

CRAIL

CANADIAN REVIEW OF ARTIFICIAL INTELLIGENCE, LAW & SOCIETY

Vol. 1 | No. 3 | Fall 2026, Part III

NOVEMBER 2026

DIGITUS NON EST IUDICIUM

crail.ca

CRAIL
CANADIAN REVIEW OF ARTIFICIAL INTELLIGENCE, LAW & SOCIETY

DIGITUS NON EST IUDICIUM

VOLUME 1 · NUMBER 3 · FALL 2026, Part III
November 2026

ISSN 3111-458X

crail.ca

JOURNAL INFORMATION

About CRAIL

CRAIL - Canadian Review of Artificial Intelligence, Law & Society is an independent Canadian interdisciplinary academic journal devoted to artificial intelligence, emerging technologies, new scientific developments, and their interaction with law, institutions and society.

Motto

Digitus non est iudicium.

Publication details

Journal	CRAIL - Canadian Review of Artificial Intelligence, Law & Society
Volume / issue	Vol. 1, No. 3
Issue	Fall 2026, Part III (November 2026)
Frequency	Quarterly
Website	crail.ca
ISSN	3111-458X
Working language	English
Publisher	Dr. Adel Kildeev
Place of publication	Richmond Hill, Ontario, Canada
Publication format	Online journal; full-issue PDF, individual article PDF and HTML

Editorial and publishing data

Editor-in-Chief: Dr. Adel Kildeev (Canada)
Editor: Mr. Alejandro Sinopoli (Argentina)
Editor: Mr. Abdullah Yuri Kang (France)
Copyright: Authors retain copyright. CRAIL does not require transfer of copyright.
Licence: Unless otherwise expressly stated, published articles are licensed under CC BY-NC 4.0.

FROM THE EDITOR

From Doctrine to Philosophy

The third issue of *CRAIL — Canadian Review of Artificial Intelligence, Law & Society* brings together three articles addressing different dimensions of a common question: what happens to human understanding, judgment, and responsibility when increasingly sophisticated technological systems become integrated into education, law, and scientific research?

The articles approach this question from distinct perspectives. Their common concern is not whether artificial intelligence can produce useful results, but whether institutions and individuals can preserve the intellectual capacities necessary to evaluate those results independently.

The issue opens with John Gao's "Cognitive Seeds as an Integrative Design Framework for Human Understanding in AI-Mediated Learning." Gao examines a fundamental distinction between producing a successful answer and developing the understanding necessary to explain, question, and apply knowledge independently. His proposed Cognitive Seeds framework emphasizes relational understanding, progressive elaboration, and the learner's ability to revise conclusions in response to new evidence.

Importantly, Gao does not present the framework as a validated theory of cognition. Instead, he identifies its conceptual foundations, distinguishes it from established educational approaches, and proposes a research agenda for evaluating its effectiveness. The article thus raises an educational question of growing significance: how can artificial intelligence assist learning without reducing the learner to a recipient of externally generated answers?

The second contribution, "AI Governance in Review: Competing Doctrines, Legal Arguments, and Emerging Research," examines selected scholarship circulated through the SSRN Legal Scholarship Network in September 2026. Rather than offering an exhaustive literature survey, the article critically compares competing approaches to legal responsibility, technological autonomy, professional accountability, and institutional governance.

Its central concern is the distinction between technological capability and legal authority. The ability of a computational system to generate conclusions or execute complex operations does not, by itself, resolve questions of authorship, responsibility, or legitimate decision-making power. The review also emphasizes that effective governance requires more than abstract declarations of human oversight: legal doctrines must be evaluated against evidence, institutional realities, and the practical conditions of their implementation.

The issue concludes with "Academia Without Scholars: Artificial Intelligence and the Erosion of Independent Judgment." This article moves from legal and institutional questions toward the philosophical foundations of scholarly activity itself. It examines the possibility that academic institutions may continue producing publications, degrees, and measurable research outputs while weakening the processes through which independent intellectual judgment is developed and transmitted.

Drawing on philosophical analysis and empirical indicators from higher education, the article distinguishes documented institutional transformations from claims that artificial intelligence

has already caused cognitive deterioration. Its argument is therefore not that universities have ceased to produce scholars, but that the increasing separation between the production of academic texts and their independent human verification presents a serious institutional risk.

Taken together, these contributions trace a progression from the cultivation of understanding, through the legal organization of responsibility, to the preservation of independent judgment within the institutions responsible for producing knowledge.

This progression explains the title of the present editorial: *From Doctrine to Philosophy*. Legal rules and governance architectures remain indispensable, but their effectiveness ultimately depends upon human beings capable of understanding, questioning, verifying, and accepting responsibility for their decisions.

CRAIL does not regard technological progress and human intellectual independence as opposing objectives. The challenge is to develop institutions in which technological assistance strengthens, rather than displaces, the exercise of human judgment.

Digitus non est iudicium.

Dr. Adel Kildeev
Editor-in-Chief
CRAIL - Canadian Review of Artificial Intelligence, Law & Society

AIMS & SCOPE

About CRAIL

CRAIL - Canadian Review of Artificial Intelligence, Law & Society is an independent Canadian interdisciplinary academic journal devoted to the study of artificial intelligence, emerging technologies, new scientific developments, and their interaction with law, institutions and society.

The journal provides a forum for rigorous scholarship across law, technology, science, philosophy, psychology, sociology and the broader social sciences. While artificial intelligence is one of CRAIL's principal areas of inquiry, its scope is deliberately broader. CRAIL welcomes research on emerging technologies and scientific developments, their philosophical and ethical foundations, their psychological and social consequences, and their impact on law, institutions, professional practice and human judgment.

CRAIL publishes doctrinal, theoretical, empirical and interdisciplinary research, as well as carefully developed contributions from legal, scientific and technical practice.

Our editorial approach is neither technological evangelism nor technological pessimism. CRAIL seeks independent thought, verifiable argument, intellectual openness and serious dialogue across disciplines. We particularly welcome work that challenges established assumptions, brings together fields that too often develop in isolation, and examines not only what new technologies can do, but what their development means for people and institutions.

Technology and science are the subject matter; law and society are the context; the human person remains at the centre.

CRAIL is published quarterly in Richmond Hill, Ontario, Canada. The working and primary publication language of the journal is English. Scholarly works in other languages may be considered, but any work published in a language other than English must also be published with a complete English translation.

CONTENTS

Volume 1 · Number 3 · Fall 2026, Part III

1. **Cognitive Seeds as an Integrative Design Framework for Human Understanding in AI-Mediated Learning** **11-30**

John Gao (USA)

2. **AI Governance in Review: Competing Doctrines, Legal Arguments, and Emerging Research** **31-83**

Dr. Adel Kildeev (Canada)

3. **Academia Without Scholars: Artificial Intelligence and the Erosion of Independent Judgment** **84-152**

Dr. Adel Kildeev (Canada)

Article page ranges refer to the complete CRAIL issue.

EDITORIAL POLICIES

Core Editorial Principles

Effective September 2026. The following is a concise statement of CRAIL's principal editorial rules. Full and current policies will be maintained at crail.ca.

EDITORIAL SCREENING

Submission does not create an entitlement to peer review. Every manuscript must first pass CRAIL's editorial screening for relevance, scholarly value, integrity and compliance with submission requirements. Manuscripts may be declined without external review.

PEER REVIEW

CRAIL operates a single-blind peer-review process. Manuscripts selected for peer review are normally evaluated by one qualified reviewer. Additional review may be sought where warranted.

PUBLICATION FEES AND INDEPENDENCE

CRAIL does not charge submission or peer-review fees. A publication fee may be charged only after a manuscript has been formally accepted for publication. Payment does not influence editorial evaluation, peer review or acceptance. CRAIL may reduce or waive the publication fee.

AI USE AND DISCLOSURE

Substantive use of artificial intelligence by authors or reviewers must be disclosed. Authors and reviewers remain personally responsible for their work, claims, sources, evaluation and conclusions. AI assistance may support judgment; it may not substitute for judgment.

CONFIDENTIALITY

Submitted and unpublished manuscripts are confidential scholarly material. Editors and reviewers must not disclose or improperly use manuscript content, including through external AI or third-party services where confidentiality, intellectual property, personal information or author rights may be compromised.

CITATION AND SOURCE INTEGRITY

Authors are responsible for the accuracy, authenticity and traceability of citations, quotations, references, data and source materials. Fabricated, materially inaccurate or unverifiable authorities are incompatible with CRAIL's standards. AI-generated citations must be independently verified.

COPYRIGHT AND LICENSING

Your work remains yours. CRAIL does not require authors to transfer copyright. Authors grant CRAIL a non-exclusive publication licence. Unless otherwise expressly stated, CRAIL articles are published under CC BY-NC 4.0.

REPOSITORIES AND REPUBLICATION

Prior circulation as a preprint or deposit in recognized repositories such as SSRN, Zenodo, arXiv and comparable platforms does not, by itself, preclude consideration. Previously published work may be republished only where the necessary rights and permissions are clear and documented.

LANGUAGE

English is CRAIL's working and primary publication language. Works in other languages may be considered, but any non-English work published by CRAIL must also appear with a complete English translation.

LENGTH AND DISCIPLINES

CRAIL imposes no fixed word or page limits. Scholarly value, relevance and substance - not length - govern editorial consideration. CRAIL welcomes rigorous interdisciplinary work across law, technology, science, philosophy, psychology, sociology and related fields.

EDITORIAL BOARD

Editorial Board membership is a responsibility, not a publication privilege. Board members must disclose material conflicts of interest and must not participate in editorial evaluation or publication decisions concerning manuscripts in which they have such a conflict.

CORRECTIONS AND SCHOLARLY RECORD

CRAIL may issue corrections, clarifications, editorial notices or retractions where necessary. Substantive changes to the published scholarly record will not be made silently.

Digitus non est iudicium.

PUBLICATION RECORD

Rights, Access & Publication

AUTHOR RIGHTS

Authors retain copyright in their work. CRAIL does not require transfer of copyright.

LICENCE

Unless otherwise expressly stated, articles are published under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) licence.

PUBLICATION FEES

CRAIL charges no submission or peer-review fee. Any publication fee is considered only after formal acceptance and may be reduced or waived.

REPOSITORIES

Preprints and repository deposits, including SSRN, Zenodo, arXiv and comparable academic repositories, do not by themselves prevent consideration by CRAIL.

AI DISCLOSURE

Substantive use of AI by authors and reviewers must be disclosed. Human authors and reviewers remain responsible for verification, judgment and conclusions.

SCHOLARLY RECORD

CRAIL maintains corrections and other post-publication notices transparently. Substantive changes to the published scholarly record are not made silently.

Your work remains yours.

Full and current editorial and publication policies: crail.ca

Cognitive Seeds as an Integrative Design Framework for Human Understanding in AI-Mediated Learning

John Gao

Humanity to Eternity (H2E)

Independent Researcher / Convener, H2E

United States

October 2026

Crossref DOI: Pending Registration

Abstract

Generative artificial intelligence can improve an assisted answer without improving the understanding its user can subsequently exercise. This conceptual article presents Cognitive Seeds as an integrative design framework, not a new explanatory theory of cognition. A cognitive seed denotes an initial, revisable relational understanding whose proposed generative property requires evidence of further questioning, application, and learning from new information. The framework distinguishes external carriers, a selected minimal knowledge set, and a learner-developed minimal cognitive core. A micro-worldview is an unmeasured pedagogical heuristic for connecting aspects of that core across self, others, action, and uncertainty. Comparison with spiral curricula, threshold concepts, big ideas, and conceptual-change approaches locates the contribution in coordinated design commitments rather than novel psychological mechanisms. These commitments include explicit relation-to-carrier maps, context-bound omission tests, and assessment of independent use and preparation for future learning. Seven activity descriptions support observation without constituting a validated scale. A staged research agenda separates feasibility from effectiveness, controls prompt–assessment alignment, specifies the AI's instructional role, and distinguishes justified revision from model deference. No participant data or efficacy claims are presented. The framework earns additional value only if these commitments improve design decisions or outcomes beyond credible existing approaches.

Keywords: cognitive seeds; instructional design; generative artificial intelligence; relational understanding; preparation for future learning; transfer

1. Introduction

An educational answer has at least two possible achievements: it can satisfy a task requirement, and it can contribute to the learner's subsequent ability to understand and act. These achievements may coincide, but their relationship requires examination. When generative AI supplies explanations, arguments, or completed solutions, evaluating the

resulting artifact alone cannot establish which parts of the understanding remain available to the person who used it.

Learning and performance can diverge (Soderstrom & Bjork, 2015). In high-school mathematics, Bastani et al. (2025) compared ordinary instruction with GPT Base, a general-purpose chat interface, and GPT Tutor, which used teacher-informed instructions and materials to guide assistance. Assisted practice improved in both AI conditions. On a subsequent examination without AI, GPT Base performed worse than the control; GPT Tutor did not show that disadvantage. This is evidence about those instructional conditions, not a general finding that generative AI reduces cognition or that safeguards guarantee durable learning.

Cognitive Seeds asks what initial structures of understanding help a learner question, interpret, and revise externally supplied material. The author traces the proposal to an educational work written as Si Jiu, referred to here by the published manuscript "New Education: Cultivating Seeds of Cognition (English only): ChatGPT Era " (Gao, 2024). The curriculum uses concise poetry and progressively elaborated activities as starting materials. This article develops a conceptual framework and an evaluation agenda; it does not report an evaluation of that curriculum.

The author also describes a published manuscript, "A Guide for Planting the Future of Education: Cognitive seeds textbook toolkit" (Gao, 2025), containing instructional modules and lesson examples.

The contribution is an integrative design framework, expressed through a practical vocabulary. It coordinates three commitments: identify relations separately from their carriers; test a starting set's adequacy and dispensability for a specified learner group; and examine independent application, revision, and subsequent learning rather than assisted output alone. Its components have antecedents. The claim is that making their coordination explicit may improve instructional specification and evaluation, not that a new cognitive mechanism has been discovered. Childhood cultivation and adult AI interaction remain separate applications requiring separate evidence.

2. Conceptual approach and scope

The article uses conceptual analysis and design-oriented synthesis rather than a systematic review. Sources were selected for relevance to knowledge organization, learner-generated elaboration, assistance, transfer, and instructional design. The argument compares the proposed terms with established approaches, identifies counterexamples, and specifies tests of design commitments. This selective process cannot establish that the literature lacks another framework combining similar commitments; novelty is consequently limited to the particular integration and its explicit evaluation requirements.

Three kinds of claim are distinguished throughout. Descriptive claims summarize cited research. Construct definitions specify how the proposal uses its terms. Design propositions state expectations that remain to be tested. Examples illustrate proposed relationships; they are not observations of participants. No new human-subject data are reported.

The term cognitive system is used at a functional and educational level. It does not identify a single neural structure, establish a theory of consciousness, or imply that all cognitive functions have one common mechanism. Likewise, the proposed seven activity dimensions are a working vocabulary for task design. They are not asserted to be exhaustive, independent psychological faculties or a validated measurement scale.

3. The cognitive system and its interaction with the world

3.1 A functional definition

A personal cognitive system is provisionally defined here as the evolving organization of perceptual, mnemonic, conceptual, affective, and goal-directed processes through which a person interprets experience, forms judgments, and guides action. The definition emphasizes relationships among processes. It allows the same information to have different significance for people with different experiences, concerns, and available knowledge.

For educational purposes, cognitive development can be examined through recurring interactions among experience, interpretation, action, feedback, and revision. A learner observes an event, relates it to prior understanding, proposes an explanation, acts or seeks further information, and reconsiders the explanation in light of consequences. This is a heuristic sequence, not a claim that cognition always proceeds in discrete stages. Some responses are implicit, affective, habitual, or simultaneous, and social interaction can alter every part of the sequence.

Cultivation targets conditions that educators can influence within this recurring process: which features receive attention, what vocabulary and examples support interpretation, whether action produces interpretable feedback, and whether learners have a meaningful opportunity to revise. Instruction should specify those opportunities rather than infer development from exposure to the carrier alone.

A child rebuilding a fallen block tower illustrates the distinction between an event and its interpretation. The initial explanation might attribute collapse to an accidental touch. Further attempts might reveal the importance of base width. The educational question is whether the learner can explain why changing the base matters and recognize comparable instability in a new construction. Successful reconstruction by copying an adult is a different observation.

3.2 Knowledge organization and external support

The importance of organized knowledge is compatible with schema-based approaches to learning. Sweller (1988) links problem-solving demands and schema acquisition, warning that activity can consume processing resources without producing the desired learning. A compact curriculum therefore cannot justify itself by giving learners less information. It must still supply the knowledge needed for the intended reasoning, and its tasks must be manageable for the particular learner.

An account centered on the learner need not treat external resources as threats. Clark and Chalmers (1998) argue that environmental resources can participate in cognitive processes, while Risko and Gilbert (2016) examine cognitive offloading and its

metacognitive regulation. These perspectives challenge an equation between cognitive agency and performing every operation internally. The concern of Cognitive Seeds is narrower: whether learners can understand the purpose and limits of assistance, evaluate relevant claims, and retain capabilities required by the educational goal.

Accordingly, the framework distinguishes a temporary dependence appropriate to learning from an unexplained loss of opportunity to practice the target activity. A worked example, a teacher's explanation, and an AI-generated suggestion may each support learning. Their educational value depends partly on what the learner does with them. Removing assistance can help examine a specified capability; it is not a universal test of human worth or cognitive independence.

4. Cognitive seeds and a minimal cognitive core

4.1 The seed as a relational understanding

A cognitive seed is defined as an initial, revisable relational understanding that can support further explanation, questioning, and application. The seed is a learner's understanding, whereas a poem, story, diagram, or experience is a possible carrier. Remembering a carrier does not establish that its proposed relations have been understood.

For example, a story can carry a relation between actions and consequences. Recalling the story is evidence of recall; explaining the relation and its limits in an unfamiliar case supplies different evidence. The proposed generative property goes further: the learner should be able to formulate a relevant new question, identify information that could resolve it, and use new evidence to extend or revise an explanation. Section 8.4 specifies a preparation-for-future-learning task. No single successful answer establishes a seed as a distinct internal entity.

Four requirements constrain the proposal. The relation must be understandable at the intended level of development. It must connect with more than one possible experience or question. It must permit elaboration beyond reproduction of the original material. Finally, it must remain open to correction. These are proposed selection and evaluation criteria, not established necessary and sufficient conditions for learning.

Operational support for generativity should require more than a novel phrase. A relevant question must address an uncertainty in the case; an information choice must distinguish plausible alternatives; and an extension must use newly supplied evidence appropriately. Independent assessors should accept more than one defensible answer and check whether the same behavior occurs after replacing the carrier. Rote question templates, irrelevant novelty, and confident repetition do not meet these criteria.

A concept can provide an external carrier when presented as a term, rule, or worked example. The seed designation concerns the learner's initial, revisable use of the relation, not the concept's label or an assumed stage of mastery. More developed conceptual understanding can retain the same expandable and revisable properties; seed and mature understanding need not be sharply separated psychological kinds.

4.2 What minimal means

A minimal knowledge set is an externally selected collection of materials judged sufficient to begin a specified learning pathway. A minimal cognitive core is the learner's developing organization of relations that makes those materials usable. The two cannot be assumed to coincide: educators can deliver the same materials while learners construct different understandings.

Minimality is relative to a goal, population, and context. It does not mean the smallest universal list of facts, nor does it imply that foundational knowledge can replace subject expertise. Selection involves a trade-off between accessibility and adequate explanatory resources. A defensible starting set should be compact enough to revisit but rich enough to distinguish relevant cases, support reasons, and expose its own limits.

This contextual requirement accords with How People Learn II, which emphasizes interactions among prior knowledge, learning strategies, materials, goals, and cultural contexts (National Academies of Sciences, Engineering, and Medicine, 2018). A reduction in curriculum size alone is consequently not evidence of improved cognitive development. The claim must concern what learners become able to do and the conditions under which they can do it.

An operational test can begin by defining a target task family, an eligible learner population, an instructional budget, and an adequacy threshold before comparing material sets. For illustration, adequacy might require at least 80% of learners to reach a specified reasoning standard on parallel unfamiliar tasks. The percentage is a design example, not an established educational benchmark; its justification and the acceptable loss margin must be set for the actual context.

Planned removal comparisons must target relations as well as materials. Before comparison, construct a relation-to-carrier matrix: each target relation is linked to every poem, diagram, example, teacher prompt, and practice activity that could teach it. Removing one poem while another activity teaches the same relation tests that poem's contribution, not the relation's necessity. To test omission of a relation, remove or redesign its planned instructional routes while preserving other target content and documenting residual exposure. If that separation is infeasible, necessity remains untested. Learners' existing knowledge cannot be deleted; the estimand concerns omission from the specified instruction.

