

Redesigning the Paper as the Unit of Scholarly Intermediaries

Abstract

The academic paper was designed as a human readable unit of knowledge under conditions of print scarcity, slow dissemination, and journal centered gatekeeping. As AI systems increasingly generate manuscripts, summarize literatures, and allocate attention, the paper functions less as a narrative artifact and more as an intermediary that feeds machine triage, aggregation, and synthesis. Under these conditions, the binding constraint shifts from rhetorical persuasion to verification under abundance. This article argues that the paper should be redesigned as a structured research object with an evidence dominant core: explicit claim registries, assumption registries, machine readable claim to evidence mapping, executable provenance, and versioning for post publication updating. Narrative remains important, but primarily as an interpretive layer that organizes evidence rather than as the principal carrier of claims. The argument builds on evidence focused publication systems in applied sciences and on analyses of how publication systems shape accumulation and impact (Zhang & Ertug, 2025). The paper concludes with design principles and governance implications for journals, repositories, and evaluation systems across disciplines.

1. Introduction

The academic paper is simultaneously a cognitive device and an institutional object. It structures how researchers make claims, how communities certify them, and how knowledge accumulates across time. The dominant format, often operationalized as IMRaD variants, was engineered for sequential human reading and for review under scarcity: limited space, limited attention, and slow diffusion.

Two changes destabilize this fit. First, the marginal cost of producing scholarly text is declining as AI systems draft, paraphrase, and elaborate with minimal effort. Second, the marginal cost of first pass reading is also declining as AI systems summarize, cluster, and rank papers for humans. These shifts do not merely increase speed. They change selection pressure. When text is abundant, the scarce resource is credible verification.

This article develops a systems claim: the paper is becoming a unit of scholarly intermediaries. That is, it increasingly serves as an interface that mediates between (a) human authors and reviewers, (b) AI systems that triage and synthesize, and (c) downstream users who treat the literature as an evidentiary base for decisions, models, and policy. Under these conditions,

narrative first design becomes fragile because it is easy to generate and hard to audit at scale.

The core proposal is not "less theory." It is a reallocation of primacy: evidence and auditability become first class, while narrative becomes a secondary rendering layer. This orientation is consistent with how many applied sciences maintain rigor through structured evidence accumulation and heterogeneous contribution types rather than requiring each individual paper to deliver theoretical novelty (Moher et al., 2009; Guyatt et al., 2008; Vincenti, 1990).

2. The paper as infrastructure, not merely text

Scholars in the history and sociology of science have long treated scientific outputs as infrastructural artifacts rather than as neutral containers of ideas. Kuhn (1962) emphasized that scientific development depends on shared exemplars and evaluative standards that coordinate a community. Latour (1987) argued that facts travel through networks of inscriptions and devices that stabilize claims. The paper is one of the central inscriptions that allows claims to travel, be contested, and be institutionalized.

In a human centric regime, the paper's infrastructural role was closely tied to narrative competence: motivations, novelty

claims, and discursive positioning. In a machine mediated regime, the infrastructural role shifts toward structured interoperability. AI systems that summarize, embed, and retrieve papers require stable identifiers for constructs, claims, methods, and evidence. Without such structure, the paper remains readable but becomes computationally lossy.

This perspective aligns with the view that publication systems shape what gets produced and how it accumulates. Zhang and Ertug (2025) frame publication systems as interdependent elements that affect balance between evidence and theory, diversity of research types, responsiveness to real world challenges, and the coupling of exploration and replication. While their argument is grounded in organization research, the system logic generalizes: when evaluation systems reward narrative performance and novelty signaling, the literature becomes theoretically prolific but evidentially thin; when systems reward structured evidence and replication, theory can emerge from accumulation.

3. Abundance and the verification bottleneck

A central implication of AI text generation is that plausibility is no longer a scarce output. If many papers can be produced that look methodologically orthodox and rhetorically competent,

then surface signals lose discriminating power. The problem is not that AI generated text is inherently unreliable. The problem is that volume breaks the current audit regime.

This is an instance of what Borgman (2015) describes in data intensive scholarship: as the scale and heterogeneity of outputs increase, infrastructures for curation, provenance, and reuse become central to credibility. Nielsen (2011) similarly emphasizes that networked science changes the division of labor, shifting value toward tools that make knowledge discoverable, recombinable, and verifiable at scale.

