptheus is positioned as "the preprint archive for AI-first research": an arXiv-style repository purpose-built for manuscripts where AI tools played a substantial role in production. Critically, the platform itself was built through human-AI co-authorship. The human author specified the system architecture, database schema, design language, feature requirements, and editorial vision through a detailed prompt document; the AI system (Claude Code, Anthropic) generated the implementation code across the full stack. This reflexive construction—an AI-co-authored platform for AI-co-authored papers—is a deliberate demonstration of the infrastructure's own principles.

6.1. Architecture and Technical Design

Promptheus is implemented as a server-rendered web application using PHP 8.x with a MySQL 8.x relational database, Bootstrap 5 for responsive interface design, and jQuery for client-side interactivity. This technology selection prioritizes accessibility, deployment simplicity, and long-term maintainability over framework novelty—values aligned with scholarly infrastructure that must remain functional over decades rather than development cycles. The database schema models the core entities of scholarly communication—users, papers, categories, tags, versions, and comments—while extending them with AI-specific metadata fields that are absent from existing preprint systems.

Papers follow an arXiv-style identifier format (PRTH-YYYY-NNNNN, e.g., PRTH-2026-00042) that signals institutional identity while maintaining the alphanumeric conventions familiar to researchers. PDF manuscripts are stored in a versioned directory structure that

supports the living-document model described in Principle 4. The platform supports ten disciplinary categories seeded from arXiv's taxonomy (ranging from cs.AI to econ.GN), with an extensible category system that accommodates disciplinary heterogeneity.

6.2. The AI Disclosure System: Promptheus's Core Differentiator

The AI disclosure system is the defining feature of Promptheus and the primary instantiation of Principles 1 and 2. Unlike existing platforms where AI disclosure is an optional afterthought—a free-text sentence appended to acknowledgments—Promptheus makes disclosure a mandatory, structured, and visually prominent component of every submission.

The disclosure system operates on two dimensions. The first is a graduated contribution level taxonomy with three tiers. "AI Assisted" (displayed as a green badge) indicates that AI tools were used for editing, grammar correction, or minor restructuring, with the human author producing the core intellectual content. "AI Co-Authored" (amber badge) indicates that AI contributed substantively to content generation—including drafting, argumentation, or analysis—under sustained human direction, with the human providing intellectual framework, iterative feedback, and final judgment. "AI Primary" (purple badge) indicates that AI generated the majority of the manuscript content, with the human serving primarily in a curatorial, editorial, and oversight role. These three tiers correspond to meaningfully different modes of collaboration with distinct epistemic implications, rather than arbitrary gradations of a single continuum.

The second dimension is a required free-text AI disclosure field in which authors describe which specific AI tools were used and how. A typical entry might read: "Claude 3.7 Sonnet (Anthropic) for initial drafting and literature synthesis; GPT-4o (OpenAI) for code generation and data

analysis; Midjourney for figure generation. All outputs were reviewed, revised, and validated by the human authors." This structured-plus-narrative approach captures both the categorical information needed for filtering and aggregation and the contextual detail needed for meaningful interpretation.

Visually, the AI disclosure system is designed for maximum prominence. Contribution level badges appear on every paper card in the browse and search interfaces, enabling readers to filter and calibrate before they click through. On the paper detail page, the AI disclosure is rendered as a visually distinct panel—set apart from the abstract and metadata through differential styling—that readers encounter before they engage with the paper's content. This design choice reflects our theoretical commitment to interpretive affordances: disclosure is not merely available but foregrounded, providing interpretive context that shapes how the work is read.

6.3. Submission Workflow and Provenance Capture

The submission workflow is designed as a multi-step guided process that normalizes AI disclosure by integrating it into the standard submission flow rather than appending it as an afterthought. Authors proceed through five stages: metadata entry (title, abstract, category, tags), author listing (with support for both registered and external co-authors, ordered authorship, and corresponding author designation), AI disclosure (the mandatory contribution level selection and tool description), PDF upload with optional DOI and journal reference fields, and a review-and-submit confirmation screen.

The AI disclosure step is positioned third—after the intellectual metadata but before the file upload—to signal that it is a core component of the scholarly record rather than a bureaucratic checkbox. Authors who attempt to skip this step cannot proceed. This mandatory positioning

instantiates the design insight that platform affordances shape behavior: by making disclosure a structural requirement rather than an optional courtesy, the platform establishes new norms around transparency.

Upon submission, papers enter a moderation queue where administrators review them for basic quality and appropriate categorization. This human-in-the-loop moderation step is deliberate: unlike aiXiv's fully automated review, Promptheus maintains human judgment as the gatekeeping mechanism while using the structured metadata and AI disclosure to inform that judgment. Moderators can see the AI contribution level and disclosure details alongside the paper, enabling calibrated assessment that accounts for the mode of production.