The time-filling rule is part of the comparison. If the removed material is replaced with a neutral activity that does not teach the target relation, the contrast examines omission within that stated activity budget. If the time is used for another relevant carrier, explanation, or practice opportunity, the contrast examines relative instructional value against that substitute. Neither route creates a pure absence-of-input condition; time, instructor attention, workload, and additional exposure should be documented rather than assumed equivalent. Omission and replacement comparisons should be reported separately.

An inclusion-minimal set is adequate while none of its tested proper subsets remains adequate under the same criteria. For a relation-level test, compare full instruction with a mapped omission condition; retain a separate carrier-replacement comparison when relevant. Check compensatory cues, overlapping relations, workload, and teacher explanations. A

meaningful loss after omission supports instructional importance within that tested pathway, not logical or universal necessity. One-at-a-time removals are insufficient when relations are redundant or interact; selected joint removals and replication would be needed. A precise loss within an acceptable margin supports local dispensability; imprecision leaves the question unresolved.

4.3 A micro-worldview and the position of the self

A micro-worldview is an unmeasured pedagogical heuristic for connecting aspects of a learner's minimal cognitive core across self, other people, environment, action, and uncertainty. The core names the learner's developing organization of usable relations; micro-worldview names a proposed way of discussing connections among them, not an additional psychological faculty or a scored outcome. Table 1 illustrates possible relations. The seven activities in Section 6 describe ways understanding may be exercised; they are neither components of a validated micro-worldview scale nor a developmental sequence.

Table 1. Candidate relations for an initial micro-worldview

Relation	Illustrative question
Observation and inference	What did I actually observe, and what did I add to it?
Action and consequence	What might change if the action or circumstances change?
Self and other perspectives	What could another person notice or value differently?
Evidence and uncertainty	What would help distinguish competing explanations?
Understanding and revision	What new information would require me to reconsider?

Beginning with the self means recognizing the learner as a situated observer and agent who experiences consequences. It does not grant personal preference authority over evidence or make other people peripheral. The educational task includes learning that one's perspective is partial and that others possess experiences, interests, and capacities to disagree. Personal agency and perspective-taking must therefore develop together.

The metaphor of a world model can help discuss connections between representations and experience, but an analogy with AI does not establish mechanistic equivalence. The present proposal makes no inference about human consciousness from AI architecture. It also leaves open whether one coherent micro-worldview is always desirable; productive understanding may require several context-sensitive models whose tensions are acknowledged rather than prematurely resolved.

4.4 Relationship to existing constructs and discriminating tests

Spiral curriculum design revisits a subject's underlying structures at increasing complexity (Bruner, 1960). Cognitive Seeds shares revisitation and progressive elaboration; its additional specification is to map candidate relations to all instructional routes and test local omission and generative use. This is an operational emphasis, not a property absent from spiral curricula.

Threshold concepts concern understandings that transform participation in disciplinary thinking and can be troublesome to acquire (Meyer & Land, 2003). A seed need not be a disciplinary gateway or a transformative threshold. Its selection is provisional and its instructional value can be tested without claiming a discontinuous transformation. Threshold-concept instruction can also promote transfer; transfer alone cannot distinguish the frameworks.

Big-ideas and backward-design approaches already connect central understandings, transfer, assessment, and instructional activities (Wiggins & McTighe, 2005). Cognitive Seeds does not claim to invent relation-centered curriculum or alignment. Its proposed increment is a compact, explicitly mapped starting set with context-bound dispensability tests and a separate test of learning from new resources. Whether those requirements improve a design must be evaluated.

Conceptual-change research examines how prior conceptions are restructured when learners encounter reasons to reconsider them (Posner et al., 1982). Cognitive Seeds shares attention to prior understanding and warranted revision, but does not offer an alternative account of the mechanism of conceptual change. It supplies selection and assessment commitments that can be implemented within such an account.

Cognitive Seeds is an integrative design framework, not an explanatory theory or a psychological kind distinct from schemas, mental models, or conceptual understanding. Its vocabulary specifies what designers should select, make observable, and compare. A substantive contribution would consist in better specified decisions, useful predictions about a stated design contrast, or improved outcomes beyond an adequate comparison. Renaming familiar teaching is not such a contribution.

Table 2 translates the commitments into comparisons. Recall-oriented and unconnected-content conditions can establish a contrast but are weak tests of added value. A stronger active comparison should also provide explanation, feedback, and transfer opportunities. If the

framework adds no practical guidance or meaningful benefit beyond such instruction, it should remain an organizing vocabulary rather than be promoted as an independent theory.

Table 2. Discriminating comparisons for the proposed design commitments

Design contrast	Expected observation	Interpretive limit
Recall-matched teaching with or without explicit boundary cases	The boundary-case condition better distinguishes when a learned relation applies and when it does not.	Supports boundary-case teaching; does not uniquely establish a seed mechanism.
Time-matched connected relation maps versus the same facts taught separately	Connected maps support more justified links and new questions on unfamiliar cases.	Tests relational organization beyond fact coverage; schema approaches may predict the same result.
Alternative carriers mapped to the same target relation	Learners can explain and appropriately revise the relation when the familiar carrier is replaced.	Examines carrier dependence; transfer across cultures still requires separate evidence.
The full design versus an active elaboration and transfer comparison	Prespecified outcomes improve beyond a well-supported active baseline at comparable cost and time.	An advantage supports incremental design value; identifying a distinct mechanism requires further analysis.

These comparisons also create failure conditions. Recall without transfer weakens the claim that a carrier has functioned as a seed; indiscriminate application weakens the claim of revisability; and no incremental benefit over a strong active comparison weakens the claim that the framework adds instructional value. None of these outcomes warrants denying that the learner has other forms of understanding.

A testable incremental prediction is that an explicit relation map plus planned replacement of familiar carriers will improve justified application to unfamiliar carriers and recognition of boundary cases beyond a specified active comparison. Match target relations, instructional time, examples, explanation opportunities, and feedback, and document the actual differences. Schema-based, big-ideas, and conceptual-change instruction may predict the same outcome. If those features already occur in the comparison, the proposed contrast is absent. A precise lack of advantage weakens the added-value claim; an imprecise estimate leaves it unresolved. Success would support the tested design package, not establish a unique seed mechanism.

5. Cultivation through progressive elaboration

The practical proposal is to revisit a limited selection of carriers while changing what the learner is asked to do. Early encounters can establish vocabulary and familiarity. Later encounters can invite explanation, comparison, alternative cases, reflection, and revision. Familiarity becomes a resource for inquiry rather than an endpoint.

Chi et al. (1989) associate self-generated explanations with understanding and reduced reliance on examples in a mechanics-learning context. Chi and Wylie (2014) distinguish modes of engagement through the ICAP framework, including constructive and interactive activity. These studies provide reasons to investigate learner-generated elaboration, but they do not validate Cognitive Seeds, poetry as its carrier, or a particular developmental schedule. The proposal must show what it adds to existing approaches.

Its candidate contribution is a design constraint: selected starting materials should have an explicit map of relations, opportunities for learner-generated elaboration, and planned occasions for revision and transfer. Without such a map, the intervention could simply rename familiar reflective teaching. Comparison studies must therefore specify whether the proposed benefit comes from content selection, repetition, explanation prompts, instructional time, or their combination.

The author describes a source curriculum using 26 Chinese quatrains across themes including emotions, values, nature, society, and science (Gao, 2024). Its six options—recitation, understanding, learning, appreciation, exploration, and thinking—illustrate progressive elaboration, not verified age-bound stages. Poetry is one candidate carrier; figurative meaning, cultural familiarity, translation, and interpretation accuracy can affect its suitability. Stories, pictures, games, and practical activities are alternatives, not demonstrated equivalents.

The six instructional levels organize teaching choices, while the seven activity dimensions in Section 6 describe candidate responses. They do not require a one-to-one mapping: an appreciation activity may involve connection, judgment, and reflection, and a relation diagram may involve observation, reasoning, and revision. These organizational schemes should not be interpreted as validated developmental stages.

In technical or non-poetic settings, a short program, a mathematical relation, a diagram, or a paradox could be considered as a carrier when it meets the same accessibility, elaboration, and revision criteria. For example, changing a program's input can test the limits of a rule. This is a proposal for alternative materials, not evidence that the framework generalizes across media or disciplines.

The Three Little Pigs illustrates how cognitive and values-education objectives must be separated. A cognitive task might ask learners to compare explanations of a building's stability, vary resources, or identify evidence that could distinguish alternatives. A values discussion might consider responsibility, fairness, or cooperation. These objectives need separate instructional statements and evidence. A reasoned moral disagreement with the teacher must not lower a reasoning score; agreement with a preferred moral is not evidence of understanding a causal relation.

Progression also requires adaptation. A novice unable to explain an event may need demonstration or vocabulary support before being asked to generate alternatives. Extra friction can exhaust rather than assist. The framework therefore treats challenge as something to calibrate against task demands and prior understanding, not as an educational good in itself.

6. From proposed capacities to observable activity

Seven verbs organize candidate observations: observe, connect, question, reason, judge, reflect, and revise. Table 3 offers task-specific operational candidates. They describe activities, not seven established faculties or the measured structure of a micro-worldview. Each requires an elicitation opportunity and task-specific checks. A behavioral indicator concerns performance under stated conditions, not direct access to an internal cognitive state.

Table 3. Candidate activity dimensions and observable indicators

Dimension	Working definition	Candidate indicator
Observe	Attend to relevant features and distinguish observation from inference.	Separates stated evidence from an added assumption.
Connect	Relate information through a meaningful structure.	Explains a relevant similarity and its limits in a new case.
Question	Identify a specific gap or uncertainty worth investigating.	Asks for information that could distinguish explanations.
Reason	Develop an inference from stated grounds.	Links premises to a conclusion and considers a counterexample.
Judge	Assess a claim using relevant criteria and evidence.	Identifies a specific unsupported inference without rejecting sound claims.
Reflect	Examine one's own interpretation, strategy, or confidence.	Names an assumption and explains how it influenced the answer.
Revise	Change an interpretation in response to relevant grounds.	Explains what changed, why, and what remains uncertain.

Metacognition supplies an important connection to the last two dimensions. Flavell (1979) distinguishes knowledge about cognition from experiences of monitoring it. A learner's statement of confidence, an account of a strategy, and an actual correction should

accordingly be recorded separately. Reporting reflection is not necessarily accurate monitoring, and changing an answer is not necessarily a justified revision.

The dimensions can overlap. Identifying a flawed causal inference may involve observation, questioning, reasoning, and judgment. Until their measurement properties are established, summing them into one score could conceal such overlap and create unjustified precision. A preferable initial report presents task scores and qualitative evidence by dimension, explicitly marking activities that a task did not elicit.

No empirical study presented here establishes discriminant validity for these seven dimensions. The first studies should report dimension-specific evidence and scores without a seven-dimension total. Code a dimension as NE, not elicited, when the task provides no meaningful opportunity to demonstrate it. For example, a one-shot evaluation with no initial response or subsequent new evidence cannot adequately elicit revision and should receive NE for that dimension, not zero. A separate planned step supplying conflicting evidence can make revision assessable.

When the task does elicit the activity and a valid response shows no relevant demonstration, a rubric score of zero may be appropriate. Missing responses, technical failures, and not-elicited activities require distinct codes and prespecified handling. When the same passage supports more than one dimension, coders should retain the shared evidence identifier and avoid treating the resulting codes as independent observations. A narrower task-specific endpoint can combine explicit reasoning criteria; that does not validate a global cognitive-capacity score.

A preliminary rubric can distinguish no relevant demonstration, a partially justified demonstration, and a coherent demonstration supported by task-specific evidence. It needs examples of acceptable alternatives and common false positives. Assessors should score anonymized responses without knowing the experimental condition where feasible, independently double-score a planned subset, and document disagreement. Agreement between assessors would establish consistency of scoring, not the validity of a theory of cognition.

Nonverbal evidence also needs task-specific anchors. For a tower-stability task, learners could choose between two unfamiliar base configurations, construct a stable alternative after a failure, and point to or demonstrate the changed relation. A candidate rubric should specify which choices and changes count as relevant, which can be explained by copying or motor skill alone, what assistance is permitted, and when the task did not elicit the intended activity. A successful construction is not automatically evidence of causal understanding. Scores from different response modes should not be assumed interchangeable without checking their demands and interpretation.

7. Implications for interaction with generative AI

The framework proposes that a learner should sometimes formulate an initial interpretation before consulting AI, then compare the response with that interpretation and test the resulting understanding in another setting. This sequence creates opportunities to make assumptions and revisions visible. It does not establish that prior formulation is

necessary for every useful AI interaction, or that an unrestricted interface always prevents learning.

The prior-formulation, AI-comparison, and independent-application sequence remains a design hypothesis. Self-explanation research (Chi et al., 1989) motivates eliciting a learner's own explanation. Bastani et al. (2025) motivates distinguishing a general-purpose assistant from a teacher-informed tutor and measuring unassisted outcomes. Neither study directly tests the sequence proposed here. An immediate-access comparison must therefore use a specified AI role, not be equated with every form of unrestricted generative AI use.

Kapur (2016) distinguishes productive and unproductive combinations of initial performance and later learning. The relevant rationale is that an initial generation attempt can prepare learners to benefit from subsequent instruction, even when the attempt is unsuccessful; subsequent instructional consolidation matters. This does not establish that any delay, difficulty, or failed attempt is useful. Prior formulation here need not involve failure, and a live AI response may not supply effective consolidation. Thus productive failure motivates a hypothesis and checks on instructional support, workload, and prior knowledge; it does not validate this AI sequence.

Three outcomes deserve separate attention: the quality of the assisted artifact, the quality of the learner's contributions during interaction, and subsequent performance on a comparable task without the same assistance. A fluent final response can coexist with weak independent explanation; an imperfect assisted response can coexist with useful questioning and correction. These combinations make a single productivity metric inadequate for the educational question.

Independent tasks must still be defined relative to the target capability. A no-AI assessment can examine whether participants retain an argument structure or detect an unsupported inference, while allowing ordinary accessibility supports that do not supply the tested reasoning. An AI-free task cannot establish total cognitive autonomy. Conversely, performance with a tool may be valuable even if the learner is not expected to reproduce every operation unaided.

For adults, the emphasis may be on exercising and revising existing understanding rather than constructing an initial educational foundation. For children, content, support, language, and consent procedures require separate consideration. Evidence from an adult analytical pilot cannot validate a preschool curriculum or establish developmental benefits. The relationship between these applications is a research question, not an assumed generalization.

8. Research propositions and a staged evaluation agenda

8.1 Four propositions

Proposition 1 concerns relational elaboration. Under comparable instruction, explanation, comparison, and revision opportunities are expected to support better independent application than carrier reproduction alone. A stronger test compares the fully specified package with existing active instruction. No practically meaningful advantage with adequate precision would weaken its incremental instructional-value claim in that setting.

Proposition 2 concerns compactness. A compact, connected starting set may support integration when it supplies adequate domain knowledge. Removing material that learners need should weaken rather than improve application. This proposition rejects an unrestricted claim that less knowledge is better and requires comparison with a well-designed broader curriculum.

Proposition 3 concerns interaction structure. Prior formulation and explicit comparison with an AI response may improve specified subsequent reasoning activities relative to direct AI use. The effect could be absent, negative, or limited to learners with sufficient prior knowledge. A time-matched preparation condition is needed to distinguish additional thinking time from the particular structure of the intervention.

Proposition 4 concerns generative use. After initial instruction, learners should be able to ask a diagnostic question, use new information to learn an extension, and revise only when the evidence warrants it. Independent transfer without improved learning from new resources would support a narrower application claim, not the fuller generative claim. Rigid repetition, indiscriminate revision, or gains confined to a familiar response template would count against the proposed interpretation. A lack of incremental advantage on PFL (Preparation for Future Learning) tasks—even when accuracy is already sufficiently high—serves to weaken the claim for generative use.

8.2 Feasibility before effectiveness

An internal walkthrough with two or three willing reviewers checks readability, timing, elicitation opportunities, AI fidelity, and scoring disagreements. It cannot estimate effectiveness or stable reliability. Reviewers familiar with intended flaws may test usability, but their detection scores are not unbiased outcome evidence. Repeating a case across conditions likewise introduces exposure effects. Keep calibration responses separate from later outcome data and record changes before freezing materials, prompts, anchors, and planned contrasts.

Curriculum evaluation and AI interaction evaluation should initially be separated. A curriculum study can compare teaching methods with the same carriers, content, and instructional time. A later comparison can examine alternative carriers or curriculum breadth. An interaction study can hold the learning material constant while comparing direct AI use, time-matched unstructured preparation followed by AI, and structured prior formulation followed by AI. An eventual adequately powered factorial design could examine how curriculum and interface features interact.

8.2.1 Interaction conditions and the tutor's role

An illustrative adult comparison uses common three-minute reading followed by a ten-minute interaction window. Immediate access provides ten minutes with the specified AI tutor. Free preparation provides four minutes of unrestricted private notes without AI, then six minutes with the tutor. Structured preparation provides four minutes to state an initial explanation, distinguish evidence and assumptions, identify an uncertainty, and name evidence that could change the explanation, then six minutes with the same tutor. All arms receive the same final-response instruction; the two preparation arms also receive the same

comparison instruction after AI access. Timings remain calibration choices, not an adopted protocol.

The primary structure contrast is structured versus free preparation, matching preparation time, AI exposure, final instructions, and materials. Record spontaneous structuring in free notes rather than prohibit it. Immediate access estimates an overall package contrast that also changes AI exposure; it cannot isolate friction. Log actual active time, tutoring turns, reading time, interruptions, and deviations. Equal scheduled time does not ensure equal cognitive effort. A short workload rating is supplementary and cannot by itself identify a mechanism.

For this comparison, the AI is a reasoning tutor rather than an answer generator. A candidate common instruction is: "Help the learner examine the report using questions and requests for justification. Distinguish reported facts from missing information. Do not invent allocation procedures or other facts. Do not provide a ready-made critique or reveal the target flaw. Ask the learner to consider plausible alternatives and the limits of the conclusion; acknowledge uncertainty." This instruction requires auditing and testing; it does not guarantee compliance. Use the same tutor instructions and capabilities in all arms, disable browsing and external tools in the bounded task, and preserve complete task-relevant conversations.

A separate operator manual must fix the allowed user prompts, turn or time limits, resources, handling of clarification requests, and stopping rules. Any named interface cannot be assumed to be implemented by these general principles; any named procedure needs its own accepted specification. Internal walkthrough participants should inspect whether tutor behavior changes across the preparation conditions before an outcome pilot begins.

These timings specify a controlled interaction component, not a complete lesson. Any implementation in a longer class must document surrounding instruction and keep it comparable across conditions. No lesson timetable or related essay substitutes for a frozen experimental protocol.

Use short adult analytical reports with enough information to distinguish observations from causal and adoption claims. Include both an overreaching report and one with a justified, appropriately limited conclusion. Lack of reported randomization is missing information, not proof of selection bias. An immediate test can be informative without establishing durable learning. Pretest readability and difficulty with people outside the outcome sample. Use parallel baseline items without disclosing the study case's target inference, and record prior domain knowledge, reading proficiency, and AI experience.

8.2.2 Independent tasks and follow-up

The first delayed endpoint is a new task approximately 24 hours later without generative AI. Change surface content and response format while retaining the reasoning demand. Section 8.3 describes the non-isomorphic format; Section 8.4 adds preparation for future learning. These are different endpoints: transfer asks what can be done on a new problem, whereas preparation for future learning asks how participants use a subsequent learning opportunity.

A follow-up around one week later can add durability information, but no interval alone establishes retained understanding. Fix allowable timing windows, reminders, accessibility supports, and resource restrictions before collection. Record deviations and possible intervening practice; remote compliance cannot simply be assumed. Report attrition by condition and prespecify missing-data analyses. Neither a 24-hour nor a one-week difference establishes general cognitive preservation, augmentation, or substitution.

8.3 Measurement and interpretive boundaries

For the proposed interaction study, the primary endpoint is the prespecified, task-specific 24-hour non-isomorphic transfer score; the primary contrast is structured versus free preparation. A preparation-for-future-learning score is an additional endpoint, not a post hoc replacement for an unsuccessful primary test. Immediate output quality, false criticism, confidence, workload, protocol fidelity, and later follow-up are secondary or process measures. Preregister the endpoint hierarchy and multiplicity plan before an outcome study; the present article provides no validated scoring instrument.

8.3.1 Assessment alignment and format familiarity

A candidate transfer task presents a new domain with two causal diagrams, alternative explanations, and possible follow-up comparisons. Participants choose which claim the report supports, select a diagnostic comparison, and predict a result under a changed condition, giving brief reasons. Instructions do not repeat the preparation categories or ask for the same four-part template. A task-specific 0–6 score can assign 0–2 for each of three targets: correctly limiting the inference, giving a case-grounded explanation or prediction, and selecting evidence that distinguishes alternatives. Each target requires its own correct, partial, and incorrect anchors. These targets share the intended reasoning content with instruction; changing format does not eliminate all alignment.