In empirical fields, the reproducibility movement can be interpreted as an early signal of this bottleneck. The Open Science Collaboration (2015) demonstrated that even in a human authored regime, many published findings do not replicate cleanly. When we add massive growth in output and machine mediated diffusion, the cost of false positives and overstated generalizations rises. The question becomes: what design of the paper makes verification cheaper?

4. The paper as a structured research object

A research object is not a PDF. It is a set of linked, versioned components that allow claims to be queried, evaluated, and

recombined. The argument here is that the “paper” should evolve into a canonical rendering of such an object, rather than being the object itself.

A minimal evidence dominant architecture has four layers.

4.1 Claim layer: explicit claim registry

The primary failure mode of narrative first papers is that claims are entangled with motivation, literature positioning, and hedging. Machines and humans both struggle to identify what is actually asserted. The claim layer makes this explicit.

A claim registry should include, at minimum: (a) the focal claim(s), (b) whether each claim is descriptive, predictive, causal, or design oriented, (c) the implied scope conditions, and (d) the intended level of generalization. In causal work, this maps to explicit estimands (Shaver, 2020). In theory work, it maps to explicit construct definitions, proposed mechanisms, and discriminating predictions.

4.2 Assumption layer: explicit assumption registry

Claims are only as credible as their assumptions. Yet assumptions are often implicit, scattered, or framed as limitations rather than as formal conditions for inference. Making assumptions explicit is central to machine mediated

accumulation because it enables downstream systems to compare claims under comparable conditions.

For causal inference, this includes identification assumptions, measurement assumptions, and model dependence. For observational research, reporting standards such as STROBE exist precisely to increase transparency and comparability (von Elm et al., 2007). For systematic synthesis, PRISMA institutionalizes explicit reporting of inclusion, exclusion, and bias considerations (Moher et al., 2009). These standards are not rhetorical. They are designed to reduce ambiguity in cumulative evidence systems.

4.3 Evidence layer: claim to evidence mapping with uncertainty

An evidence dominant paper binds each major claim to the outputs that support it and makes uncertainty a first class object. In many empirical papers, uncertainty is present but not mapped to claims. The result is selective emphasis: readers infer confidence from rhetoric rather than from structured evidence.

The health sciences offer a mature model of evidence grading, in which strength of recommendation is separated from quality of evidence, often using frameworks such as GRADE (Guyatt et al., 2008). The key design principle is separability: claims can be strong, tentative, or speculative, but the evidence basis is explicitly graded.

In a machine mediated regime, this mapping enables automated meta inference: aggregation of effect sizes, detection of heterogeneity, and identification of boundary conditions. It also enables rapid human audit: a reviewer can interrogate the evidence object directly rather than reconstructing it from narrative.

4.4 Provenance and executability layer: data, code, and lineage
Evidence dominance requires that the analysis pipeline is inspectable. This is now routine in many computationally intensive fields. In medicine, machine learning systems increasingly require careful reporting of training data, evaluation, and drift (Rajkomar et al., 2019). In climate and Earth system science, deep learning is integrated with process understanding, but credibility depends on data provenance and model validation (Reichstein et al., 2019).

For the paper, the implication is straightforward: data schemas, preprocessing, and code should be treated as integral components of the research object, not as optional supplements. This is not only about replication. It is about enabling machine systems to identify comparable constructs and to reuse evidence without reinterpreting prose.

5. Where does theory sit in an evidence dominant design

A common objection is that evidence dominance risks narrowing scholarship to what is measurable and executable, marginalizing conceptual innovation. This concern is real, but it misstates the tradeoff.

In evidence dominant design, theory remains essential as an organizing schema. Simon (1996) characterized design as the science of the artificial, emphasizing that complex human made systems require iterative cycles of problem formulation, artifact creation, and evaluation. A parallel point applies here: theory is the schema that organizes what counts as evidence, which contrasts are meaningful, and what boundary conditions are plausible.

The key shift is functional. Narrative theory work becomes more valuable when it produces structure: precise constructs, explicit mechanisms, and discriminating predictions that can be linked to evidence objects. In other words, theory moves from rhetorical positioning toward schema design for accumulation.

This view is consistent with how engineering knowledge develops through epistemic accumulation rather than isolated theoretical leaps (Vincenti, 1990). It is also consistent with calls in strategy and organization research to emphasize cumulative identification over single study claims (Shaver, 2020) and to

treat replication as a core component of knowledge development (Bettis et al., 2016).