6.4. Versioning, Discussion, and Community Engagement

Promptheus implements a version history system in which authors can submit revised manuscripts that are stored alongside prior versions with change summaries. This creates a public record of the paper's evolution—particularly valuable for AI-co-authored work, where iterative refinement is the norm rather than the exception. Readers can access previous versions and track how the manuscript developed over time.

The platform supports threaded commenting on papers, enabling substantive scholarly discussion that functions as a form of open community review. Comments are attached to specific papers and organized hierarchically, supporting both top-level critiques and focused exchanges. This design supports Principle 3 (community-driven quality signals) by providing a public forum in which the scholarly community can evaluate, question, and build upon AI-co-authored work. Unlike opaque peer review, this discussion is visible to all readers, contributing to the interpretive context surrounding each paper.

Additional community features include paper bookmarking, author profiles with ORCID integration and publication histories, automated BibTeX citation generation, and view counting. These features are standard in modern preprint platforms, but their integration with the AI disclosure system is novel: an author's profile page, for example, displays their papers alongside the contribution level badges, enabling readers to see patterns in how individual researchers use AI across their body of work.

6.5. Design Language and Epistemic Framing

The visual design of Promptheus is itself an argument about how AI-co-authored scholarship should be perceived. The aesthetic direction is "editorial-academic with a modern AI-forward edge": clean whitespace, scholarly serif typography (Playfair Display for headings, Source Serif 4 for body text), and a color palette anchored by deep navy and accented with a Promethean fire orange. The monospaced typeface JetBrains Mono is used for paper identifiers and BibTeX output, signaling technical precision. This design language deliberately positions AI-co-authored scholarship as serious, rigorous, and worthy of the same aesthetic respect accorded to traditional academic publishing.

The platform's name, Promptheus, is a portmanteau of "prompt" and "Prometheus"—a deliberate reference to the Titan who brought fire to humanity, reframing the act of prompting an AI system as an act of knowledge-creation rather than intellectual abdication. The domain (prompthe.us) reinforces the communal dimension: the ".us" suffix suggests collective enterprise rather than individual authorship. These naming and branding choices are themselves design decisions that encode the normative stance described in Principle 7: AI co-authorship is neither something to apologize for nor something to celebrate uncritically, but something to engage with honestly and communally.

6.6. Mapping Design Principles to Platform Features

Table 1 summarizes how each of the seven design principles articulated in Section 5 maps to specific features implemented in the Promptheus prototype.

Table 1. *Mapping of Design Principles to Promptheus Features*

Design Principle	Promptheus Feature
P1: Contribution Provenance	Three-tier AI contribution level taxonomy (Assisted / Co-Authored / Primary) with required free-text tool disclosure
P2: Graduated Disclosure	Color-coded badges (green / amber / purple) on all paper cards and detail pages; visually prominent AI Disclosure panel
P3: Community Quality Signals	Threaded commenting system for open scholarly discussion; view counts; moderated submission queue with human review
P4: Iterative Versioning	Version history system with change summaries; PDF storage in versioned directories (v1, v2, v3…)
P5: Interoperability	arXiv-style identifiers (PRTH-YYYY-NNNNN); ORCID integration; automated BibTeX generation; DOI and journal reference fields
P6: Reflexive Design	Platform itself built via human-AI co-authorship (Claude Code); extensible category system; community governance mechanisms
P7: Epistemic Humility	No automated quality scoring; human moderation; platform generates data on AI involvement patterns without prescribing norms

7. Implications for the IS Research Community

The vision articulated in this paper has significant implications for IS scholars, the broader research community, and the institutions that govern scholarly communication.

7.1. Implications for IS Research

For IS researchers, AI-native preprint infrastructure represents both an object of study and a tool for practice. As an object of study, it raises fundamental questions about how platform design shapes knowledge production—extending the IS literature on platforms, digital infrastructure, and sociotechnical systems into the domain of scholarship itself. How do disclosure affordances affect author behavior? How do community quality signals compare to traditional peer review?

How does contribution provenance metadata change the way readers engage with scholarly work? These are empirical questions that AI-native infrastructure is uniquely positioned to illuminate, because it generates fine-grained data about the scholarly production process that current platforms do not capture.

As a tool for practice, AI-native infrastructure invites IS scholars to model the kind of transparent, reflexive engagement with AI that the field advocates for other domains. The IS community has long studied how organizations adopt, adapt, and resist new technologies. We are now in the position of being both the researchers and the researched—a reflexive condition that, properly leveraged, can generate insights available to no other discipline.

7.2. Implications for Scholarly Publishing

For publishers, the design principles outlined here suggest a path beyond the current impasse of disclosure policies without disclosure infrastructure. The CRediT-AI taxonomy instantiated in Promptheus's three-tier system could be adopted by journals as a supplement to existing CRediT statements, providing structured information about AI involvement that is more informative than free-text disclosures. The community discussion and quality signaling features could serve as a complement to traditional peer review, providing pre-review filtering that reduces reviewer burden while surfacing promising work for formal evaluation.