Before collection, all groups should receive equal practice with the assessment's diagram and selection format using unrelated content and no instruction in the target inference. This reduces format unfamiliarity without teaching the test solution. Use blind review to check that neither wording nor scoring rewards reproduction of the structured prompt. Freeze task-specific alternatives, sound-claim anchors, and opportunity checks after calibration. A bounded 0–6 score is not a seven-activity total; similarly structured scores across topics are not automatically interchangeable. Prespecify double scoring and adjudication.

8.3.2 Exposure, justified revision, and study controls

Separate report-flaw detection from AI-output evaluation. For detection rates, define diagnostic flaws and sound claims in advance, use justified identifications as the numerator, and state the opportunity denominator. A common audited AI response containing a controlled error, paired with a sound response, can test evaluation under equal exposure; it does not estimate a model's natural error rate. Keep this controlled stimulus separate from live tutor conversations. Expert review of live outputs should record natural errors and direct disclosure of the target flaw, including wording, turn, timestamp, and whether disclosure preceded the participant's identification.

To distinguish justified revision from copying or deference, retain a time-stamped initial response, the AI text actually seen, and a final response with reasons for changes. In a separate common-response check, present evidence that warrants a change on one item and an incorrect AI recommendation on another; score both evidence-sensitive revision and justified resistance. Add a new application item requiring a prediction or counterexample that the AI has not supplied. Verbatim or near-verbatim uptake is an exposure flag, not proof of absent understanding; changing one's answer or paraphrasing the model is likewise insufficient. Blind raters should assess the grounds and the new application separately from surface similarity.

Direct tutor disclosure is a protocol-fidelity event, not an automatic participant exclusion. Prespecify a primary analysis retaining randomized participants in their assigned conditions and descriptive disclosure rates, with clearly labeled sensitivity analyses if justified. Because disclosure is a post-assignment event, conditioning on it can bias a causal comparison. Log it rather than silently relabel the trial as successful mediation.

An exploratory pilot should prioritize completion, usability, fidelity, score distributions, and uncertainty, with its sample justified by those goals. Confirmatory effectiveness requires a separate justified sample-size calculation, a locked primary contrast and endpoint, random assignment or another defensible identification strategy, and prespecified exclusions and missing-data handling. Use baseline parallel-task performance and domain knowledge for balance checks and prespecified adjustment; do not select covariates after seeing favorable results. Calibrate task difficulty, avoid ceiling and floor effects, and keep calibration participants out of the confirmatory sample.

A delayed score difference is an observation. Calling it retention, augmentation, or substitution requires additional support. Prompt alignment, prior knowledge, AI disclosures, exposure, motivation, and task difficulty remain alternative explanations. If structured preparation does not outperform time-matched free preparation with adequate precision, the structure-specific claim is weakened. If gains disappear when format changes or merely reproduce AI wording without successful application, the broader interpretation is weakened. Imprecise null estimates establish neither absence of benefit nor equivalence.

Permit favorable, unfavorable, mixed, and inconclusive results. Report uncertainty and process deviations alongside outcomes. Freeze model identity or service version, date, system and user instructions, settings where exposed, and tools enabled. Use separate sessions without cross-participant memory and record any provider update. If the model cannot be pinned, hold collection windows narrow, log changes, and prespecify sensitivity analyses or a pause rule. Do not pool a substantive model change silently. A model-mediated result applies to the tested instructional arrangement; it is not a universal property of AI.

8.4 Preparation for future learning and weakening results

Bransford and Schwartz (1999) argue that transfer assessment should include how prior learning prepares people to learn from later resources, not only what they can reproduce unaided. Here, preparation for future learning is an additional endpoint: using unfamiliar information to extend or revise a relation after initial instruction. It is conceptually distinct from the primary delayed-transfer score. It does not measure general readiness to learn.

A proposed task first presents a new domain and a short baseline prediction, then offers a fixed menu of information sources under an equal selection budget. Sources include diagnostic comparisons, useful background, and plausible but nondiagnostic details; their accuracy and relevance are independently checked. Record which information the learner requests and why. All learners subsequently receive the same concise new lesson, including a boundary condition not supplied in the original teaching. They then solve a different case without AI, predict an outcome, explain how the new information matters, and state when their initial relation would need modification. Alternate forms and counterbalancing are needed to limit item and order effects.

Candidate secondary scores are relevance of the information choice, integration of the new lesson in a new prediction, and warranted revision at the boundary case, each with task-specific anchors. Record before-lesson performance, but do not treat an unadjusted gain score as a causal mechanism. A future outcome study should prespecify between-condition contrasts, baseline adjustment, acceptable alternative answers, and uncertainty reporting. Pilot item difficulty and score interpretation separately; a composite should not be assumed valid merely because its parts are coded 0–2.

The fuller generative claim would be weakened by accurate delayed reproduction accompanied by poor resource use, no meaningful advantage in learning the extension, failure to recognize its boundary, or an advantage confined to the original carrier or template. With adequate precision, no added benefit over active instruction would weaken the framework's incremental-value claim. Imprecise results remain inconclusive. Even favorable preparation-for-future-learning results would support the tested instructional design and target relations, not uniquely identify a cognitive seed mechanism.

9. Limitations and ethical considerations

The framework integrates existing ideas and may add no explanatory power beyond them. The empirical issue is whether its explicit relation maps, omission rules, and evaluation commitments improve a specified design or outcome. Its micro-worldview remains an unmeasured organizing heuristic; task-level findings cannot validate that broad network. Favorable results would support tested components or a design package, not demonstrate a distinct mental object named a cognitive seed.

A second limitation concerns content selection. A supposedly minimal foundation can reproduce the selector's cultural assumptions or moral preferences. Materials should expose alternative interpretations and disclose normative choices. Learners' disagreement should not be coded as cognitive failure. Multiple localized implementations may be preferable to one centrally defined canon.

Cross-language implementation must revalidate the mapping from carrier to target relation, including implied values, figurative meanings, familiarity, and possible alternative interpretations. Translating a Chinese quatrain does not establish that it elicits the same relation or level of difficulty elsewhere. Local co-design, interpretation checks, and adapted examples should precede comparisons. Claims of cross-cultural generality would require evidence that the tasks and scoring criteria remain interpretable across the populations studied.

Third, a demand for explicit explanation can favor verbal fluency and overlook embodied, practical, or nonverbal understanding. Assessment should allow appropriate demonstrations, diagrams, or oral responses and investigate language and accessibility effects. A learner's inability to produce the expected wording cannot by itself establish lack of understanding.

Any participant study should obtain appropriate institutional ethics review or a documented determination before recruitment, secure informed consent, and minimize identifiable data. Studies involving children require developmentally appropriate assent and guardian permission where applicable. Planted errors should appear only in approved, low-stakes materials and be explained afterward. Interaction logs should record task-relevant actions without claiming to reveal private mental states. Participation in discussion does not imply agreement to recruitment, recording, or an evaluator role.

A child-specific feasibility starting point could use a small group, one brief session, and no AI, examining whether instructions and materials are understandable and whether the activities can be completed. Record assistance, discomfort, and reasons for stopping. Such a walkthrough does not estimate educational effects or establish developmental stages, and it should precede considering an AI interaction study with that age group.

10. Conclusion

Cognitive Seeds is an integrative design framework for specifying relations, instructional carriers, and evidence of learner use. Its distinctions prevent recall from being equated with understanding, a supplied knowledge set with a learner's cognitive core, or an assisted product with independent capacity. Micro-worldview remains an unmeasured pedagogical heuristic rather than an additional outcome.

The next step is a bounded feasibility check of the relation map, materials, tutor instructions, non-isomorphic tasks, and scoring anchors. Later studies can test independent application and learning from new information against credible existing instruction. No new mechanism or educational benefit is established here. The framework should be retained, simplified, or rejected according to whether its explicit commitments add useful guidance and withstand those comparisons.

Declarations

Data availability: This conceptual paper reports no newly collected participant data. The examples and proposed protocols are illustrative.

Ethics: No human-subject study is reported. The research agenda describes procedures for possible future studies, not an approved or completed experiment.

Funding: No funding is reported.

Author's Position and Interests: This framework was proposed by John Gao, whose works on foundational education are authored under the pseudonym "Si Jiu". The author maintains an academic/intellectual interest in the development of this proposal.

AI Assistance: OpenAI's ChatGPT model was utilized during this revision process to assist with drafting the English text, refining concepts, conducting literature searches,

verifying references, and preparing the document. No participant observation was conducted, and no new data or research findings were generated. This statement does not imply the introduction or evaluation of any specific "study tutor" mechanism.

Literature and Sources: The literature search and synthesis process was selective rather than systematic. The selected literature is based on conceptual relevance, as outlined in Section 2. The revised interpretations of the literature, as well as any newly added references, have been verified against records published by the respective publishers, relevant institutions, or the author himself. John Gao bears responsibility for the assumptions upon which this framework is built, as well as for its final arguments.

References

1. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), e2422633122. https://doi.org/10.1073/pnas.2422633122
2. Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. Review of Research in Education, 24, 61–100. https://doi.org/10.3102/0091732X024001061
3. Bruner, J. S. (1960). The process of education. Harvard University Press.
4. Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1
5. Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational Psychologist, 49(4), 219–243. https://doi.org/10.1080/00461520.2014.965823
6. Clark, A., & Chalmers, D. J. (1998). The extended mind. Analysis, 58(1), 7–19. https://doi.org/10.1093/analys/58.1.7
7. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist, 34(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906
8. Gao, J. (2024). New Education: Cultivating Seeds of Cognition (English only): ChatGPT Era (CS1). Amazon. https://www.amazon.com/gp/product/BODM6HJT8N
9. Gao, J. (2025). A Guide for Planting the Future of Education: Cognitive seeds textbook toolkit (CS1 4). Amazon. https://www.amazon.com/gp/aw/d/B0FBKGMCHM
10. Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Educational Psychologist, 51(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

11. Meyer, J. H. F., & Land, R. (2003). Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines (Occasional Report 4). Enhancing Teaching–Learning Environments in Undergraduate Courses Project, University of Edinburgh. https://pressbooks.atlanticoer-relatlantique.ca/app/uploads/sites/803/2021/07/ETLreport4.pdf
12. National Academies of Sciences, Engineering, and Medicine. (2018). How people learn II: Learners, contexts, and cultures. The National Academies Press. https://doi.org/10.17226/24783
13. Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education, 66(2), 211–227. https://doi.org/10.1002/sce.3730660207
14. Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
15. Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. Perspectives on Psychological Science, 10(2), 176–199. https://doi.org/10.1177/1745691615569000
16. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
17. Wiggins, G., & McTighe, J. (2005). Understanding by design (Expanded 2nd ed.). Association for Supervision and Curriculum Development.

AI GOVERNANCE IN REVIEW: COMPETING DOCTRINES, LEGAL ARGUMENTS, AND EMERGING RESEARCH

Kildeev Adel
Leningrad State University named after A.A. Zhdanov (1990)
PhD in Law
ORCID: 0009-0001-2211-4944

Crossref DOI: Pending Registration
SSRN DOI: https://doi.org/10.2139/ssrn.7607562

Abstract: This article presents a selective critical review of AI governance scholarship circulated through the SSRN Legal Scholarship Network during September 2026. From a corpus of more than one hundred research entries identified in the monitored SSRN alerts, the author selects a limited number of works considered particularly significant for their legal arguments, competing doctrinal approaches, methodological implications, and potential contributions to the development of AI governance. The review is deliberately selective and does not purport to provide an exhaustive survey of SSRN publications or AI governance scholarship beyond the specified period.

The analysis examines competing approaches to legal attribution, artificial intelligence accountability, human oversight, professional responsibility, and institutional governance. Particular attention is given to the originality and coherence of legal arguments, the reliability of cited authorities, the adequacy of supporting evidence, and the distinction between technological capabilities and legally attributable responsibility.

The article evaluates both the contributions and limitations of the selected works, identifying areas of doctrinal convergence, unresolved disagreements, methodological weaknesses, and emerging research directions. Its purpose is not to establish a hierarchy of scholarly publications but to advance a substantive, evidence-based academic discussion grounded in critical examination of original texts and verifiable sources.

Keywords: Artificial Intelligence Governance; Legal Scholarship; Comparative Legal Analysis; AI Accountability; Human Oversight; Legal Responsibility; Scholarly Verification; Research Integrity; SSRN; Instrumentum Vocale.

The use of the generative models ChatGPT (OpenAI), Gemini (Google), Grok (xAI) in drafting this article illustrates, rather than undermines, its central thesis: LLMs function as instrumenta vocalia—tools assisting in linguistic production without possessing understanding, intention, or authorship. Their involvement does not alter the locus of legal or intellectual responsibility, which remains exclusively with the human author, who makes all substantive, structural, and evaluative decisions. LLM outputs cannot be accorded independent legal or intellectual status apart from the human author.

Chapter I. Introduction: Scope and Methodology of the Review

The present review examines a deliberately limited selection of academic works on artificial intelligence governance circulated through the SSRN Legal Scholarship Network in September 2026. During this period, more than one hundred research entries were identified in the monitored alerts. However, the purpose of this article is neither to catalogue these publications nor to provide a comprehensive assessment of all research appearing on SSRN.

Instead, the review concentrates on selected works that warrant closer examination because of their doctrinal significance, originality of argument, methodological approach, or relevance to unresolved questions of legal responsibility and AI governance.

The September 2026 timeframe refers to the circulation of research through the monitored SSRN alerts, rather than necessarily to the original publication dates of the individual works. Accordingly, the review distinguishes between original publication, subsequent SSRN dissemination, and inclusion in the monitored September alerts.

The selection is qualitative rather than statistically representative. Its purpose is to enable a focused comparative examination of substantive legal arguments, competing theoretical frameworks, and the evidentiary foundations upon which proposed AI governance models rest.

Selection does not imply endorsement; omission does not imply rejection. The scope of a critical review is determined by its stated research purpose, not by expectations of authors whose works fall outside its selection.

Chapter II. Legal Personality, Agency and Attribution

2.1. Distinguishing Legal Personality, Attribution, and Responsibility

One of the most consequential methodological difficulties in contemporary AI governance scholarship is the tendency to treat legal personality, functional autonomy, attribution, and responsibility as closely related aspects of a single problem. Although these concepts intersect, they perform different functions within legal doctrine and cannot be substituted for one another without altering the legal question under examination.

Legal personality determines whether an entity may occupy a legally recognized position as a bearer of rights and duties. Attribution concerns the legal connection between particular conduct, an institutional arrangement, and a recognized actor. Liability concerns the conditions under which legal consequences may be imposed upon that actor. Judgment raises a further question: whether a determination can be recognized as the exercise of legally relevant evaluative authority.

The distinction is not merely terminological. A system may produce an output that has legal consequences without becoming a legal person. A company may incur liability for the deployment of an automated system without necessarily being regarded as the intellectual author of every individual output. Equally, the technical capacity to execute an operation cannot, by itself, establish the legal authority to make a binding decision.

The selected literature reveals substantial agreement on the need to separate these categories. The more significant disagreements arise over what follows once the separation has been made.

2.2. Attribution of Judgment Versus Allocation of Consequences

Jon Polenberg's *Judgment Without a Subject: Artificial Intelligence and Legal Attribution* offers one of the more conceptually ambitious approaches to this problem.[1]

Its principal contribution is the distinction between the legal allocation of consequences and the attribution of judgment. The former identifies the person or institution upon whom the law imposes responsibility for an outcome. The latter asks whether the determination itself can be recognized as an act of judgment attributable to a legally identifiable subject.

Polenberg develops this distinction through philosophical accounts of personhood and their implications for different legal institutions, including copyright, corporate law, civil liability, employment law, criminal responsibility, and evidence.[2]

The argument is particularly relevant where machine-generated determinations are presented not merely as technical information but as conclusions upon which legal institutions are invited to rely.

The distinction exposes a genuine conceptual difficulty. Assigning liability to a manufacturer or deployer does not necessarily establish that the manufacturer or deployer personally exercised the judgment represented by a particular machine-generated conclusion.

Nevertheless, two separate questions remain.

First, must every legally admissible machine-generated conclusion constitute judgment in the sense proposed by Polenberg? Secondly, can the law recognize the evidentiary significance of an automated determination without simultaneously recognizing the generating system as a subject of judgment?

These questions are particularly important because legal systems routinely distinguish between evidence produced by technical instruments and judgments rendered by persons exercising legal authority.

A machine-generated measurement, classification, or probabilistic assessment may be relevant evidence without possessing independent juridical authority. Its admissibility and weight can be evaluated through rules concerning authenticity, reliability, methodology, and procedural fairness. The final legal judgment remains attributable to the competent human or institutional decision-maker.

Accordingly, Polenberg's warning against the uncritical promotion of automated output into legal authority deserves serious consideration. However, the argument requires a careful

[1] Polenberg J. Judgment without a Subject: Artificial Intelligence and Legal Attribution. – 2026. – 58 p. – DOI: 10.2139/ssrn.7248440. – URL: https://ssrn.com/abstract=7248440 (accessed: 11.10.2026).

[2] Polenberg J., pp.11-49.

distinction between recognizing machine-generated information as evidence and treating the machine itself as an independent juridical decision-maker.

The former does not necessarily entail the latter.

This qualification does not undermine Polenberg's central distinction. Rather, it identifies the doctrinal boundary at which that distinction becomes most useful.

2.3. The Attribution Vacuum and the Problem of Principal Displacement

Northon Salomão de Oliveira approaches the same general problem from a different direction.

In *Autonomous AI Agents and Legal Personality: Agency, Attribution, and the Limits of Artificial Intelligence Liability*, he explicitly rejects the proposition that granting legal personality to autonomous systems would resolve practical difficulties of accountability.[3]

Instead, he develops the concept of an *attribution vacuum*, understood as a situation in which the functional autonomy exercised by an AI system exceeds the scope of authority specifically manifested by an identifiable human or corporate principal.

His proposed Delegated Autonomy Attribution Standard (DAAS) distinguishes three forms of AI-mediated conduct: scripted attribution, bounded discretionary attribution, and displaced discretionary attribution. The framework seeks to connect increasing operational discretion with appropriate liability doctrines while avoiding the creation of a new legal subject.[4]

This approach has a notable practical advantage. It directs legal analysis toward the conditions of authorization, deployment, monitoring, and institutional responsibility rather than toward speculative questions of artificial personhood. The difficulty lies in the characterization of principal displacement as an attribution vacuum. A distinction must be drawn between the absence of specific authorization for an individual machine-generated act and the absence of a legally relevant human or institutional actor.

These conditions are not equivalent.

Where a corporation deliberately deploys an autonomous system, establishes its operating environment, authorizes access to resources, defines permissible objectives, and derives economic benefits from its operation, the absence of prior approval for every individual output does not necessarily eliminate the legal relevance of those earlier decisions.

The question may concern the scope of a duty of care, the foreseeability of harm, the adequacy of supervision, the allocation of enterprise risk, or the applicable standard of liability. None of these questions requires treating the instrument as a legal subject.

---

[3] Oliveira N. S. de. Autonomous AI Agents and Legal Personality: Agency, Attribution, and the Limits of Artificial Intelligence Liability. – 2026. – 19 p. – DOI: 10.2139/ssrn.7355821. – URL: https://ssrn.com/abstract=7355821 (accessed: 11.10.2026).

[4] Oliveira N.S., pp.11-13.

Indeed, Oliveira's own proposed third tier assigns residual risks to the principal benefiting from the deployment. This is a substantive proposal for allocating responsibility to an identifiable institutional actor rather than evidence that such an actor has disappeared.[5]

The distinction matters because a perceived gap in existing remedies does not necessarily establish a gap in legal attribution itself.

The strongest version of Oliveira's argument is therefore not that operational autonomy produces an ontological absence of responsible actors, but that existing legal doctrines may require adjustment to distribute responsibility effectively among actors exercising different forms of control.

This interpretation also reflects an important limitation expressly recognized in the article: the proposed framework's highest-displacement tier remains an untested doctrinal construction rather than a standard demonstrated through decided cases.[6] Its value should consequently be assessed as a proposal for legal development, not as an established description of current law.

2.4. Emergent Behavior and the Limits of Causal Reconstruction

A related but distinct approach appears in Oliveira's earlier study, *Artificial Intelligence, Autonomous Systems and the Future of Legal Responsibility.*[7]

Here, the organizing concept is not principal displacement but the difficulty of reconstructing the causal pathway between an autonomous system's operation and a harmful outcome. The article proposes the Emergent Autonomy Liability Standard (EALS), combining risk-related liability, behavioral logging, and a compensation mechanism for harms that resist conventional causal reconstruction.

The central distinction is between harms whose causal development can be reconstructed and harms attributed to emergent system behavior. This proposal raises an important problem in AI governance: technical opacity may substantially increase the difficulty of proving how a particular outcome occurred, especially in systems whose behavior changes after deployment.

However, technical reconstructibility and legal causation are not identical concepts.

Legal responsibility does not invariably depend upon identifying the precise internal mechanism through which a harmful event developed. Depending on the applicable jurisdiction and cause of action, liability may arise through legally defined duties, presumptions, enterprise risk allocation, or other established mechanisms. Consequently, the inability to reconstruct a specific computational sequence does not, without further argument, demonstrate the inability of legal doctrine to identify a responsible actor.