6. Implications for journals, repositories, and evaluation systems

If papers become research objects with structured cores, then journals' comparative advantage changes. In a scarcity regime, journals curate narrative contributions and allocate prestige. In an abundance regime, journals can differentiate by certifying auditability: completeness of claim registries, explicitness of assumptions, quality of evidence mapping, and executability.

This logic parallels Zhang and Ertug's (2025) argument that publication systems differ in how they balance evidence and theory and how they enable exploration and replication. Under evidence dominance, exploration becomes easier to publish because exploratory claims can be explicitly labeled and bounded, and because replication has a defined place in the system rather than being treated as a second class contribution (Bettis et al., 2016). Responsiveness can also improve because evidence objects can be disseminated quickly, with subsequent versioned updating.

For applied fields, relevance becomes less about rhetorical appeals and more about portability of evidence across contexts.

Entrepreneurship research, for example, faces an ongoing relevance challenge (Wiklund et al., 2019). Evidence dominant objects can support relevance by making effect sizes, heterogeneity, and boundary conditions explicit, enabling practitioners and policymakers to judge applicability without relying on narrative translation.

7. Governance risks: formalization, inequality, and gaming

Redesign is not merely technical. It has governance risks.

First, formalization can become performative. Authors may “fill fields” without increasing transparency. This risk is minimized when fields are linked to executable artifacts and when audit trails exist.

Second, evidence dominance can increase inequality if infrastructure access determines credibility. This mirrors concerns in data intensive scholarship (Borgman, 2015).

Governance responses include shared repositories, standardized tooling, and journal level support for reproducibility checks.

Third, structured systems can be gamed. Any metricized field is vulnerable to optimization. The remedy is plural certification: multiple signals tied to replication, calibration, and post publication performance rather than to single score proxies.

Improvement science offers a pragmatic analogy: building reliable systems requires iterative feedback loops and attention to system design, not only to individual performance (Bryk et al., 2015). Applying that logic here implies that publication reform should be treated as system engineering.

8. Conclusion

The paper is evolving from a narrative artifact into a unit of scholarly intermediaries that feeds machine triage, synthesis, and downstream decision systems. Under AI mediated abundance, the scarce resource is verification. The design response is to move toward an evidence dominant research object architecture: explicit claim registries, explicit assumption registries, structured claim to evidence mapping, and executable provenance, with narrative serving as an interpretive rendering layer.

The argument is compatible with the broader idea that publication systems shape accumulation and impact, and that evidence centric designs can enable more responsive, diverse, and cumulative science (Zhang & Ertug, 2025). The generalization is that this logic now applies to scholarship as a whole, not only to any one discipline, because AI systems increasingly intermediate both production and attention.

References

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal*, 37(11), 2193–2203. <https://doi.org/10.1002/smj.2580>

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2), 424–455. <https://doi.org/10.1257/jel.48.2.424>

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650), 924–926. <https://doi.org/10.1136/bmj.39489.470347.AD>

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269.

<https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

Nielsen, M. (2011). *Reinventing discovery: The new era of networked science*. Princeton University Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358. <https://doi.org/10.1056/NEJMra1814259>

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204. <https://doi.org/10.1038/s41586-019-0912-1>

Shaver, J. M. (2020). Causal identification through a cumulative body of research in the study of strategy and organizations. *Journal of Management*, 46(7), 1244-1256. <https://doi.org/10.1177/0149206320931131>

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). MIT Press.

Vincenti, W. G. (1990). *What engineers know and how they know it: Analytical studies from aeronautical history*. Johns Hopkins University Press.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *The Lancet*, 370(9596), 1453–1457. [https://doi.org/10.1016/S0140-6736\(07\)61602-X](https://doi.org/10.1016/S0140-6736(07)61602-X)

Wiklund, J., Wright, M., & Zahra, S. A. (2019). Conquering relevance: Entrepreneurship research's grand challenge. *Entrepreneurship Theory and Practice*, 43(3), 419–436. <https://doi.org/10.1177/1042258719844146>

Zhang, S. X., & Ertug, G. (2025). CROSSROADS: Organization research as an applied science: Lessons from fields that shape practice and policy. *Organization Science*, 36(5), 2028–2039. <https://doi.org/10.1287/orsc.2025.20459>