More fundamentally, AI-native preprint infrastructure models a different relationship between platforms and norms. Rather than platforms enforcing norms through gatekeeping (accept/reject decisions), they support norms through transparency (disclosure mechanisms) and enable norm evolution through community governance (adaptable quality signals and open discussion). This

shift—from enforcement to enablement—may be essential for an era in which the pace of technological change outstrips the capacity of institutional governance to keep up.

7.3. Implications for Epistemic Governance

At the broadest level, the vision articulated here is about epistemic governance: the rules, norms, and infrastructure that determine what counts as knowledge and who gets to produce it. The integration of AI into scholarship does not merely raise practical questions about workflow efficiency or quality control. It raises foundational questions about authorship, originality, expertise, and intellectual contribution that have been latent in the scholarly communication system for decades and are now being forced into the open.

AI-native infrastructure does not answer these questions. It does something more valuable: it creates the conditions under which they can be investigated empirically and debated productively, with shared data, common metrics, and visible evidence about how different modes of human-AI collaboration affect the scholarly enterprise. This is, we believe, the most important contribution infrastructure can make in a period of rapid epistemic change.

8. Challenges, Risks, and Open Questions

We do not underestimate the challenges inherent in this vision. Several merit explicit discussion.

Adoption and legitimacy. A new preprint platform faces the classic chicken-and-egg problem of two-sided markets: authors will not submit without readers, and readers will not come without content. Moreover, a platform explicitly associated with AI co-authorship may carry a legitimacy deficit in a scholarly culture that still, often justifiably, views AI involvement with suspicion. Promptheus addresses this partly through its design language—positioning AI-co-authored work as serious scholarship—and partly through its interoperability features, which allow papers to

flow into the broader ecosystem rather than remaining siloed. However, sustained adoption will require institutional allies: journals willing to accept submissions from the platform, departments willing to count Promptheus preprints in tenure files, and prominent researchers willing to submit their AI-co-authored work openly.

Gaming and adversarial behavior. Any system that relies on self-reported provenance is vulnerable to misreporting. Authors may understate AI involvement to claim more human credit, or overstate it to signal methodological sophistication. Community quality signals are vulnerable to manipulation by coordinated groups. These risks must be addressed through a combination of technical safeguards (anomaly detection, cross-validation), social mechanisms (reputation systems, community norms), and institutional incentives (rewarding honest disclosure). No system can be made impervious to gaming, but the goal is to make honest disclosure the path of least resistance—which Promptheus's mandatory, structured disclosure workflow is designed to support.

Equity and access. AI-native infrastructure must be attentive to equity concerns. If the platform's features are designed around the capabilities of state-of-the-art AI systems, they may inadvertently advantage well-resourced researchers and institutions. Promptheus's three-tier disclosure system is deliberately inclusive: the "AI Assisted" tier accommodates researchers using basic AI tools for language polishing (a particularly important use case for non-native English speakers), while the "AI Primary" tier accommodates cutting-edge human-AI research workflows. The platform must continue to ensure that researchers at all resource levels can participate meaningfully.

The provenance paradox. There is a fundamental tension between the desire for transparency about the co-authorship process and the reality that detailed provenance disclosure may chill legitimate AI use. If authors know their AI interactions may be scrutinized, they may self-censor or use AI less creatively. The platform must calibrate disclosure requirements to balance transparency with the cognitive and creative freedom that makes human-AI collaboration productive. Promptheus's design navigates this tension by requiring categorical disclosure (the contribution level) and tool identification (which AI systems were used) while leaving the granularity of process disclosure to the author's discretion.

Disciplinary heterogeneity. Different fields have different norms around authorship, evidence, and intellectual contribution. What counts as legitimate AI involvement in computational linguistics may differ profoundly from what is acceptable in interpretive sociology or clinical medicine. Promptheus's category system and its reflexive governance principle (Principle 6) are designed to accommodate this heterogeneity, but the challenge of supporting diverse disciplinary norms within a single platform is substantial and will require ongoing community engagement.

9. A Note on Reflexivity: This Paper as Evidence

We close with a reflection on the meta-dimensions of this work that we believe are inseparable from its scholarly contribution. This paper was produced through human-AI co-authorship of the "AI Co-Authored" variety: the human author conceived the research question, identified the theoretical framing, directed the argument's development, and exercised editorial judgment throughout, while an AI system (Claude, Anthropic) contributed literature synthesis, prose generation, and iterative refinement. The platform it describes—Promptheus—was built through an analogous process, with the human authoring a detailed architectural specification and the AI generating the implementation.