---

[5] Oliveira N.S., p.12.

[6] Oliveira N.S., p.15.

[7] Oliveira N. S. de. Artificial Intelligence, Autonomous Systems and the Future of Legal Responsibility. – 2026. – 16 p. – DOI: 10.2139/ssrn.7305740. – URL: https://ssrn.com/abstract=7305740 (accessed: 11.10.2026).

There is also a conceptual tension between treating unforeseeable system behavior as evidence of a structural failure in liability law and proposing mandatory logging and institutional compensation as corrective measures. The proposed measures may be normatively justified. They could improve evidentiary access, distribute risks, and strengthen compensation. Yet their usefulness does not independently prove the claimed doctrinal impossibility. This is an important methodological distinction: a proposed reform may be valuable even where the diagnosis used to justify it is broader than the available evidence establishes.

Read together, Oliveira's two studies illustrate a recurring challenge in contemporary AI liability research. The identification of a difficult attribution or causation problem can lead rapidly to the assertion of a structural legal vacuum, even where the proposed solution itself continues to rely upon identifiable human or corporate actors.

The existence and precise extent of that supposed vacuum therefore require independent examination rather than acceptance as a foundational premise.

2.5. Human–AI Coupling and the Problem of Individual Judgment

Deborah B. Glosser's *The Chip of Theseus: Human-AI Coupling and the Erosion of Legal Individuation* introduces another dimension to the debate.[8]

Unlike approaches primarily concerned with autonomous systems acting externally, Glosser examines continuous interaction between human users and AI systems capable of influencing their decisions. The central problem concerns the difficulty of separating independently formed human judgments from recommendations supplied through persistent technological assistance.

Glosser distinguishes between the user's awareness of AI involvement, the external detectability of that involvement, and the attribution of an individual judgment to its source.[9] Her argument becomes particularly interesting where a legal institution assumes that an identifiable person's decision represents an independently formed act of judgment.

Contract formation, voting, and agency relationships provide different examples of this underlying concern. Yet the problem of cognitive influence should not automatically be equated with the disappearance of legal individuality. Human judgment has never developed in complete informational isolation. Individuals act through advice, education, professional consultation, institutional procedures, and other external influences. The law does not generally require proof that a person's decision originated independently of all such influences. Instead, particular doctrines establish conditions under which external influence becomes legally relevant.

AI-mediated cognitive influence may present new problems of intensity, continuity, transparency, or commercial control. These features may justify additional legal safeguards.

[8] Glosser D. B. The Chip of Theseus: Human-AI Coupling and the Erosion of Legal Individuation. – 2026. – 21 p. – DOI: 10.2139/ssrn.7279242. – URL: https://ssrn.com/abstract=7279242 (accessed: 11.10.2026).

[9] Glosser D. B., pp.1-2.

Nevertheless, their presence does not automatically invalidate the attribution of an act to the person who performed or legally adopted it.

Glosser's proposed safeguards, including disclosure mechanisms, temporal separation of automated suggestions, and evidentiary presumptions, are therefore more persuasive as possible governance instruments than as proof that established legal individuality has ceased to function.

Their practical effectiveness remains a separate empirical question.

2.6. Comparative Doctrinal Assessment

The examined studies identify different aspects of a common institutional problem.

Polenberg asks whether an automated determination can acquire the status of judgment without an attributable subject. Oliveira examines the difficulties of assigning responsibility where technical autonomy and causal uncertainty complicate existing doctrines. Glosser considers whether continuous machine assistance can weaken the practical distinction between human judgment and externally generated recommendations.

These approaches should not be treated as interchangeable theories.

Their most important common contribution is the demonstration that operational performance alone does not resolve questions of legal authority, attribution, causation, or institutional responsibility. Their principal disagreement concerns whether technological developments expose correctable deficiencies within existing legal frameworks or establish more fundamental failures in the conceptual architecture of law.

The distinction requires particular care. Technical autonomy is not necessarily legal agency. Causal uncertainty is not necessarily the absence of a responsible party. Cognitive influence is not necessarily the disappearance of human legal individuality. Equally, assigning financial consequences does not necessarily establish the intellectual or juridical authorship of an automated determination. The selected scholarship is strongest where it preserves these distinctions and identifies the specific legal operation requiring clarification or reform. It becomes more vulnerable where uncertainty within one category is treated as evidence of failure in another.

A coherent approach to AI governance must therefore maintain analytical separation between the technical capabilities of automated systems, the human and institutional decisions enabling their operation, the requirements of legal attribution, and the rules governing the consequences of harmful conduct.

This provides the doctrinal foundation for examining the next question: the extent to which automated systems can support, reproduce, or replace human judgment in judicial and administrative decision-making.

Chapter III. Human Judgment and Automated Adjudication

3.1. The Distinction Between Legal Prediction and Judicial Judgment

The increasing capacity of artificial intelligence systems to classify legal information, identify precedents, generate legal arguments, and predict judicial outcomes has revived a fundamental question concerning the nature of adjudication. If a computational system can produce a determination substantially similar to that of a human judge, does functional equivalence justify institutional substitution?

The selected scholarship suggests that this question cannot be resolved by comparing the accuracy of machine-generated and human decisions alone. Judicial adjudication is not exclusively an exercise in information processing. It is also a legally constituted procedure through which public authority is exercised, competing arguments are evaluated, and binding determinations acquire institutional legitimacy.

Eric Alston and William Lehr develop this distinction in *On (Human) Adjudication*, arguing that the difficulty of algorithmic adjudication becomes particularly apparent where the applicable legal rules depend upon contested or changing human values.[10]

Their argument does not depend upon demonstrating that artificial intelligence is incapable of performing useful legal analysis. On the contrary, the authors acknowledge significant developments in model capabilities, including structured internal representations, improved performance on reasoning benchmarks, and the production of novel solutions through systems combining generation with external evaluation.

The central distinction concerns the difference between applying an established evaluative standard and exercising authority to determine what that standard should mean in circumstances where its interpretation remains contested.

This distinction is significant because legal prediction and judicial judgment answer different questions.

A predictive system seeks to estimate an outcome based upon available information and identifiable patterns. A court must determine the legally justified disposition of a particular dispute within an institutional framework that authorizes, constrains, and permits challenges to the exercise of judicial power.

Predictive accuracy may improve the informational basis of adjudication. It does not necessarily establish the legitimacy of the resulting decision.

This position is consistent with, but not identical to, the argument developed by Jon Polenberg. In *Judgment Without a Subject*, Polenberg distinguishes the production of a determination from its attribution to a legally identifiable subject whose judgment it represents.[11]

The two approaches converge in resisting the automatic transformation of functional performance into legal authority. They differ in analytical emphasis. Polenberg concentrates

[10] Alston E., Lehr W. On (Human) Adjudication. – 2026. – pp.1-5 – DOI: 10.2139/ssrn.7334643. – URL: https://ssrn.com/abstract=7334643 (accessed: 11.10.2026).

[11] Polenberg J. Cited paper, pp.3-10.

on the attribution of judgment, whereas Alston and Lehr focus on the normative and institutional conditions under which judgment is exercised.

Taken together, they demonstrate why the question of whether an AI system can generate a legally plausible answer must be distinguished from whether that answer should possess the authority of a judicial decision.

3.2. Novelty, Imprecision, and Incommensurability in Legal Adjudication

Alston and Lehr identify three recurring categories of legal disputes in which the limits of algorithmic adjudication become particularly significant: novelty, imprecision, and incommensurability.[12]

Novelty arises when technological or social developments produce circumstances not specifically anticipated by existing legal rules. The authors illustrate this difficulty through *United States v. Jones* (2012), examining the application of established constitutional protections to new surveillance technologies.

The relevant problem is not simply the availability of precedents. It concerns the selection and interpretation of legal principles whose application to new circumstances may itself require authoritative judgment.

Imprecision presents a different difficulty. Here, the relevant facts may be substantially established, but the governing legal standard does not mechanically determine the outcome.

The authors use *Palsgraf v. Long Island Railroad Co.* (1928) to illustrate the problem. Competing interpretations of foreseeability and the scope of legal responsibility demonstrate that disagreement may concern the normative boundaries of a legal standard rather than uncertainty about the underlying facts.

Incommensurability arises when adjudication requires consideration of competing values that cannot readily be reduced to a common quantitative measure. Alston and Lehr discuss *United States v. Rahimi* (2024) to illustrate the interaction between constitutional interpretation and competing public interests.

The analytical significance of these examples lies in their common structure. Each concerns a situation in which additional information alone may be insufficient to determine the appropriate legal result.

A computational system may identify relevant precedents, summarize competing interpretations, or estimate how a court would resolve the dispute. Nevertheless, the selection of the legally authoritative interpretation remains a separate institutional act.

This analysis provides a useful corrective to proposals that treat sufficiently accurate prediction as a functional equivalent of adjudication.

At the same time, the three categories should not be understood as establishing that every legal dispute possesses the same degree of interpretive uncertainty. A significant proportion

[12] Alston and Lehr, pp.9-15.

of judicial and administrative work involves comparatively settled rules, repetitive procedures, and factual questions for which computational assistance may be valuable.

The relevant distinction is therefore not between legal work that machines can perform and legal work that humans can perform. It is between tasks involving the application of sufficiently determinate criteria and decisions requiring the authoritative resolution of questions that those criteria leave open.

3.3. Algorithmic Consistency and the Problem of Judicial Discretion

An important strength of the analysis by Alston and Lehr is its recognition that human judicial performance should not automatically be treated as an ideal benchmark.

The authors discuss experimental research by Posner and Saran comparing the responses of language models with patterns observed in human judicial decision-making.[13]

According to their account of these experiments, some more capable models followed formal legal rules more consistently than the participating human judges, whose responses were also affected by considerations not reducible to the stated legal criteria.

This finding complicates simplistic arguments concerning the superiority of human judgment. If a computational system applies a valid legal rule more consistently than a human decision-maker, its greater consistency may represent an institutional advantage rather than a deficiency.

The authors acknowledge this objection. Their argument is not that human judges necessarily produce more accurate, coherent, or desirable outcomes. Rather, it concerns the institutional difference between consistent rule application and the exercise of contextual discretion.

This distinction deserves particular emphasis. Judicial discretion can serve legitimate purposes where legal standards leave room for interpretation or where circumstances require the adaptation of general principles to particular facts. However, discretion can also introduce bias, inconsistency, and unequal treatment.

An adequate theory of AI-assisted adjudication must account for both possibilities.

The relevant objective cannot be the preservation of human discretion regardless of its consequences. Nor can it be the elimination of discretion merely because automated systems offer greater consistency. Instead, the institutional task is to identify the circumstances in which discretion is legally required, the constraints under which it should operate, and the procedural mechanisms through which its exercise may be challenged.

The experimental evidence discussed by Alston and Lehr contributes to this debate by demonstrating possible differences between human and machine decision patterns. Yet differences in observed behavior do not independently establish which pattern is normatively preferable.

[13] Alston and Lehr, pp.16-17.

Such experiments also cannot, by themselves, resolve questions about future technological capabilities or the proper constitutional allocation of judicial authority. Their strongest contribution is therefore comparative rather than conclusive: they expose a potential tension between formal consistency and context-sensitive legal judgment.

3.4. Public Authority, Procedural Legitimacy, and the Limits of Automation

A related institutional approach appears in J. G. Allen and Florian Gamper's *Artificial Intelligence, Liability and Agency*.[14]

Although their study addresses a broader range of responsibility and attribution questions, its discussion of public administration and judicial decision-making is particularly relevant to the present inquiry.[15]

The authors distinguish the substantive correctness of an administrative or judicial determination from the procedural legitimacy of the process through which it is reached. Within public law, an outcome cannot necessarily be justified solely by demonstrating that it is accurate or efficient. The decision-making process must also satisfy applicable requirements concerning authority, procedural fairness, public justification, and reviewability. This distinction reveals a structural tension between the optimization objectives commonly associated with machine-learning systems and the normative requirements of public administration. A system may produce an apparently correct result while leaving unresolved whether the relevant decision was reached through a legally acceptable procedure.

Allen and Gamper do not advocate the categorical exclusion of automation from public decision-making. Instead, they distinguish between determinations involving relatively predictable factual or administrative operations and decisions that create legal consequences through substantive normative judgment.

This approach provides a useful point of comparison with Alston and Lehr.

Both studies recognize a role for computational systems in supporting routine or sufficiently determinate operations. Both also identify limits to the substitution of automated processes for legally constituted human authority.

Their emphases nevertheless differ.

Alston and Lehr locate the difficulty primarily in the evolving and contested values underlying legal interpretation. Allen and Gamper approach it through established principles of administrative legality, procedural accountability, and the institutional requirements governing the exercise of public power.

The approaches are complementary rather than mutually exclusive.

---

[14] Allen J. G., Gamper F. Artificial Intelligence, Liability and Agency : SMU Centre for Digital Law Research Paper No. 02/2025. – 2025. – 18 p. – DOI: 10.2139/ssrn.7457561. – URL: https://ssrn.com/abstract=7457561 (accessed: 11.10.2026).

[15] Allen and Gamper, pp.13-16.

One addresses why judicial reasoning may resist complete reduction to predetermined criteria. The other addresses why even a technically competent automated process must operate within a legal structure capable of establishing its authority and reviewing its consequences.

3.5. Machine-Generated Evidence and the Attribution of Legal Conclusions

The distinction between technical reliability and judicial authority becomes particularly important when automated outputs enter legal proceedings as evidence.

Polenberg examines this question through the interaction between machine-generated evidence, evidentiary reliability, and the legal concepts of witnesses and declarants.[16]

His analysis of the proposed Federal Rule of Evidence 707 illustrates the difficulty of applying legal mechanisms developed around human testimony to determinations generated by computational systems. The proposed rule sought to address the reliability of certain machine-generated evidence through standards associated with expert testimony. As Polenberg records, the proposal was not adopted in its published form during the May 2026 consideration and remained subject to further study. The underlying difficulty nevertheless extends beyond the particular proposal.

Evidence may be produced by a technical system whose operation can be documented, tested, and evaluated. Such evidence does not necessarily express the judgment of a human witness in the conventional sense. The absence of a human declarant, however, does not automatically mean that machine-generated information lacks evidentiary value.

Legal proceedings routinely admit information obtained through instruments and technical processes. The appropriate procedural treatment depends upon the nature of the evidence, its purpose, the applicable rules, and the ability of the parties to challenge its reliability. The essential distinction is between establishing that an automated process produced a particular result and establishing that the result merits acceptance as a legally significant conclusion.

These propositions require different forms of justification.

Technical authentication may demonstrate provenance, integrity, or conformity with specified operating conditions. It does not necessarily establish the substantive correctness of the underlying conclusion. Likewise, successful performance against a benchmark does not invariably prove that a system is appropriate for the particular legal question before a court.

Polenberg's argument draws attention to the danger of treating procedural recognition of machine-generated evidence as an implicit recognition of machine judgment. The more defensible institutional approach is to preserve the distinction between an instrument producing evidence and a legally competent decision-maker evaluating that evidence.

[16] Polenberg J. Cited paper, pp.49-55.

This does not require rejecting automated evidence. It requires that its legal significance remain subject to applicable evidentiary standards, adversarial challenge, and authoritative human or institutional determination.

3.6. Automation, Human Oversight, and Institutional Responsibility

The selected studies point toward an important distinction between automating components of adjudication and transferring adjudicative authority itself.

Alston and Lehr propose a complementary role for AI in judicial practice, including information organization, support for routine decisions, and the identification of inconsistencies in comparable outcomes.[17]

Such applications may improve efficiency and assist human decision-makers in identifying relevant information. However, the introduction of automated assistance also creates institutional dependencies. The choice of computational models, training data, evaluation criteria, and operational constraints may influence the information presented to judges and the alternatives they consider. The authors correctly emphasize that these choices are not necessarily neutral. They can relocate significant elements of institutional discretion from the adjudicative process to the design and governance of the technological system.

This observation has implications beyond the courtroom.

Automated systems used in administrative decision-making may similarly influence the interpretation of rules, the classification of individuals, and the practical exercise of public authority. The existence of human oversight is therefore insufficient unless the oversight mechanism provides a genuine capacity to understand, contest, and, where appropriate, reject the system's output.

A nominal human approval requirement does not necessarily establish meaningful judicial control. At the same time, requiring human participation at every technical stage would risk eliminating many of the practical advantages that computational systems can provide. A more coherent institutional distinction concerns the nature and legal significance of the task being performed.

Where governing criteria are sufficiently determinate and legal authorization permits automation, carefully bounded computational processes may perform substantial functions without requiring an individual human judgment at every operational step.

Where legal standards remain contested, material rights are affected, or authoritative interpretation is required, the preservation of accountable and reviewable institutional judgment becomes particularly important.

Alston and Lehr themselves recognize the possibility of democratically authorized automation for narrowly specified categories of decisions, provided that the relevant

[17] Alston and Lehr, pp. 20-23.

arrangements remain transparent, challengeable, and subject to appropriate institutional controls.[18]

This qualification prevents their argument from becoming an absolute prohibition on automated adjudication. It also clarifies the central governance question. The objective is not to preserve every traditional human activity against technological substitution. It is to identify which legal functions may properly be automated, which require continuing institutional judgment, and what safeguards are necessary to maintain lawful authority over the resulting process.

3.7. Comparative Assessment

The selected publications provide a substantive foundation for distinguishing computational competence from judicial authority.

Alston and Lehr demonstrate the difficulties encountered when legal rules depend upon contested standards and evolving social values. Allen and Gamper emphasize the procedural and institutional requirements governing the exercise of public authority. Polenberg examines the implications of treating machine-generated determinations as legally significant conclusions without clearly identifying the subject of judgment. The approaches converge upon a common proposition: the technical production of a legal determination does not, by itself, establish the authority of that determination.

Nevertheless, their differences are equally important.

Alston and Lehr ground their argument in the nature of difficult legal decisions and the institutional location of normative discretion. Allen and Gamper emphasize accountability and procedural legality. Polenberg questions the conceptual foundations upon which automated outputs may be recognized as judgment. None of these approaches requires the rejection of computational assistance in legal proceedings. The more difficult question is where the boundary between assistance and legally authoritative decision-making should be drawn, and whether that boundary can be implemented in institutional practice rather than merely asserted in theoretical terms.

The reviewed scholarship identifies several necessary considerations: the determinacy of applicable rules, the nature of the rights affected, the availability of procedural safeguards, the capacity to challenge automated evidence, and the identification of actors exercising meaningful control over the relevant technological and legal processes.

These considerations suggest that the future of AI-assisted adjudication cannot be evaluated through technological performance metrics alone. It must also be examined through the legal architecture governing authorization, institutional accountability, procedural legitimacy, and the practical preservation of reviewable judgment.

The resulting research problem extends beyond judicial decision-making. Similar questions arise wherever professional or public institutions rely upon automated systems to produce conclusions that influence legally significant outcomes. This provides the basis for

[18] Alston and Lehr, pp.22-23.

the subsequent examination of professional responsibility, manufacturer liability, and the allocation of duties across the technological and institutional chains through which AI-generated outputs acquire practical legal effect.

Chapter IV. Professional Responsibility and Manufacturer Liability

4.1. The Problem of Distributed Responsibility in AI-Assisted Legal Practice

The integration of generative artificial intelligence into professional legal practice has exposed a fundamental tension between established duties of professional competence and the increasingly complex technological infrastructure through which legal services are delivered.

Traditional principles of professional responsibility presume that a licensed lawyer exercises independent judgment, supervises the preparation of legal documents, verifies relevant authorities, and assumes responsibility for representations made to clients and courts. The introduction of generative AI does not eliminate these obligations. It nevertheless complicates the institutional conditions under which they are discharged.

The selected scholarship approaches this problem from several distinct perspectives.

Drew Simshaw examines the professional and institutional incentives affecting the responsible adoption of legal AI. Michael J. Bryant questions whether emerging verification requirements impose disproportionate burdens upon smaller legal practices. Sana Sohail emphasizes the perceived imbalance between professional sanctions and developer accountability. Adel Kildeev develops a producer-liability approach that considers the independent responsibilities of those designing, marketing, and distributing legal AI systems.[19]

These approaches identify a shared difficulty: legal practice increasingly relies upon technological products whose internal design, information sources, limitations, and operational controls are not entirely accessible to the professionals expected to verify their outputs.

The resulting problem should not be reduced to a choice between lawyer liability and manufacturer liability. Professional discipline, civil liability, contractual obligations, product safety, and institutional governance constitute related but legally distinct mechanisms. Each requires its own analysis of applicable duties, the relevant conduct, causation, and the conditions under which responsibility may arise. A coherent framework must therefore distinguish the lawyer's responsibility for the legal act from the producer's potential responsibility for the instrument through which that act was prepared.

[19] **Conflict-of-Interest Disclosure.** One of the works discussed in this review was authored by the author of the present article. Its inclusion serves comparative and doctrinal purposes and should not be interpreted as an independent assessment or endorsement of that publication. Any separate evaluation of its scholarly quality should be conducted by an independent reviewer.

4.2. Professional Competence and the Limits of Technological Reliance

The American Bar Association's Formal Opinion 512, issued on 29 July 2024, provides an important reference point for contemporary discussion of lawyers' use of generative AI.[20]

The opinion addresses existing professional duties rather than establishing a separate comprehensive system of AI-specific civil liability. These duties include competence, confidentiality, communication with clients, supervision, candour toward tribunals, and reasonable fees. Their application depends upon the circumstances of representation and the applicable professional rules. The underlying principle is that technological assistance does not relieve a lawyer of the obligations associated with professional practice.

Simshaw develops this position within a broader institutional analysis in *Legal Ethics of AI Snake Oil: Navigating the Hype, Harm, and Hope of Legal AI*, published in the *Michigan Law Review*.[21]

His study is a scholarly review of Arvind Narayanan and Sayash Kapoor's *AI Snake Oil*, rather than a new empirical benchmark of legal AI systems. This distinction matters when evaluating its contribution.

Simshaw's central concern extends beyond inaccurate AI outputs. He examines how institutional weaknesses, commercial incentives, technological enthusiasm, and professional expectations can encourage the adoption of systems whose actual benefits and limitations are insufficiently understood.

The problem, in this account, is not simply that lawyers may misunderstand a technology. It is that legal institutions may become receptive to exaggerated claims because they face existing pressures involving efficiency, cost, access to justice, and professional competition. This perspective introduces an important qualification to the conventional discussion of professional competence.

A lawyer's failure to verify an AI-generated legal authority may constitute a professional failure. However, the circumstances encouraging reliance upon unreliable systems may include organizational and commercial factors extending beyond the individual lawyer.

Simshaw also draws attention to lawyers' obligations when selecting and supervising third-party technology providers, particularly in connection with confidentiality and the use of external services.[22]

---

[20] American Bar Association. Standing Committee on Ethics and Professional Responsibility. Formal Opinion 512: Generative Artificial Intelligence Tools. – July 29, 2024. – URL: https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf (accessed: 11.10.2026).

[21] Simshaw D. Legal Ethics of AI Snake Oil: Navigating the Hype, Harm, and Hope of Legal AI // Michigan Law Review. – 2026. – Vol. 124, no. 8. – P. 1393–1423. – DOI: 10.36644/mlr124.8.legal. – URL: https://repository.law.umich.edu/mlr/vol124/iss8/14 (accessed: 11.10.2026).

[22] Simshaw, pp. 1420-1422.

His argument therefore connects professional responsibility with institutional due diligence. Nevertheless, the existence of an ethical duty to evaluate a vendor should not be confused with the existence of a corresponding civil claim against that vendor. Professional conduct rules principally regulate lawyers. Separate legal grounds are required to establish whether a manufacturer or service provider has breached an actionable duty toward an injured party. This distinction becomes decisive when the analysis moves from professional supervision to producer liability.

4.3. The Verification Burden and the Risk of Regulatory Asymmetry

Michael J. Bryant's *The AI Supervision Paradox: Double Standards, Algorithm Aversion, and the Unequal Burden of Legal Tech Regulation* introduces a substantial counterargument to increasingly demanding verification requirements.[23]

Bryant questions whether legal institutions are developing standards that treat AI-generated errors more severely than comparable errors arising through conventional human research or drafting.

The argument has two interconnected dimensions.

The first concerns consistency in professional regulation. If the relevant professional obligation is to exercise reasonable care in preparing accurate legal submissions, the mere involvement of AI should not automatically establish a more culpable form of misconduct.

The second concerns the economic distribution of compliance costs.

A large law firm may distribute research and verification tasks among associates, professional support staff, and specialized research teams. A sole practitioner may lack comparable resources. Verification requirements that are relatively inexpensive for one institution may substantially reduce the practical benefits of technological assistance for another.

Bryant argues that this asymmetry may reinforce existing inequalities within legal practice and restrict access to more affordable legal services.[24]

This position deserves serious consideration.

Professional regulation should not impose technologically specific burdens without adequate justification. Nor should regulatory design disregard differences in institutional capacity. At the same time, the existence of unequal compliance costs does not establish that verification requirements are substantively unnecessary. An important methodological difficulty concerns the comparison of human and machine error rates.

Legal research benchmarks may measure answer accuracy, completeness, authority selection, or task performance. Judicial sanctions concern different questions, including the conduct of a lawyer, the seriousness of an inaccurate representation, and the applicable

[23] Bryant M. J. The AI Supervision Paradox: Double Standards, Algorithm Aversion, and the Unequal Burden of Legal Tech Regulation. – 2026. – 51 p. – URL: https://ssrn.com/abstract=7337558 (accessed: 11.10.2026).

[24] Bryant, pp.28-32.

procedural obligations. A composite benchmark score cannot automatically be converted into the probability that a particular citation in a court filing is fabricated or materially misleading. Similarly, a statistical comparison between human and AI performance on selected research tasks does not independently establish that every category of legal error is equivalent in its causes or consequences.

Bryant recognizes important limitations in the available empirical record, including incomplete information about professional adoption, inconsistent survey methodologies, and the absence of certain demographic comparisons.[25] These limitations should inform the strength of the conclusions drawn from the evidence. The most persuasive element of Bryant's argument is therefore the demand for proportionate, evidence-based professional regulation rather than an assumption that all existing verification requirements are unjustified.

4.4. Certification, Safe Harbors, and the Meaning of Verification

Bryant proposes several reforms intended to reconcile technological efficiency with professional accountability.

These include a harmonized negligence standard, independent certification of sufficiently reliable legal AI systems, rebuttable safe harbors for practitioners using certified tools, and sanctions calibrated to professional conduct rather than technological origin.[26] The proposals raise an important institutional question: can independently demonstrated product reliability justify a reduction in the intensity of case-specific human verification?

In principle, certification may provide useful evidence concerning system quality. Independent testing can establish whether a defined product performs satisfactorily on specified tasks, within identifiable operating conditions and against stated evaluation criteria. Such evidence may properly inform procurement decisions, professional standards, insurance arrangements, and regulatory assessments.

However, certification and verification perform different functions.

Certification evaluates characteristics of a system or process. Verification of a legal submission concerns particular propositions, authorities, quotations, factual representations, and procedural assertions. A certified system may still produce an inaccurate statement in an individual case. Conversely, an uncertified system may produce a correct result that can be independently established through primary sources. The distinction becomes especially important where certification depends upon aggregate error rates. A product's performance within an acceptable statistical range does not necessarily establish that a specific legal authority exists, remains applicable, or supports the proposition for which it is cited.

The legal significance of certification must therefore be defined with care. A rebuttable safe harbor may be appropriate where it reflects demonstrated diligence and preserves meaningful case-specific obligations. It becomes more problematic if certification is treated

[25] Bryant, pp.8-10, 49.

[26] Bryant, pp.43-48.

as a substitute for examining material assertions in a filing. This does not require exhaustive duplication of every operation performed by an automated system. Rather, the required standard of verification should reflect the legal significance of the assertion, the risks associated with reliance, the reliability of the available sources, and the professional's ability to identify and correct errors.

The disagreement is thus not necessarily between efficiency and verification. It concerns the conditions under which technological reliability may legitimately inform, but not automatically replace, the exercise of professional judgment.

4.5. From Professional Misconduct to Independent Producer Liability

The question of manufacturer responsibility emerges more directly in Sohail's *AI Hallucinations and Lawyer Liability*[27] and Kildeev's *The Tool Has a Manufacturer: Producer Liability for Legal AI.*[28]

Both works challenge the proposition that professional responsibility exhausts the legal consequences of defective AI-assisted legal work. Their approaches nevertheless differ in emphasis and doctrinal development.

Sohail concentrates on incidents involving fabricated legal authorities and the regulatory responses directed toward lawyers who submitted them. Her argument is that professional sanctions address the conduct of individual users without necessarily addressing failures in the technological systems producing unreliable legal material.[29]

The study calls for shared accountability, stronger technological safeguards, and greater regulatory coordination. Its central concern is important: professional misconduct and technological failure may coexist within the same incident. However, the distinction between disciplinary responsibility and civil liability must be preserved.

A proceeding imposing professional sanctions upon an attorney does not necessarily determine whether a software provider could face an independent negligence, misrepresentation, contractual, or product-liability claim. The absence of manufacturer liability from the particular disciplinary proceeding therefore cannot, by itself, establish a general legal exemption for manufacturers. The proper inquiry requires identification of the potentially applicable cause of action and its constituent elements.

Kildeev's study addresses this question through a differentiated analysis of producer liability under the legal systems of the United States, Canada, and England.

It examines product classification, negligence, misrepresentation, warning duties, contractual limitations, and the relationship between professional reliance and retained technological control. A central proposition is that the lawyer's responsibility for verifying a

[27] Sohail S. AI Hallucinations and Lawyer Liability: A Critical Analysis of Professional Responsibility, Regulatory Frameworks, and the Accountability Gap. – 2026. – 11 p. – DOI: 10.2139/ssrn.7010178. – URL: https://ssrn.com/abstract=7010178 (accessed: 11.10.2026).

[28] Kildeev A. The Tool Has a Manufacturer: Producer Liability for Legal AI. – 2026. – 38 p. – DOI: 10.2139/ssrn.7388178. – URL: https://ssrn.com/abstract=7388178 (accessed: 11.10.2026).

[29] Sohail, pp.1-4.

legal submission does not necessarily exclude independent responsibility on the part of the entity that designed or marketed the system.

The framework treats liability as potentially cumulative rather than necessarily exclusive. This approach has a significant conceptual advantage. It does not require an artificial intelligence system to possess legal personality or independent legal agency. Instead, it directs attention toward the identifiable natural and juridical persons participating in the technological and professional chain. Nevertheless, the recognition of potentially cumulative responsibility does not itself determine the legal outcome. Different claims remain subject to different requirements concerning duty, defect, breach, causation, damage, contractual allocation of risk, and available defences.

A producer's superior technical knowledge may be relevant to establishing the scope of a duty. It does not automatically establish breach. Similarly, the foreseeability of professional reliance may support an argument concerning reasonable product design or risk disclosure. It does not independently prove that an actionable defect caused a particular loss.

The substantive contribution of this producer-liability approach is therefore best understood as a framework for identifying additional legally relevant actors and possible grounds of liability, rather than as proof that manufacturers must be liable whenever AI-assisted legal work fails.

4.6. Product Classification and the Limits of Existing Liability Regimes

The manufacturer's potential liability depends partly upon the legal classification of the technology and the nature of the alleged harm.

An AI system used in legal practice may combine software, continuously supplied services, access to proprietary databases, generated information, and functionality integrated into professional workflows.

Different elements of this arrangement may engage different areas of law.

Kildeev examines this difficulty through decisions concerning commercially distributed software, technical information products, and the provision of inaccurate information.[30]

The distinction is substantial. The characterization of software as a commercially supplied good does not necessarily establish that every software service qualifies as a product for purposes of tort liability. Likewise, an inaccurate statement generated by an AI system may support a misrepresentation analysis without necessarily establishing a defect actionable under strict product-liability doctrine.

Comparative legal analysis must therefore avoid treating contractual classification, product liability, and negligence as interchangeable mechanisms.

[30] Kildeev, pp.10-13, 19-22.

The European Union's revised Product Liability Directive, Directive (EU) 2024/2853, provides an important example of legislative adaptation.[31]

The Directive expressly brings software within its concept of products, including certain AI systems, irrespective of whether the relevant functionality is supplied through conventional distribution or cloud-based services.

However, the Directive distinguishes software from information as such and contains specific requirements concerning damage, defectiveness, causation, and the persons potentially liable.

Its temporal scope is also material. The revised regime applies to products placed on the market or put into service after 9 December 2026. As of the September 2026 monitoring period, that application date had not yet arrived.

Accordingly, the Directive demonstrates a legislative approach to future producer responsibility; it should not be presented as an already applicable liability regime governing all earlier AI-related losses. The relevance of this example extends beyond the European Union. It illustrates how legislators can adapt existing liability categories to technological developments without necessarily creating a new legal person or an entirely separate body of AI law. At the same time, the diversity of national legal systems prevents the unqualified transfer of one jurisdiction's product-liability rules into another. A manufacturer-liability framework must remain sensitive to the actual conditions imposed by the governing law.

4.7. Institutional Safeguards and the Distribution of Control

The reviewed scholarship also exposes a distinction between allocating responsibility after harm and designing institutions capable of reducing the probability of that harm.

Sohail emphasizes the need for technological and regulatory reforms extending beyond the punishment of individual lawyers.

Simshaw examines responsible vendor selection, professional supervision, and the institutional incentives underlying technology adoption.

Bryant proposes independent benchmarking and certification to improve professional confidence while reducing unnecessary compliance costs.

Kildeev argues for an external architecture separating AI-generated assertions from verified legal information, with source-level checking, attribution, qualified professional review, and reconstructible records.

These proposals operate at different levels.

Professional supervision addresses the conduct of the lawyer and the firm. Product evaluation concerns the demonstrated characteristics of the technology. Verification addresses the reliability of particular assertions. Auditability concerns the reconstruction and

[31] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on Liability for Defective Products and Repealing Council Directive 85/374/EEC // Official Journal of the European Union. – 2024. – 18 November. – URL: https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng (accessed: 11.10.2026).

examination of the process through which legally significant material was produced and adopted.

No single mechanism necessarily performs all of these functions.

For example, an audit record can demonstrate that a verification procedure was performed. It does not automatically establish that the substantive legal conclusion was correct. Likewise, independent benchmarking can provide evidence about system reliability under specified conditions without establishing the validity of every subsequent output.

Human review can preserve professional responsibility, but only where the reviewing professional has sufficient information, competence, authority, and opportunity to exercise meaningful judgment. A governance architecture that records a formal approval without establishing the substance of that review risks preserving the appearance of accountability without its actual function.

This difficulty is particularly relevant where technical systems are designed to generate confidence indicators, verification labels, or automatically produced compliance records. The existence of such records may improve accountability. Their evidentiary value nevertheless depends upon what they actually establish and whether their accuracy and integrity can be independently examined.

The common methodological lesson is that the mechanisms of oversight must be evaluated according to their demonstrated functions rather than their formal descriptions. The need for effective institutional safeguards does not determine in advance whether the appropriate solution should take the form of judicial rules, professional guidance, product regulation, contractual standards, or independent certification. That choice requires further analysis of institutional authority, implementation costs, effectiveness, and the distribution of legal responsibility.

4.8. Professional Responsibility Beyond Accuracy and Liability

A further contribution of Simshaw's work deserves recognition. Much of the contemporary debate concerning legal AI concentrates upon accuracy, confidentiality, efficiency, and liability for erroneous outputs. Simshaw extends professional-ethics analysis to the social, labor, and environmental consequences of AI deployment.[32]

His argument is that responsible technological adoption should not be evaluated solely through its immediate effect upon a lawyer's productivity or a client's financial interests. The broader institutional context also matters. This position should not be confused with a claim that every social or environmental cost automatically gives rise to a presently enforceable duty under existing professional-conduct rules.

Simshaw's analysis includes proposals for the future development of ethical standards and professional practices. Its importance lies in identifying dimensions of legal AI governance that are easily overlooked when analysis is restricted to courtroom hallucinations

[32] Simshaw, pp.1413-1423.

or conventional product liability. At the same time, proposed professional obligations must be distinguished from established legal requirements.

The strongest contribution of this broader ethical perspective is therefore its insistence that legal technology should be evaluated within the institutional and public-interest purposes of the profession.

4.9. Comparative Assessment

The selected studies reveal several competing priorities within the emerging governance of legal AI.

The protection of clients and courts requires competent professional conduct and reliable legal submissions. The development of accessible legal services requires regulatory arrangements that do not unnecessarily eliminate the efficiency gains of appropriate technological assistance. The prevention of systemic technological failures requires attention to the responsibilities of manufacturers, vendors, and institutional deployers.

Finally, the legitimacy of legal practice requires that accountability remain connected to identifiable persons and institutions capable of exercising meaningful control over the processes for which they are responsible.

These priorities are not necessarily incompatible. The more productive distinction is between different obligations arising at different points within the professional and technological process. A lawyer may be responsible for submitting an inaccurate legal representation. A firm may be responsible for inadequate supervision or procurement. A producer may incur independent obligations concerning the design, testing, representation, or operation of the system. Each proposition requires evaluation under the applicable legal framework.

The reviewed scholarship is strongest where it distinguishes these obligations and identifies the conditions under which they arise. It is less persuasive where professional sanctions are treated as proof that manufacturers enjoy universal immunity, where aggregate benchmark performance is treated as equivalent to case-specific legal verification, or where proposed technical safeguards are assumed to provide guarantees not established by independent testing. The central research question therefore concerns neither the exclusive liability of lawyers nor the automatic liability of manufacturers. It concerns the development of a legally coherent and operationally realistic framework in which duties correspond to the functions, knowledge, representations, and control exercised by identifiable actors.

The analysis also reveals that responsibility cannot be secured through legal allocation alone. Where automated systems increasingly influence professional legal work, the practical ability to inspect sources, challenge generated assertions, reconstruct relevant processes, and demonstrate meaningful verification becomes part of the institutional conditions of accountability. This conclusion leads to the next area of research: the relationship between verification, technical evidence, and the governance architectures proposed to control the risks of increasingly autonomous AI systems.

Chapter V. Verification, Evidence and Governance Architectures

5.1. Verification as a Legal and Technical Problem

The development of artificial intelligence governance has increasingly shifted attention from the reliability of individual model outputs toward the institutional and technical arrangements through which those outputs are produced, examined, authorized, and used.

This shift is particularly significant where automated systems generate legal information, interact with external services, execute delegated operations, or produce conclusions that may affect the rights and obligations of identifiable persons.

The selected scholarship reveals that verification is not a single operation. Its meaning depends upon what is being verified and the legal consequences attached to the result. An identity system may establish that a particular computational process issued a request. An authorization mechanism may establish that the process possessed the necessary technical permissions. An audit record may establish that an operation occurred under a particular configuration. A source examination may establish that a cited document supports a factual or legal proposition.

These findings are related, but none necessarily establishes the others. The distinction becomes especially important when the existence of a technically valid record is treated as evidence that an automated operation was properly authorized, substantively correct, or legally justified.

Farrukh Zaheer's *The Closed Intent Loop: Identity and Authorization in Autonomous Tool-Using AI Agents* examines this problem from the perspective of computer security and delegated authority.[33]

Jon Polenberg approaches a related difficulty through evidence law, particularly the recognition and evaluation of machine-generated conclusions.[34]

Kildeev considers the relationship between generated legal assertions, external verification, professional responsibility, and producer-controlled safeguards.[35]

Although the studies differ in disciplinary orientation, they expose a common analytical problem: the existence of a successful technical procedure does not necessarily establish that the procedure has satisfied the legal or epistemic requirement attributed to it. The implications extend beyond the reliability of language models. They concern the design of institutions capable of distinguishing the production of information from its validation and the execution of delegated functions from the legitimate exercise of authority.

[33] Zaheer F. The Closed Intent Loop: Identity and Authorization in Autonomous Tool-Using AI Agents. – 2026. – 42 p. – DOI: 10.2139/ssrn.7418158. – URL: https://ssrn.com/abstract=7418158 (accessed: 11.10.2026).

[34] Polenberg J., Cited.

[35] Kildeev, Cited.

5.2. Identity, Authorization, and the Closed Intent Loop

Zaheer's principal contribution is the distinction between establishing the identity of an AI agent and establishing that a particular action corresponds to the intention of the human principal who delegated authority to it.

Authentication answers whether a request originated from an identified agent or computational process. Authorization evaluates whether an operation falls within the permissions granted to that process.

Neither necessarily establishes whether the specific operation was wanted by the principal. The distinction is familiar in human-facing digital transactions. A user may authenticate successfully while a separate confirmation process is required before a consequential transaction is executed.

Zaheer argues that autonomous tool-using agents frequently lack an equivalent independent confirmation channel at the level of individual actions.[36]

An agent may possess valid credentials, operate within its assigned technical permissions, and generate records demonstrating compliance with established security rules. Yet the particular operation it performs may differ from the principal's actual objective.

The problem is not necessarily that authentication or authorization has failed. The controls may function exactly as designed while remaining unable to evaluate whether an otherwise permitted action reflects the principal's intended choice.

This constitutes the core of what Zaheer describes as the closed intent loop.

The concept offers an important correction to accounts of AI governance that equate successful authentication with legitimate delegation.

Nevertheless, the argument requires careful delimitation.

The absence of independent confirmation for every operation does not establish that all meaningful authorization is impossible. Human principals can establish objectives, define permissible operations, restrict resources, impose thresholds, and retain authority over specified categories of consequential decisions.