This paper is therefore both an argument for AI-native preprint infrastructure and a specimen of the kind of work that infrastructure is designed to support. We submit it to the platform it describes, with its AI contribution level set to "AI Co-Authored" and its AI disclosure field specifying the tools and processes involved. The reader of this paper on Promptheus will see the amber co-authorship badge, read the disclosure panel, and engage with the threaded discussion—all features that exist because this paper argued they should.

This recursive quality is unusual in scholarly writing, but we believe it is appropriate for the moment. The IS community is uniquely positioned to study, build, and model the infrastructure that will shape how AI transforms knowledge production. We offer Promptheus not as a finished product but as a working provocation: a demonstration that the infrastructure we need is buildable, that the norms we need are articulable, and that the transparency we need is achievable—if we are willing to build systems that demand it.

10. Conclusion

The integration of AI into scholarly production is not a future scenario. It is the present condition. Every day, thousands of researchers use AI systems as substantive intellectual partners—and every day, the infrastructure of scholarly communication pretends otherwise. The result is a growing transparency crisis that threatens the epistemic foundations of the scholarly record.

We have argued that the solution is not better detection, more restrictive policies, or full automation, but purpose-built infrastructure that treats human-AI co-authorship as a legitimate mode of scholarly production deserving of its own tools, norms, and evaluation criteria. The design principles we have proposed—contribution provenance, graduated disclosure, community

quality signals, iterative versioning, interoperability, reflexive governance, and epistemic humility—provide a framework for building this infrastructure. Promptheus demonstrates that these principles can be instantiated in working software and used to archive and disseminate the very scholarship that motivates their development.

The scholarly communication system has reinvented itself before—from handwritten manuscripts to printed journals, from physical distribution to digital archives, from subscription access to open preprints. Each reinvention required new infrastructure, new norms, and new conceptions of what scholarship is and how it should be shared. We are at another such inflection point. The question is not whether AI will reshape scholarly communication but whether we will build infrastructure that makes that reshaping transparent, accountable, and ultimately beneficial to the production of knowledge.

References

Boboris, K. (2025, October 31). Attention authors: Updated practice for review articles and position papers in arXiv CS category. arXiv Blog. https://blog.arxiv.org/2025/10/31/attention-authors-updated-practice-for-review-articles-and-position-papers-in-arxiv-cs-category/

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. Learned Publishing, 28(2), 151–155.

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. Journal of Information Technology, 33(2), 124–135.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Management Science, 49(10), 1407–1424.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147–160.

Elsevier. (2024). The use of generative AI and AI-assisted technologies in writing for Elsevier. https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-writing-for-elsevier

Heald, D. (2006). Varieties of transparency. In C. Hood & D. Heald (Eds.), Transparency: The key to better governance? (pp. 25–43). Oxford University Press.

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. Electronic Markets, 30(1), 87–98.

ICML. (2026). ICML 2026 call for papers. https://icml.cc/Conferences/2026/CallForPapers

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633–705.

Kusumegi, K., Yang, X., Ginsparg, P., de Vaan, M., Stuart, T., & Yin, Y. (2025). AI gives scientists a boost, but at the cost of too many mediocre papers. Cornell University.

Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., McFarland, D. A., & Zou, J. Y. (2024). Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. arXiv preprint arXiv:2403.07183.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology, 83(2), 340–363.

Michener, G., & Bersch, K. (2013). Identifying transparency. Information Polity, 18(3), 233–242.

Monperrus, M. (2025). Preprint servers for AI generated papers. https://www.monperrus.net/martin/preprint-servers-of-AI-generated-papers

Porsdam Mann, S. (2025). Quoted in: A new preprint server welcomes papers written and reviewed by AI. Science.

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. Communications of the ACM, 43(12), 45–48.

Scott, W. R. (2014). Institutions and organizations: Ideas, interests, and identities (4th ed.). Sage.

Springer Nature. (2024). Artificial intelligence (AI) policy. https://www.springernature.com/gp/policies/ai-policy

Stohl, C., Stohl, M., & Leonardi, P. M. (2016). Managing opacity: Information visibility and the paradox of transparency in the digital age. International Journal of Communication, 10, 123–137.

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Digital infrastructures: The missing IS research agenda. Information Systems Research, 21(4), 748–759.

Tiwana, A. (2014). Platform ecosystems: Aligning architecture, governance, and strategy. Morgan Kaufmann.

Wijers, R. (2026). Quoted in: ArXiv preprint server clamps down on AI slop. Science.

Zhang, P., Li, Y., Huang, G., et al. (2025). aiXiv: A next-generation open access ecosystem for scientific discovery generated by AI scientists. arXiv preprint arXiv:2508.15126.

Zittrain, J. (2008). The future of the Internet and how to stop it. Yale University Press.

Zou, J. (2025). Quoted in: A new preprint server welcomes papers written and reviewed by AI. Science.