Zaheer expressly recognizes such mechanisms, while distinguishing them from independent confirmation of each individual action. The resulting problem is therefore one of authorization granularity and the institutional placement of control. The relevant question becomes how much operational discretion may be delegated within a bounded environment, which decisions require renewed approval, and what evidence can subsequently establish the scope of authority under which an operation occurred.

[36] Zaheer, pp. 1-4.

This distinction has direct legal significance. A technically valid credential is evidence of an authorization arrangement. It is not necessarily proof that every operation performed under that credential constituted an independently approved exercise of the principal's will.

At the same time, legal responsibility does not always depend upon proving specific subjective approval of every operational decision. The law may attribute consequences according to prior authorization, duties of supervision, control over the relevant activity, or applicable rules governing risk. Accordingly, Zaheer's framework should be understood as identifying limitations in technical evidence of authorization, rather than demonstrating a universal failure of legal attribution.

5.3. Compensating Controls and Their Evidentiary Limits

Zaheer proposes an eight-layer architecture intended to address risks associated with autonomous agents operating under organizational delegation.

The framework includes supply-chain provenance, agent identity, delegated authority, credential management, runtime authorization, decision auditability, lifecycle governance, and post-compromise forensic reconstruction.[37]

Its significance lies not simply in the number of layers but in their functional differentiation. Some controls restrict what an agent can do before an operation occurs. Others establish the technical origin of an action, preserve evidence, or support investigation after a failure.

These functions should not be treated as equivalent. A provenance mechanism may establish that a deployed artifact corresponds to a particular signed version. It does not necessarily establish that the artifact performs safely or produces correct conclusions.

An identity mechanism can distinguish one computational instance from another without proving that its actions are substantively appropriate. A runtime authorization mechanism may prevent operations outside predetermined limits while permitting an unwanted operation within those limits. An audit mechanism can establish what occurred without demonstrating that the recorded action was justified. The analysis is particularly valuable because Zaheer explicitly distinguishes preventive controls from mechanisms supporting attribution and reconstruction.

This distinction prevents a common conceptual error: assuming that the capacity to document a harmful operation necessarily entails the capacity to prevent it. The empirical component of the article is a documentary assessment of eighteen products, protocols, and governance frameworks against the author's specified coverage criteria.

Under the stated coding method, no individual solution in the sample enforces more than three of the eight layers. The assessment identifies limited documented coverage for certain

[37] Zaheer, pp. 17-24.

forms of runtime authorization and particularly weak coverage for post-compromise forensic capabilities.[38]

However, the evidentiary status of these findings must be preserved. The study examines documented capabilities rather than independently testing the products under operational or adversarial conditions. Its coding was conducted by a single researcher, and the proposed architecture has not been implemented or validated through deployment experiments. The author also reports sensitivity to the criteria used to classify coverage. A more permissive threshold changes the maximum number of layers attributed to an individual solution.

Accordingly, the findings cannot establish that existing commercial systems universally lack the relevant capabilities. They establish a narrower proposition: within the selected documentary sample and under the stated assessment criteria, the author identified significant gaps between the proposed architecture and the capabilities evidenced by the reviewed materials. This is a legitimate contribution to architectural research, provided it is not confused with demonstrated operational effectiveness. The distinction between a proposed control, a documented vendor capability, and an independently validated security property is fundamental to any credible evaluation of AI governance architectures.

5.4. Auditability, Independent Verification, and the Integrity of Evidence

A particularly important element of Zaheer's framework concerns the independence of the mechanisms that produce and preserve evidence.

The study examines situations in which an adversary compromises not merely the agent but components responsible for enforcing or recording security decisions. Under such conditions, a control may continue to operate and report successful compliance while processing information supplied or manipulated by the adversary.[39]

This produces a distinctive evidentiary difficulty. A record may be technically well formed, cryptographically valid, and consistent with the system's stated procedures without providing reliable evidence of an uncompromised authorization process. The existence of a log therefore cannot be equated with the integrity of the events it purports to document.

Zaheer's analysis distinguishes the preservation of an existing record against subsequent alteration from the more difficult problem of establishing that the record was trustworthy at the moment of creation. The distinction is especially important where authorization, monitoring, and logging share dependencies or administrative control. If the component responsible for certifying compliance is itself compromised, its certification may not provide independent evidence against that compromise.

This concern has a close conceptual relationship with verification problems identified in AI-assisted legal research. An automated system may generate a legal proposition and then

[38] Zaheer, pp. 22-25.

[39] Zaheer, pp. 14-15, 20-22.

produce an apparently confirming explanation. Another component may label the proposition as verified. A further system may record that the verification procedure was completed. Yet these operations do not necessarily establish that the relevant legal authority exists or supports the proposition. The decisive question is whether the purported verification introduces evidence capable of challenging the original assertion independently of the process that generated it. Independence should not be understood merely as the use of different software components or multiple language models.

It concerns the relevant sources of evidence, access permissions, administrative control, and the capacity to identify errors without relying exclusively upon the representations of the system being examined. This does not mean that every verification function must be performed manually. Automated controls can establish document identity, retrieve authoritative records, detect mismatched quotations, compare citations, and preserve relevant information. Their contribution is substantial where the underlying evidence and the validity of the checking procedure are independently examinable. The methodological distinction is between generating further assertions about correctness and establishing correctness through appropriately grounded evidence.

5.5. Machine-Generated Evidence and the Limits of Procedural Recognition

The legal consequences of this distinction become particularly apparent in Polenberg's examination of machine-generated evidence.

The problem arises when a technical output is introduced into legal proceedings not merely as an object requiring authentication but as a conclusion offered to establish a disputed proposition.

Polenberg examines the relationship between evidentiary reliability and the identification of a subject whose judgment the conclusion represents.[40] His analysis of the proposed Federal Rule of Evidence 707 is particularly relevant. The published proposal sought to address certain forms of machine-generated evidence offered without a human expert by requiring consideration of the reliability standards associated with expert testimony.

Following public consultation, the Advisory Committee on Evidence Rules withdrew the published proposal from immediate advancement and continued consideration of a revised approach. As of October 2026, Rule 707 had not entered into force.[41] The proposal illustrates a substantive regulatory difficulty. Traditional expert-evidence procedures are built around identifiable witnesses who may explain their methods, address objections, and be

[40] Polenberg J., pp.49-55

[41] Judicial Conference of the United States. Advisory Committee on Evidence Rules. Agenda Book for the October 15, 2026 Meeting. – 2026. – 221 p. – URL: https://www.uscourts.gov/sites/default/files/document/2026-10_evidence_rules_agenda_book_9-23_final.pdf (accessed: 11.10.2026). – Text : electronic.

challenged through adversarial procedures. Machine-generated determinations may perform similar informational functions without having the same relationship to a human expert.

Polenberg argues that procedural reliance upon technical reliability may obscure the distinct question of whose judgment is being accepted. His concern deserves attention, but the legal implications require further differentiation. An instrument can produce evidence without becoming a witness or a legal subject. Evidentiary systems have long recognized information derived from scientific and technical processes. The absence of a human declarant therefore does not necessarily defeat the admissibility or probative value of machine-generated information.

At the same time, authentication, reliability, admissibility, evidentiary weight, and the final judicial determination remain separate legal operations. Authentication concerns whether the item is what its proponent claims it to be. Reliability concerns whether the relevant methods and their application provide an adequate foundation for relying upon the result. Admissibility concerns whether the evidence satisfies the applicable procedural requirements.

Evidentiary weight concerns the significance that the decision-maker should attach to the admitted information. Judicial judgment concerns the authoritative resolution of the dispute. No single one of these operations necessarily establishes the others. The difficulty with machine-generated evidence is therefore not resolved merely by producing a technical reliability certificate. Nor is it resolved by declaring that all machine-generated conclusions are legally equivalent to human testimony.

A defensible approach must identify the precise evidentiary function of the output, the procedures through which its reliability may be tested, the persons responsible for presenting and evaluating it, and the available means of adversarial challenge. This approach preserves the value of technical evidence without requiring recognition of the producing system as an independent legal actor.

5.6. Source Verification in AI-Assisted Legal Research

The distinction between procedural verification and substantive accuracy is especially important in professional legal research.

An AI-assisted legal research system may perform several operations within one user interaction: locating potentially relevant materials, extracting passages, summarizing holdings, comparing authorities, and generating a proposed legal argument.

These operations differ in evidentiary significance. The retrieval of an existing judicial decision establishes that a document has been located. It does not necessarily establish that the document supports the proposition for which it is cited. A correct quotation may be taken from a decision that is procedurally distinguishable or no longer controlling. A valid citation may identify the appropriate document while omitting qualifications essential to the legal significance of the passage.

Likewise, a reliable summary of an earlier decision may become misleading when applied without considering subsequent legislation or changes in precedent. The problem is therefore not exhausted by detecting fabricated citations.

Kildeev's *The Tool Has a Manufacturer* develops a proposed architecture in which generated legal assertions are distinguished from externally verified information, with attention to source-level validation, qualified professional review, and reconstructible records.[42]

This approach shares with Zaheer's security framework an emphasis on separating generation from the controls intended to constrain or verify it.

Nevertheless, the two frameworks address different problems. Zaheer's controls primarily concern whether autonomous operations fall within delegated technical authority and whether their execution can be reconstructed. The legal verification architecture concerns whether generated propositions are supported by appropriate legal sources and whether a qualified professional has exercised the judgment required for their use.

The distinction is significant.

A system could prevent an AI agent from accessing unauthorized resources while still allowing it to generate an inaccurate legal argument. Conversely, a system might successfully verify the accuracy of every cited authority while failing to control an agent's unauthorized external operations. Both risks may arise in the same professional workflow, but they require different safeguards.

A comprehensive governance architecture should therefore distinguish between operational authorization and epistemic validation rather than assuming that one mechanism necessarily secures both.

The proposed separation between generated and verified material is conceptually useful. Its practical effectiveness, however, depends upon the accuracy of the source-checking mechanisms, the integrity of the records, and the substantive competence of the professional performing the final assessment. An architectural proposal cannot be treated as validated merely because its theoretical structure corresponds to a recognized legal duty.

5.7. Verification Without Creating a System of Generalized Surveillance

The demand for reconstructible records introduces another governance problem: the relationship between accountability and privacy.

Deborah Glosser's *The Chip of Theseus: Human-AI Coupling and the Erosion of Legal Individuation* offers a relevant counterpoint.

[42] Kildeev, pp.29-31.

Her analysis concerns continuous interaction between human users and AI systems, particularly circumstances in which technological assistance may influence decisions without creating an externally visible record of that influence. One possible regulatory response would be to require continuous logging and retention of relevant interactions. Such records could assist subsequent investigations by establishing whether an AI system was operating and what information was communicated.

Glosser nevertheless rejects comprehensive retrospective logging as her preferred approach, emphasizing the intrusion that persistent records of human conversations and cognitive assistance may create.[43] This concern has broader relevance to AI governance. The ability to reconstruct an automated process may require preserving prompts, retrieved documents, tool calls, model versions, and records of authorization. Yet indiscriminate retention can create additional risks involving confidentiality, third-party information, unauthorized disclosure, and the secondary use of sensitive records. The solution cannot simply be to maximize the volume of information retained.

Instead, the purpose of each record must be connected to a defined governance or evidentiary requirement.

An operational record necessary to establish the authorization of a consequential transaction may not require retention of the complete conversational history of every participant. Likewise, verification of a legal citation may require preserving the relevant source document and checking result without necessarily preserving unrelated confidential materials. The appropriate scope of documentation depends upon the purpose of the system, the rights affected, the applicable legal obligations, and the nature of the risk being controlled.

This introduces an important qualification to proposals for comprehensive AI auditability. More extensive recording does not automatically produce more reliable verification. Nor does the accumulation of information necessarily strengthen accountability where the resulting records are excessive, difficult to interpret, or incapable of establishing the substantive proposition under examination. Effective auditability requires the preservation of legally relevant evidence under conditions that protect its integrity while respecting the rights of persons whose information may be captured.

5.8. Independent Evaluation and the Problem of Recursive Verification

The selected research also raises a methodological question concerning the use of AI systems to evaluate other AI systems. Automated evaluation may be useful where the criteria are sufficiently defined and the evaluation procedure can be checked against external evidence. The difficulty arises where a generated conclusion is treated as independently verified because another probabilistic system has approved it.

This problem is particularly relevant to research methodologies employing language models to grade or compare the outputs of other models.

[43] Glosser, pp.17-18.

Aran, Perry, and Shelly's *Clinical Liability Cases Reveal a Coupling Between Legal Defensibility and Medical Procedure Escalation in Large Language Models* provides a concrete example of such methodology.[44]

The authors construct a benchmark using medical malpractice decisions and employ three language models to evaluate simulated clinical recommendations against criteria derived from judicial findings. The study incorporates documentary grounding and quality-control procedures, but the use of multiple model-based evaluators remains an important methodological consideration. Agreement among evaluating models may increase consistency within the evaluation procedure. It does not, without additional evidence, establish that the resulting assessment corresponds to an independently correct clinical or legal judgment. The relevant question is whether the evaluation criteria are sufficiently grounded, whether the scoring procedure can be reproduced, and whether the results have been adequately checked against evidence independent of the systems being evaluated.

This distinction should not be interpreted as a rejection of automated evaluation. Model-assisted extraction, classification, and consistency checking may substantially improve research efficiency. Their evidentiary contribution depends upon the reliability and independence of the reference materials and validation procedures.

The strength of the Aran, Perry, and Shelly study must therefore be assessed through its actual methodology and demonstrated results rather than through either automatic confidence in multi-model agreement or categorical skepticism toward computational evaluation.

Its empirical implications will be considered further in the subsequent comparative analysis. The broader lesson is that the multiplication of verification procedures does not necessarily produce independent verification. What matters is whether the procedures introduce evidence capable of detecting errors that the original generating process would otherwise reproduce.

5.9. Comparative Assessment: From Documented Control to Demonstrable Accountability

The reviewed publications identify several distinct problems that are frequently grouped under the general language of AI verification.

Zaheer examines the relationship between authenticated identity, delegated authority, runtime authorization, and the reconstruction of autonomous operations.

---

[44] Aran D., Perry R., Shelly S. Clinical Liability Cases Reveal a Coupling Between Legal Defensibility and Medical Procedure Escalation in Large Language Models // Communications Medicine. – 2026. – Article in press. – DOI: 10.1038/s43856-026-01839-5. – URL: https://doi.org/10.1038/s43856-026-01839-5 (accessed: 11.10.2026).

Polenberg examines whether procedures for recognizing machine-generated evidence adequately address the legal significance attributed to automated conclusions.

Kildeev develops a framework for separating generated legal assertions from source-verified propositions and for connecting professional responsibility with producer-controlled safeguards.

Glosser emphasizes the privacy and institutional costs of comprehensive records intended to establish technological influence.

Aran, Perry, and Shelly demonstrate how model-based evaluation can be incorporated into an empirical research design, while also raising questions concerning the independence of the resulting validation.

These contributions should not be reduced to a single universal architecture.

They concern different functions and operate under different legal, technical, and evidentiary conditions.

Nevertheless, their comparison supports several general conclusions.

First, identity is not equivalent to authorization. Establishing which system performed an operation does not necessarily establish that the operation corresponded to an authorized objective.

Second, authorization is not equivalent to substantive correctness. A permitted operation may still produce an inaccurate or harmful result.

Third, auditability is not equivalent to verification. A record of the procedure performed does not necessarily establish the reliability of the underlying conclusion.

Fourth, procedural verification is not equivalent to legal judgment. The technical validation of information does not independently confer authority to determine its legal significance.

Finally, accountability cannot be reduced to the production of a compliance record. The record must be capable of supporting meaningful examination by a person or institution possessing the relevant legal authority and substantive competence.

The distinction between technical implementation and demonstrated performance is equally important. Architectural diagrams, product documentation, benchmark scores, and formal compliance procedures provide different kinds of evidence. Their probative value depends upon the claims they are invoked to support.

A documented capability is not necessarily an independently tested capability. An independently tested capability is not necessarily effective under every foreseeable condition.

And the effectiveness of a technical safeguard does not, by itself, establish the legal adequacy of the institution employing it.

The most productive direction for further research is therefore neither the categorical rejection of AI-based governance controls nor the assumption that increasingly complex systems will automatically solve the problems they are designed to address. It is the development and testing of architectures in which operational permissions, evidentiary integrity, substantive verification, and legally attributable responsibility remain distinguishable, independently examinable, and appropriately connected.

The resulting governance problem is institutional as much as technological. Its solution requires not merely instruments capable of generating records and classifications, but procedures through which those records and classifications may be challenged, their limitations established, and their legal consequences attributed to identifiable actors.

This provides the foundation for the final stage of the comparative inquiry: assessing the empirical evidence, methodological limitations, and institutional consequences of the proposed approaches to AI governance.

Chapter VI. Empirical Findings and Institutional Consequences

6.1. The Evidentiary Foundations of AI Governance Research

The preceding sections have examined competing approaches to legal attribution, professional responsibility, human judgment, and the technical architecture of AI governance. Their practical significance ultimately depends upon a further question: what empirical evidence supports the proposed legal and institutional conclusions?

This question is particularly important because contemporary AI governance scholarship draws upon substantially different categories of evidence. Controlled experiments, comparative performance benchmarks, professional surveys, documentary analyses, judicial decisions, industry reports, and normative policy proposals may all contribute to legal research. However, they do not establish the same kinds of propositions.

An experimental benchmark may demonstrate that particular systems behave differently under specified testing conditions. It does not necessarily establish that the observed differences will persist in professional practice. A survey may establish the reported attitudes or practices of its respondents. It does not necessarily establish the behavior of an entire profession. A documentary assessment may identify whether particular technical capabilities are publicly described. It does not independently demonstrate that the corresponding systems perform those functions effectively in deployment. Likewise, a doctrinal proposal may offer a coherent framework for future legal regulation without providing evidence that the proposed framework has already achieved its intended consequences.

These distinctions are central to assessing the selected scholarship.

Aran, Perry, and Shelly provide an empirical comparison of language-model performance in simulated clinical consultations based upon medical malpractice decisions.

Bryant synthesizes comparative research and professional data to evaluate the consequences of AI-specific verification requirements.

Wayne Unger reports preliminary findings from a survey concerning academic attitudes toward AI-assisted legal scholarship.[45]

Zaheer examines documented technical capabilities and identifies limitations requiring experimental validation.

Other contributions, including Bernstein's public-health framework and the proposed liability standards discussed earlier, primarily seek to derive regulatory recommendations from legal, institutional, and policy analysis.[46]

The methodological challenge is not to establish a rigid hierarchy in which experimental studies automatically displace legal scholarship. The relevant question is whether each author's conclusions remain within the evidentiary limits of the method employed. A credible critical review must therefore distinguish demonstrated findings, interpretations of those findings, and proposals for institutional reform.

6.2. Measured Performance and the Meaning of Legal Defensibility

One of the most substantial empirical contributions within the selected corpus is the study by Aran, Perry, and Shelly, published in *Communications Medicine* in August 2026.

The authors introduce the Clinical Legal Accountability Dataset (CLAD), constructed from judicial decisions concerning medical malpractice in the United States.

The study examines fifteen language models and an additional prompting variant through simulated clinical consultations. Its central objective is to compare the extent to which model-generated recommendations correspond to actions identified as necessary in the underlying judicial decisions.

The original dataset contained 268 selected cases. Seventy cases in which none of the evaluated models achieved a positive defensibility score were excluded from the principal

[45] Unger W. AI Use in Legal Scholarship [preliminary abstract, survey instrument and appendices]. – 2026. – 15 p. – DOI: 10.2139/ssrn.7290238. – URL: https://ssrn.com/abstract=7290238 (accessed: 11.10.2026).

[46] Bernstein G. From Bans to Recalls: A Public Health Framework for AI Companion Bots. – Washington, DC : Brookings Institution, 2026. – 16 p. – DOI: 10.2139/ssrn.6922118. – URL: https://www.brookings.edu/articles/from-bans-to-recalls-a-public-health-framework-for-ai-companion-bots/ (accessed: 11.10.2026).

comparative analysis. The remaining 198 cases generated 3,072 valid simulated consultations across sixteen model configurations.[47]

This distinction between the initial dataset and the final analytical sample is methodologically significant.

The reported comparative results concern the retained cases, not the entire original set of 268. Excluding cases according to the performance of the evaluated models changes the population to which the principal findings directly apply.

The procedure may be justified for comparing systems on cases that distinguish their performance. Nevertheless, it limits the interpretation of the resulting scores as measures of overall clinical competence.

Within the retained sample, the authors report substantial differences between models.

GPT-5.2 achieved a mean defensibility score of 0.71, compared with 0.34 for GPT-4o. The assessment concerns conformity with court-derived criteria, principally whether the model identified an action regarded as necessary under the relevant standard of care.[48]

These findings support a specific conclusion: under the study's defined evaluation procedure, the tested models differed substantially in their capacity to identify relevant actions. They do not establish that the highest-scoring model would necessarily produce the best clinical outcomes in real practice. The authors expressly acknowledge that their study does not establish a baseline based upon human physicians performing the same consultations. Moreover, the underlying cases represent a selected category of medical disputes rather than a representative sample of ordinary clinical decisions. The distinction between benchmark performance and practical competence is therefore essential. A model's ability to identify an action retrospectively recognized by a court does not, by itself, establish that the entire proposed course of conduct is medically optimal, economically justified, or legally sufficient in every comparable situation. The study is most persuasive when interpreted as a controlled comparison of model behavior under specified criteria rather than as a comprehensive measurement of clinical or juridical competence.

6.3. The Relationship Between Defensibility and Institutional Costs

The most consequential finding of the CLAD study concerns the relationship between higher defensibility scores and the number and estimated cost of recommended medical procedures.

Across the fifteen base models, the authors report a strong positive correlation between mean defensibility and estimated procedural costs: Spearman's $\rho = 0.95$ [15, pp. 8–11].

[47] Aran, Perry, and Shelly, pp.4-7.

[48] Aran, Perry, and Shelly, pp.7-8

GPT-5.2 recommended an average of 9.3 procedures per simulated consultation, with estimated Medicare costs of $1,118. GPT-4o recommended an average of 1.3 procedures at an estimated cost of $221.

The difference in estimated expenditure was approximately fivefold.

These results raise a broader institutional question: can improving automated performance against a particular legal or professional benchmark simultaneously increase the economic costs of the resulting decisions?

The answer suggested by the experiment is affirmative under the conditions tested. However, the interpretation of the observed relationship requires particular caution. The study does not demonstrate that all additional procedures recommended by higher-scoring models were medically unnecessary.

Indeed, the authors acknowledge that their dataset disproportionately contains situations in which the alleged professional failure involved insufficient diagnosis, treatment, monitoring, or intervention. Consequently, a recommendation containing more procedures may be more likely to identify the action recognized by the court as necessary.

The benchmark therefore incorporates an asymmetry that may reward comprehensiveness. At the same time, the authors report that the relationship between procedure count and defensibility within individual models is much weaker than the corresponding relationship between models.

This finding complicates the interpretation of higher expenditure as an unavoidable condition of better performance. Different systems may achieve comparable levels of defensibility while recommending different numbers of procedures. The authors also identify differences between model families, suggesting that the relationship between increased capability and expenditure is not uniform across development trajectories.

These observations support a more limited conclusion than the claim that increasingly capable AI necessarily increases healthcare costs. They demonstrate that the measured relationship depends upon the systems tested, their operating conditions, and the structure of the evaluation procedure. The authors consider several possible explanations, including improved identification of necessary procedures and training-related preferences for more extensive intervention. Their data do not conclusively distinguish between those mechanisms.

The institutional implication nevertheless remains substantial.

Where AI systems are evaluated primarily according to their ability to avoid certain categories of error, the resulting optimization may affect other objectives, including cost, proportionality, and resource allocation. A governance framework that measures only one dimension of performance may therefore overlook material consequences in other dimensions.

This conclusion extends beyond healthcare.

Legal research, administrative decision-making, regulatory compliance, and judicial support systems may also optimize performance against criteria that do not fully represent the institutional objectives they are intended to serve.

High benchmark performance should consequently be examined alongside the consequences of achieving it.

6.4. Independent Validation and the Limits of Automated Evaluation

The CLAD methodology also provides an opportunity to examine the evidentiary significance of model-assisted evaluation. The authors employed three language models to assess consultation outputs against criteria derived from judicial decisions. Final scores were determined through an evaluation procedure incorporating majority agreement.[49]

The study did not rely exclusively upon agreement among models.

A blinded medical expert separately reviewed 121 sampled consultations. For the 106 consultations included in the binary agreement comparison, the expert's assessments agreed with the model-evaluator majority in 89 cases, or approximately 84%.[50]

This additional validation is methodologically important. It provides a comparison against a qualified human evaluator rather than treating agreement among automated systems as sufficient evidence of correctness.

The authors also make the underlying dataset and code available through a public repository and an archived research record, permitting further examination and potential replication. These measures distinguish the study from purely conceptual or undocumented assertions about model performance.

Nevertheless, limitations remain.

The independent clinical validation involved a single expert rather than a panel capable of measuring differences among professional judgments. Agreement with one expert does not necessarily establish agreement with the broader medical profession. The model-based evaluation also remains dependent upon the quality of the extracted judicial criteria and the correspondence between those criteria and the simulated consultations.

Judicial decisions provide legally relevant records, but they do not necessarily supply complete or universally applicable clinical standards. Moreover, the interpretation of model performance remains conditional upon the chosen scoring procedure. These limitations do

[49] Aran, Perry, and Shelly, pp.3-6.

[50] Aran, Perry, and Shelly, pp.7-8.

not invalidate the experiment. They identify the boundaries within which its conclusions are supported.

An additional methodological distinction concerns reproducibility. The release of underlying data and code improves the capacity for independent evaluation. It does not itself establish that an independent replication has already confirmed the published results. Availability of evidence and independent confirmation of evidence are separate achievements.

The broader implication is that the credibility of AI-assisted empirical research should be evaluated through the transparency of its methods, the relevance of its validation procedures, the accessibility of underlying evidence, and the correspondence between its findings and conclusions.

Neither the number of automated evaluators nor the sophistication of the models involved can substitute for those requirements.

6.5. Comparative Legal Performance and the Problem of Unequal Benchmarks

The evidentiary difficulties identified in clinical evaluation also arise in comparative studies of legal AI performance.

Bryant examines research comparing human lawyers and AI systems across several professional tasks, including document review, contract analysis, and legal research. His argument challenges regulatory approaches that appear to treat AI-generated errors as categorically more serious than errors produced through conventional human work.

The empirical materials he discusses provide legitimate reasons to question the assumption that human legal research is consistently more accurate than automated assistance. However, the studies employ different tasks, evaluation criteria, and definitions of successful performance. For example, a legal research benchmark combining accuracy, authoritativeness, and appropriateness into a weighted score does not measure precisely the same property as a study examining whether particular legal citations are fabricated. A system achieving a higher aggregate research score than a human comparison group cannot automatically be regarded as less likely to produce every legally significant category of error.

Similarly, a lower human benchmark score does not necessarily establish a corresponding probability of professional misconduct. This distinction becomes particularly important when benchmark results are used to justify legal standards concerning professional negligence, sanctions, or regulatory safe harbors. The empirical question concerns the accuracy or reliability of particular work products under specified conditions. The legal question concerns whether a professional exercised the care required in the circumstances. These questions are connected, but they are not identical. A comparative performance study may inform the expected reliability of an instrument. It does not independently establish whether reliance upon that instrument was reasonable in a particular legal proceeding.

Bryant's proposal for certification-based safe harbors therefore requires more than evidence that certain systems outperform human participants on selected benchmarks. It requires an adequate connection between the certified performance characteristics and the professional obligations affected by the proposed presumption of competence. The relevance of the research is substantial, but the institutional consequences cannot be inferred directly from aggregate performance rankings. The comparison illustrates a recurring problem in AI governance scholarship: quantitative results are sometimes asked to resolve legal questions for which they provide only partial evidence.

6.6. The Institutional Distribution of Verification Costs

A separate empirical question concerns the economic consequences of legal AI regulation.

Bryant argues that extensive verification requirements may impose relatively greater burdens upon sole practitioners and small legal firms than upon larger organizations capable of distributing verification tasks among professional personnel.[51]

The institutional mechanism is plausible. A requirement involving several hours of additional research may be absorbed differently by a firm with specialized resources and by an individual practitioner operating under strict time and financial constraints. The resulting disparity may influence technological adoption, professional competition, and the affordability of legal services. Nevertheless, the existence and magnitude of a particular distributional effect require evidence appropriate to the claim. Bryant expressly identifies limitations in the available professional data, including inconsistencies between surveys, self-selection within industry research, and the absence of certain demographic comparisons.[52]

These limitations matter because differences in AI adoption or verification capacity may result from several interacting variables, including firm size, professional specialization, organizational policies, technology costs, and access to training.

The available evidence does not establish every proposed causal relationship among these factors. Nor does a potentially disproportionate economic burden automatically establish unlawful discrimination under a particular constitutional or statutory framework.

Such a conclusion requires the additional elements demanded by the applicable legal doctrine. The strongest institutional implication of Bryant's analysis is therefore the need to evaluate the proportionality of verification requirements and the economic consequences of their implementation. A professionally justified safeguard should not be treated as immune from cost-benefit examination merely because its objective is legitimate.

Conversely, a safeguard should not be considered unnecessary merely because it imposes unequal compliance costs. The research question concerns how to achieve reliable

[51] Bryant, pp.28-32.

[52] Bryant, pp.49-50.

professional verification through measures proportionate to the risks and resources involved. This requires evidence of both the benefits of verification and the costs associated with different methods of performing it. The available literature identifies the problem more clearly than it establishes a definitive regulatory solution.

6.7. Empirical Attitudes Toward AI-Assisted Legal Scholarship

The institutional consequences of AI adoption extend beyond legal practice to the production and evaluation of academic research.

Wayne Unger's *AI Use in Legal Scholarship* provides preliminary survey evidence concerning attitudes toward generative AI within the legal academy.

The material available through SSRN consists of an abstract, survey materials, appendices, and a memorandum explaining that the complete research article remains under development.

This publication status is methodologically significant.

The available document reports findings from 268 survey responses, but it does not provide the full analytical presentation normally required for comprehensive independent evaluation of the study.

According to the published abstract, 78% of respondents regarded AI-assisted grammar correction as acceptable, while 76% accepted assistance with terminology consistency.

By contrast, only 11% considered it acceptable for AI to draft analytical sections presenting novel arguments.

The reported difference suggests that the respondents distinguish between technical or editorial assistance and activities associated with substantive intellectual authorship.

The survey also reports that 83% considered the establishment of institutional AI-use norms important, whereas 7.5% indicated that their institutions had explicit policies governing AI use in scholarship.[53] These figures identify a potentially significant gap between respondents' expectations concerning academic governance and their reports of institutional policy development.

However, the evidentiary limitations must be recognized.

Without the complete research article and accompanying methodological analysis, the available material does not permit a comprehensive assessment of the sampling process, response rate, representativeness, or statistical treatment of the results. The reported

[53] Unger, pp.1-2.

percentages therefore describe the responses summarized by the author. They should not automatically be generalized to the entire population of legal academics.

Moreover, the survey measures reported acceptability and institutional perceptions. It does not establish how frequently particular academic practices actually occur, whether respondents independently verify AI-generated materials, or whether the use of AI affects the quality of their published research.

These are distinct empirical questions requiring different methods of investigation.

Nevertheless, Unger's preliminary results contribute a relevant institutional perspective. They suggest that at least within the surveyed group, distinctions between editorial assistance and substantive intellectual production remain important. They also raise questions concerning the relationship between academic expectations, institutional policy, and the practical conditions under which authors use generative systems.

The value of the study lies in identifying these questions and providing preliminary data for further research. Its incomplete publication status prevents treating the reported findings as a comprehensive empirical account of AI-assisted legal scholarship.

6.8. From Evidence of Risk to Regulatory Intervention

Empirical findings concerning AI systems frequently become the foundation for proposals involving regulation, product safety, professional duties, or institutional redesign. The transition from observed risk to proposed intervention requires particular analytical care.

Gaia Bernstein's *From Bans to Recalls: A Public Health Framework for AI Companion Bots*, originally published as a Brookings policy brief in May 2026, provides an example.[54]

Bernstein argues for treating risks associated with AI companion systems through a public-health framework emphasizing preventive oversight, product safety, and the potential removal of unsafe systems from the market. The proposal draws upon regulatory mechanisms associated with products whose use may affect health and well-being.

Its principal contribution is institutional rather than experimental. The brief synthesizes existing research, documented concerns, and legislative approaches to advance a regulatory framework. It does not establish through an original controlled experiment that a particular recall regime would reduce the specified harms by a measurable amount.

This distinction is important because a policy proposal may be legally and institutionally persuasive without possessing the same evidentiary status as an intervention that has been implemented and evaluated. The analogy between AI products and traditionally regulated

[54] Bernstein G. From Bans to Recalls: A Public Health Framework for AI Companion Bots. – Washington, DC : Brookings Institution, 2026. – 16 p. – DOI: 10.2139/ssrn.6922118. – URL: https://www.brookings.edu/articles/from-bans-to-recalls-a-public-health-framework-for-ai-companion-bots/ (accessed: 11.10.2026).

health-related products also requires attention to differences in intended use, regulatory jurisdiction, evidentiary standards, and the practical feasibility of enforcement.

The relevance of Bernstein's study lies in its examination of regulatory incentives and the possibility that product-level obligations may address risks inadequately controlled through individual user responsibility. However, the effectiveness and proportionality of the proposed regime remain questions for subsequent legislative design and empirical evaluation.

A similar distinction arises in proposals for specialized AI liability standards.

Oliveira's Clinical Algorithmic Liability Attribution Standard (CALAS), for example, addresses the allocation of responsibility and evidentiary burdens where automated systems contribute to medical decisions.[55]

Its contribution concerns the proposed organization of legal responsibility among physicians, institutions, developers, and producers. The framework should be evaluated as a normative and doctrinal proposal, not as an empirically validated compensation mechanism.

Likewise, proposals for technical verification architectures must distinguish the existence of a conceptual design from evidence that the design has successfully reduced errors, prevented unauthorized operations, or improved legal outcomes in deployment. The general methodological requirement is to preserve the distinction between a demonstrated problem, a plausible causal explanation, and a proposed institutional remedy. The existence of the first does not automatically establish the correctness of the second or the effectiveness of the third.

6.9. Implementation Evidence and the Governance Gap

The relationship between proposed safeguards and demonstrated institutional performance becomes particularly important in the technical literature.

Zaheer's documentary assessment of eighteen solutions identifies limitations in the capabilities evidenced by reviewed product documentation and technical materials. The proposed architecture organizes relevant controls into layers concerning identity, delegation, authorization, auditability, and forensic reconstruction.

Its practical significance depends upon the capacity of these controls to operate effectively when integrated into real systems. However, Zaheer expressly acknowledges that the architecture had not been implemented, deployed, or tested under adversarial conditions at the time of writing.

[55] Oliveira N. S. de. Medical Malpractice, Artificial Intelligence and Physician Liability: Who Is Responsible for AI-Assisted Medical Errors? – 2026. – 16 p. – DOI: 10.2139/ssrn.7262340. – URL: https://ssrn.com/abstract=7262340 (accessed: 11.10.2026).

The study therefore provides evidence concerning the documented state of relevant technologies and the logical organization of a proposed security framework. It does not demonstrate that the complete architecture will achieve its intended operational outcomes. This distinction applies equally to legal and institutional governance proposals. A professional verification requirement may be coherent in principle while imposing unforeseen costs or producing formalistic compliance.

A technical audit mechanism may preserve reliable records without preventing the harmful action being recorded. A certification regime may improve product selection while failing to detect particular errors. A producer-liability rule may establish incentives for better safeguards without independently demonstrating that those incentives produce the intended behavioral changes. Consequently, the evaluation of governance proposals requires attention to implementation evidence.

Relevant questions include whether a proposed mechanism has been deployed, whether its performance has been independently measured, which failure conditions were tested, what costs it imposes, and whether the resulting effects correspond to its stated objectives.

The literature reviewed here offers several useful conceptual and institutional proposals. Their empirical maturity, however, varies considerably. Recognizing this variation is not a criticism of theoretical research. It is a necessary condition for assigning each contribution its appropriate evidentiary significance.

6.10. Comparative Assessment: Evidence, Generalization, and Institutional Consequences

The selected scholarship demonstrates that the empirical foundations of AI governance are developing unevenly across legal and technological domains. Controlled experiments can identify differences in system performance, but their results remain dependent upon evaluation criteria, sample selection, and the conditions of testing. Professional surveys can reveal attitudes and reported practices, but they require adequate information concerning sampling and representativeness before broader conclusions can be drawn. Documentary assessments can identify documented capabilities and institutional gaps, but they do not necessarily establish operational effectiveness. Doctrinal and policy research can develop coherent regulatory proposals without having already demonstrated their practical consequences.

These forms of research are complementary rather than interchangeable.

The central methodological problem arises when conclusions appropriate to one form of evidence are transferred into another domain without establishing the necessary connection. A measured performance advantage does not automatically establish professional competence. An aggregate benchmark score does not establish the accuracy of a particular legal citation. A correlation between model performance and resource utilization does not establish that higher expenditure is necessarily wasteful.

Reported academic attitudes do not prove actual scholarly practices. A documented technical capability does not establish successful implementation. And a persuasive regulatory proposal does not establish that the proposed regulation will achieve its intended effects. At the same time, the reviewed research provides a useful foundation for identifying future priorities.

AI systems should be evaluated against the substantive purposes of the institutions employing them, rather than exclusively against simplified technical benchmarks. Verification requirements should be examined for both their effectiveness and their practical costs. Claims concerning professional competence and product reliability should be tied to reproducible, task-specific evidence. Regulatory proposals should distinguish their normative justification from empirical claims concerning their expected outcomes.

Finally, governance architectures should be examined not only for their formal completeness but also for their demonstrated capacity to preserve meaningful control, reliable evidence, and legally attributable responsibility. The comparison also reveals the limits of any attempt to reduce AI governance to a single technical or doctrinal principle. Different institutional environments present different combinations of legal authority, evidentiary requirements, economic incentives, privacy interests, and professional obligations.

A framework suitable for autonomous software operations may not be sufficient for evaluating clinical recommendations. A professional rule governing legal research may not resolve the allocation of manufacturer liability. A public-health approach to consumer technology may require regulatory powers and evidence different from those relevant to legal practice.

The resulting challenge is not simply to develop more comprehensive AI governance models. It is to ensure that the methods used to justify those models can support the claims being made about their necessity, proportionality, and expected effectiveness. The selected literature is most valuable where it maintains that distinction and most vulnerable where institutional conclusions exceed the evidence from which they are derived.

The final comparative assessment can therefore proceed from a common methodological foundation: the evaluation of AI governance scholarship must consider doctrinal coherence, evidentiary reliability, legal applicability, and implementation feasibility as related but independently examinable dimensions of scholarly contribution.

Chapter VII. Comparative Assessment and Conclusions

7.1. The Emerging Structure of AI Governance Scholarship

The comparative examination of selected scholarship circulated through the SSRN Legal Scholarship Network in September 2026 reveals a field characterized by considerable intellectual activity, increasingly sophisticated legal arguments, and persistent disagreements concerning the foundations of artificial intelligence governance.

The reviewed publications do not constitute a unified school of thought. They approach AI governance through different legal traditions, philosophical assumptions, technical architectures, and institutional objectives. Their disagreements concern not merely the selection of regulatory instruments but also the characterization of the legal problems those instruments are intended to resolve.

Some authors emphasize the difficulty of attributing autonomous conduct to identifiable legal actors. Others question whether machine-generated determinations can acquire the status of legally authoritative judgment. A further group concentrates upon professional responsibility, producer liability, technological verification, institutional oversight, or the distributive consequences of regulatory intervention.

These approaches illuminate different dimensions of a common problem: the increasing capacity of computational systems to perform functions previously associated with human professional activity does not automatically determine how legal authority, responsibility, and institutional control should be organized.

The principal finding of this review is that the most productive developments in AI governance scholarship arise where authors distinguish technological capability from legal attribution and recognize that accountability cannot be established through functional performance alone.

Nevertheless, the significance of this proposition should not obscure the contributions of alternative approaches. Greater computational accuracy, improved consistency, and the automation of routine tasks may provide substantial institutional benefits. The relevant legal question is not whether such benefits exist, but under what conditions they justify changes in the distribution of decision-making authority and responsibility.

The selected literature is therefore most valuable when read as a continuing debate over institutional design rather than as a competition to establish a single comprehensive theory of artificial intelligence.

7.2. Competing Doctrines and Their Common Legal Problems

Several substantive areas of convergence emerge from the comparative analysis.

First, the reviewed scholarship increasingly recognizes the limitations of treating artificial intelligence as an ordinary extension of traditional software without examining the specific functions performed by the system. Autonomous execution, probabilistic generation, continuous modification, and integration into professional workflows may create difficulties that conventional contractual and institutional arrangements do not always address explicitly. However, these difficulties do not necessarily establish the inadequacy of existing legal categories. The distinction between a limitation in the application of a doctrine and a fundamental defect in its conceptual foundations remains essential.

Second, the scholarship reveals continuing disagreement over the relationship between functional autonomy and legal agency. The capacity of an automated system to select operations, generate conclusions, or respond to changing circumstances does not, by itself, establish that the system possesses legal personality or independent juridical authority. At the same time, the absence of such personality does not eliminate the possibility that automated systems may complicate conventional rules concerning authorization, causation, evidence, and professional conduct. The relevant inquiry must therefore identify the particular legal function affected rather than infer a general crisis of responsibility from technological autonomy.

Third, the comparative analysis demonstrates that assigning legal consequences and attributing judgment constitute distinct operations. A manufacturer may bear responsibility for a defective system without becoming the intellectual author of each output. A professional may assume responsibility for an AI-assisted document without necessarily controlling every aspect of the underlying technology. A court may rely upon machine-generated evidence without recognizing the machine as an independent judicial actor. These distinctions are essential to maintaining doctrinal coherence.

Fourth, the selected studies demonstrate that institutional accountability cannot be reduced to the identification of a single responsible individual. AI systems frequently operate through technological and organizational arrangements involving developers, suppliers, deployers, professional users, and institutions exercising different forms of control. Responsibility may therefore arise at several points within the same process, subject to the requirements of the applicable legal regime.

The recognition of potentially distributed responsibility does not imply that every participant must be liable for every resulting harm. It requires the identification of the relevant duty, the conduct or defect alleged, the applicable causal relationship, and the legal grounds upon which responsibility may be imposed.

The emerging doctrinal challenge is consequently one of differentiation rather than indiscriminate expansion.

7.3. Human Judgment, Technical Verification, and Institutional Authority

A particularly significant contribution of the reviewed scholarship concerns the relationship between human judgment and automated decision-making. The studies examined in this review provide substantial reasons to distinguish the production of a technically competent determination from the legally authorized exercise of judgment.

This distinction should not be interpreted as an assertion that human decision-makers are invariably more accurate, consistent, or reliable than computational systems. Empirical research indicates that automated systems may outperform human participants on particular tasks under specified testing conditions. Such findings warrant serious consideration, particularly where technological assistance may improve access to services, reduce certain errors, or support more consistent institutional practice. However, comparative performance

does not independently settle questions concerning procedural legitimacy, professional responsibility, or the authorization of binding decisions.

The institutional significance of an automated determination depends not only upon its substantive accuracy but also upon the legal framework governing its production, examination, adoption, and challenge.

The same principle applies to verification. The selected technical and legal studies demonstrate that authentication, authorization, provenance, auditability, and substantive verification perform different functions. A valid identity credential does not establish that every action of an automated agent corresponds to the principal's intentions. A technically complete audit record does not necessarily establish the substantive correctness of the operation recorded. A retrieved legal source does not necessarily support the proposition for which it is cited. A model-generated explanation does not automatically provide independent evidence that the underlying conclusion is correct.

These distinctions suggest that the development of AI governance requires more than the accumulation of technical safeguards. It requires clarity concerning the function of each safeguard, the evidence demonstrating its effectiveness, and the institutional actor authorized to evaluate its results.

Meaningful oversight cannot be defined exclusively through the presence of a human approval mechanism. Its adequacy depends upon the decision-maker's competence, access to relevant information, practical capacity to challenge the output, and authority to reject or modify the proposed action. At the same time, comprehensive human review of every minor computational operation may be neither necessary nor institutionally proportionate. The appropriate framework must distinguish between functions suitable for bounded automation and determinations requiring the continued exercise of legally accountable judgment.

7.4. The Evidentiary Limits of Contemporary Research

The comparative analysis also identifies recurring methodological difficulties within AI governance scholarship.

The first concerns the movement from technological observation to legal conclusion. Evidence that an AI system performs a particular task does not necessarily establish that existing legal doctrine is incapable of addressing the consequences of that performance. Claims concerning structural gaps in responsibility require demonstration that the applicable legal mechanisms are inadequate, rather than merely difficult to apply.

The second difficulty concerns the generalization of empirical results. Performance benchmarks may establish differences among evaluated systems without demonstrating their comparative reliability across all professional activities. Survey results may reveal attitudes among respondents without establishing representative patterns throughout the relevant population. Experimental findings may identify correlations without conclusively establishing the mechanisms responsible for them. Documentary analyses may demonstrate

the presence or absence of publicly described technical capabilities without providing independent evidence of operational effectiveness. These limitations do not diminish the value of empirical research. They define the evidentiary conditions under which its findings may legitimately support legal and institutional conclusions.

The third difficulty concerns the validation of proposed governance mechanisms. A conceptual architecture may be logically coherent without having been implemented. An implemented safeguard may satisfy technical specifications without preventing the failures it was designed to address. An independently tested system may demonstrate satisfactory performance under specified conditions without establishing that those results will remain stable after subsequent modification or deployment in a different institutional environment. The reviewed literature provides examples of authors who expressly acknowledge these limitations and distinguish demonstrated results from prospective claims. Such transparency represents an important methodological contribution in its own right.

The fourth difficulty concerns the relationship between scholarly productivity and the verification of intellectual work. The increasing accessibility of generative technologies creates opportunities to accelerate research preparation, literature organization, drafting, and publication. These capabilities may be beneficial, but the quantity of scholarly output does not establish the quality of its intellectual contribution. Conversely, unusually high publication frequency does not, without additional evidence, demonstrate the absence of independent scholarship or appropriate source verification. The distinction is particularly relevant to open research repositories, which facilitate dissemination but do not necessarily provide the independent editorial or peer-review procedures associated with scholarly journals.

Accordingly, assessments of research quality must remain grounded in the examination of original texts, identifiable sources, legal arguments, methodological transparency, and the adequacy of supporting evidence. The number of manuscripts circulated through a repository is a bibliographic fact. It is not a substitute for a substantive scholarly assessment.

7.5. Institutional Implications and Directions for Further Research

The selected scholarship indicates that future AI governance research would benefit from a more systematic connection between legal doctrine, technical implementation, and independent empirical evaluation.

Several priorities emerge.

The first is the development of function-specific approaches to legal attribution and responsibility. Rather than treating artificial intelligence as a uniform regulatory category, legal analysis should distinguish among systems providing information, generating professional documents, recommending decisions, and executing operations with external legal or economic consequences. Different functions may require different standards of authorization, verification, professional supervision, and liability.

The second priority concerns the relationship between producer obligations and professional responsibility. The duties of a professional user should not automatically exclude possible obligations arising from the design, marketing, operation, or maintenance of the relevant technological system. Equally, the identification of potential producer responsibility should not eliminate established professional duties or obscure the legal requirements governing particular causes of action. Further research should therefore examine the allocation of duties according to actual functions, knowledge, control, foreseeable reliance, and the applicable legal framework.

The third priority is independent evaluation of verification and oversight architectures. Proposed mechanisms should be assessed not merely according to their conceptual completeness but through reproducible testing capable of identifying their operational limitations. Such evaluation should distinguish error prevention, detection, documentation, correction, and the subsequent attribution of responsibility.

The fourth priority concerns the proportionality and institutional consequences of regulation. Verification requirements, certification schemes, disclosure obligations, and technical controls impose costs that may be distributed unevenly across professional and commercial environments. Their effectiveness should therefore be examined alongside their implications for access to services, competition, confidentiality, procedural fairness, and institutional capacity.

Finally, the development of AI-assisted scholarship itself requires continued methodological attention. Generative tools can contribute to research organization, linguistic production, and the examination of alternative arguments. Their use does not establish the originality, reliability, or validity of the resulting academic work.[56]

The relevant scholarly requirements remain the identification of an original contribution, the appropriate use of primary sources, the accurate presentation of evidence, and the author's responsibility for the conclusions advanced. Questions concerning the scale, methods, and verification of AI-assisted academic production deserve independent empirical examination. They should not be resolved through assumptions based upon publication frequency alone.

7.6. Final Conclusions

A broader conclusion emerges from this comparative examination of contemporary legal scholarship. The debate surrounding artificial intelligence is undergoing a significant

[56] A non-exhaustive examination of SSRN records for the period October 11, 2025–October 11, 2026 identified at least 61 distinct SSRN identifiers attributed to Travis Gilly and 42 to Northon Salomão de Oliveira. These figures represent separately indexed records, not necessarily peer-reviewed publications or independent original works; revised versions of the same SSRN record were not counted separately. The itemized research inventory contains 103 identifiers. Publication frequency alone establishes neither generative-AI authorship nor any deficiency in scholarly verification. The methodological question concerns the demonstrable relationship between the volume of scholarly output, original intellectual contribution, and independent examination of sources. Source: SSRN, author-attributed records; author's non-exhaustive bibliographic inventory (accessed: 11.10.2026).

transformation. For years, much of its theoretical energy was directed toward questions of consciousness, autonomy, agency, and legal personhood. Could an artificial intelligence system possess rights? Could it become an independent legal subject? Could existing legal analogies, ranging from corporate personality to the recognition of non-human entities, accommodate a new category of artificial actors? These questions remain intellectually legitimate. Yet the scholarship examined in this study reveals a consequential shift in emphasis: **from the ontological status of artificial intelligence to the institutional conditions of its lawful and accountable use.**

The emerging questions are more immediate and, in many respects, more demanding. Who authorizes an automated decision? Who verifies the information upon which it relies? Who determines whether an artificial intelligence system is suitable for a particular professional or governmental function? Who bears responsibility when its outputs are inaccurate, discriminatory, procedurally defective, or otherwise harmful? And, perhaps most importantly, what institutional mechanisms can ensure that human oversight remains an effective exercise of judgment rather than a ceremonial act of approval?

This transition marks an important development in the relationship between legal theory and technological practice. General declarations of ethical principles, however necessary, cannot substitute for enforceable duties, identifiable decision-makers, independent verification, documentary traceability, and effective mechanisms of review. Nor can the mere presence of a human operator establish meaningful human control. Human-in-the-loop arrangements acquire legal significance only when the responsible person possesses the information, competence, authority, and practical opportunity to examine, challenge, correct, or reject the machine-generated result. Otherwise, human participation risks becoming a procedural fiction rather than a safeguard.

The same reasoning applies to the allocation of responsibility among developers, manufacturers, deployers, professional users, and public institutions. Their respective obligations may differ according to the applicable legal framework, but technological complexity cannot itself justify an accountability vacuum. The central task is therefore not to identify an artificial substitute for the responsible legal actor, but to construct a verifiable chain of authorization, control, validation, and attribution. In this emerging architecture, verification is not merely a technical operation. It becomes a condition of institutional legitimacy.

The present review does not suggest that debates over artificial consciousness or legal personality have been resolved, nor that they should be abandoned. Its more significant finding is that these debates are increasingly accompanied by a practical research agenda concerned with the conditions under which artificial intelligence can be deployed without dissolving established structures of legal responsibility. This development points toward a new stage in AI legal scholarship: one in which the central concern is no longer simply what artificial intelligence might become, but how human institutions must govern what it already does.

The future of AI law will be determined less by what we choose to call the machine than by what we can demonstrate about our ability to control, verify, and take responsibility for its actions.

The decisive question is no longer whether a machine can imitate human judgment. It is whether human institutions will retain the capacity — and accept the obligation — to exercise it.

References

1. Polenberg J. Judgment without a Subject: Artificial Intelligence and Legal Attribution. – 2026. – 58 p. – DOI: 10.2139/ssrn.7248440. – URL: https://ssrn.com/abstract=7248440 (accessed: 11.10.2026).
2. Oliveira N. S. de. Autonomous AI Agents and Legal Personality: Agency, Attribution, and the Limits of Artificial Intelligence Liability. – 2026. – 19 p. – DOI: 10.2139/ssrn.7355821. – URL: https://ssrn.com/abstract=7355821 (accessed: 11.10.2026).
3. Oliveira N. S. de. Artificial Intelligence, Autonomous Systems and the Future of Legal Responsibility. – 2026. – 16 p. – DOI: 10.2139/ssrn.7305740. – URL: https://ssrn.com/abstract=7305740 (accessed: 11.10.2026).
4. Glosser D. B. The Chip of Theseus: Human-AI Coupling and the Erosion of Legal Individuation. – 2026. – 21 p. – DOI: 10.2139/ssrn.7279242. – URL: https://ssrn.com/abstract=7279242 (accessed: 11.10.2026).
5. Alston E., Lehr W. On (Human) Adjudication. – 2026. – 29 p. – DOI: 10.2139/ssrn.7334643. – URL: https://ssrn.com/abstract=7334643 (accessed: 11.10.2026).
6. Allen J. G., Gamper F. Artificial Intelligence, Liability and Agency : SMU Centre for Digital Law Research Paper No. 02/2025. – 2025. – 18 p. – DOI: 10.2139/ssrn.7457561. – URL: https://ssrn.com/abstract=7457561 (accessed: 11.10.2026).
7. Bryant M. J. The AI Supervision Paradox: Double Standards, Algorithm Aversion, and the Unequal Burden of Legal Tech Regulation. – 2026. – 51 p. – URL: https://ssrn.com/abstract=7337558 (accessed: 11.10.2026).
8. Simshaw D. Legal Ethics of AI Snake Oil: Navigating the Hype, Harm, and Hope of Legal AI // Michigan Law Review. – 2026. – Vol. 124, no. 8. – P. 1393–1423. – DOI: 10.36644/mlr124.8.legal. – URL: https://repository.law.umich.edu/mlr/vol124/iss8/14 (accessed: 11.10.2026).
9. Sohail S. AI Hallucinations and Lawyer Liability: A Critical Analysis of Professional Responsibility, Regulatory Frameworks, and the Accountability Gap. – 2026. – 11 p. – DOI: 10.2139/ssrn.7010178. – URL: https://ssrn.com/abstract=7010178 (accessed: 11.10.2026).
10. Kildeev A. The Tool Has a Manufacturer: Producer Liability for Legal AI. – 2026. – 38 p. – DOI: 10.2139/ssrn.7388178. – URL: https://ssrn.com/abstract=7388178 (accessed: 11.10.2026).
11. American Bar Association. Standing Committee on Ethics and Professional Responsibility. Formal Opinion 512: Generative Artificial Intelligence Tools. – July

29, 2024. – URL: https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf (accessed: 11.10.2026).
12. Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on Liability for Defective Products and Repealing Council Directive 85/374/EEC // Official Journal of the European Union. – 2024. – 18 November. – URL: https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng (accessed: 11.10.2026).
13. Zaheer F. The Closed Intent Loop: Identity and Authorization in Autonomous Tool-Using AI Agents. – 2026. – 42 p. – DOI: 10.2139/ssrn.7418158. – URL: https://ssrn.com/abstract=7418158 (accessed: 11.10.2026).
14. Judicial Conference of the United States. Advisory Committee on Evidence Rules. Agenda Book for the October 15, 2026 Meeting. – 2026. – 221 p. – URL: https://www.uscourts.gov/sites/default/files/document/2026-10_evidence_rules_agenda_book_9-23_final.pdf (accessed: 11.10.2026).
15. Aran D., Perry R., Shelly S. Clinical Liability Cases Reveal a Coupling Between Legal Defensibility and Medical Procedure Escalation in Large Language Models // Communications Medicine. – 2026. – Article in press. – DOI: 10.1038/s43856-026-01839-5. – URL: https://doi.org/10.1038/s43856-026-01839-5 (accessed: 11.10.2026).
16. Unger W. AI Use in Legal Scholarship [preliminary abstract, survey instrument and appendices]. – 2026. – 15 p. – DOI: 10.2139/ssrn.7290238. – URL: https://ssrn.com/abstract=7290238 (accessed: 11.10.2026).
17. Bernstein G. From Bans to Recalls: A Public Health Framework for AI Companion Bots. – Washington, DC : Brookings Institution, 2026. – 16 p. – DOI: 10.2139/ssrn.6922118. – URL: https://www.brookings.edu/articles/from-bans-to-recalls-a-public-health-framework-for-ai-companion-bots/ (accessed: 11.10.2026).
18. Oliveira N. S. de. Medical Malpractice, Artificial Intelligence and Physician Liability: Who Is Responsible for AI-Assisted Medical Errors? – 2026. – 16 p. – DOI: 10.2139/ssrn.7262340. – URL: https://ssrn.com/abstract=7262340 (accessed: 11.10.2026).

ACADEMIA WITHOUT SCHOLARS: ARTIFICIAL INTELLIGENCE AND THE EROSION OF INDEPENDENT JUDGMENT

Kildeev Adel
Leningrad State University named after A.A. Zhdanov (1990)
PhD in Law
ORCID: 0009-0001-2211-4944

Crossref DOI: Pending Registration
SSRN DOI: https://doi.org/10.2139/ssrn.7620899

Abstract: This article examines the risk that contemporary academic institutions may preserve the visible products of scholarship while progressively weakening the human capacities required to produce, verify, and transmit knowledge independently. Rather than attributing academic mediocrity to generative artificial intelligence or declaring the collapse of modern universities, the study situates LLM-assisted scholarly production within longstanding problems of publication incentives, research integrity, institutional authority, and intellectual continuity.

Drawing upon epistemology, philosophy of language, legal reasoning, research on cognitive offloading and automation, and empirical evidence from the United States, Canada, and England, the article investigates the growing asymmetry between the speed of generating academically convincing texts and the human capacity to understand and independently verify their contents. Its empirical component examines changes in higher-education participation, measured educational skills, academic employment, doctoral career pathways, university financing, and research expenditure. These indicators reveal significant institutional transformations and contrasting trends, but do not directly measure independent scholarly judgment or establish causal cognitive deterioration attributable to artificial intelligence.

The analysis distinguishes legitimate technological assistance from the substantive substitution of human intellectual work. It argues that linguistic sophistication, AI-use disclosure, formal peer review, and nominal human oversight cannot independently establish scholarly competence or meaningful verification. Preserving the integrity of academic knowledge requires competent evaluation, accountable authorship, institutional independence, and sustained investment in research mentorship and intellectual continuity.

The study concludes that *academia without scholars* describes a possible institutional condition rather than an accomplished historical fact: universities may expand their measurable academic output while weakening the processes through which independent scientific judgment is cultivated and reproduced.

Keywords: Artificial Intelligence; Large Language Models; Independent Academic Judgment; Cognitive Offloading; Academic Verification; Research Integrity; Peer Review; Institutional Amnesia; Higher Education; Intellectual Continuity.

The use of the generative models ChatGPT (OpenAI), Gemini (Google), Grok (xAI) in drafting this article illustrates, rather than undermines, its central thesis: LLMs function as instrumenta vocalia—tools assisting in linguistic production without possessing understanding, intention, or authorship. Their involvement does not alter the locus of legal or intellectual responsibility, which remains exclusively with the human author, who makes all substantive, structural, and evaluative decisions. LLM outputs cannot be accorded independent legal or intellectual status apart from the human author.

Scope of Interdisciplinary Evidence

This article is a work of legal philosophy and institutional analysis. It does not undertake independent psychological, medical, or educational research, nor does it offer clinical assessments of cognitive impairment. Findings from these disciplines are examined as reported in published specialist studies, subject to their respective methodological limitations. Their inclusion serves exclusively to assess the possible implications of empirical evidence for independent scholarly judgment, epistemic responsibility, and academic institutions. No generalized cognitive deterioration or causal relationship is inferred beyond what the cited studies establish.

The question is," said Humpty Dumpty,
"which is to be master — that's all."

Lewis Carroll, *Through the Looking-Glass*
(1871), Chapter VI,
quoted by Lord Atkin in his dissent in
Liversidge v Anderson [1942] AC 206.

Chapter 1. Introduction: The Crisis of Independent Scholarship

The contemporary university occupies a paradoxical position. Never before have academic institutions possessed such extensive technological capabilities for the production, storage, distribution, and retrieval of information. Scholarly publications circulate across international repositories, educational programmes reach increasingly diverse populations, and digital systems enable access to research materials on a scale unimaginable to previous generations. Yet the expansion of access to information raises a question that cannot be answered by reference to technological progress or institutional output alone: to what extent do contemporary educational and academic institutions continue to cultivate the independent intellectual judgment upon which the production, verification, and transmission of knowledge ultimately depend?

This question becomes particularly important in the context of the widespread adoption and integration of artificial intelligence into education, scientific research, academic publishing, and professional intellectual activity. Generative large language models can produce sophisticated academic prose, summarize literature, organize arguments, and imitate many of the linguistic conventions traditionally associated with scholarly work. Their

capabilities offer considerable opportunities for research and education. At the same time, they create new possibilities for separating the production of academically recognizable texts from the independent intellectual activity historically required to produce, evaluate, and defend them. The existence of such possibilities does not establish that intellectual degradation is universal or inevitable. It does, however, require investigation into the institutional conditions under which technological assistance becomes a substitute for independent judgment.

The crisis of contemporary academia cannot be adequately explained by the widespread adoption and integration of artificial intelligence alone. Any serious investigation must place these technologies within the longer history of institutional transformation, changing educational economics, the expansion and restructuring of higher education, academic commercialization, and the growing importance of measurable institutional performance. Artificial intelligence enters an academic environment already shaped by decades of political decisions, economic pressures, administrative reforms, and changing expectations concerning the purposes of universities and scientific research.

1.1. Historical and Geographical Scope

The present study examines the period from 1990 to 2026. The selection of 1990 is deliberate, but it does not imply that the problems of contemporary education originated in that year. Earlier developments in university governance, educational financing, academic professionalization, and scientific organization remain historically relevant. A comprehensive account of those developments, however, would exceed the purposes and reasonable scope of the present article.

The beginning of the 1990s provides a meaningful historical reference point because it coincided with a major transformation of the European political and institutional order. The dissolution of the socialist political systems of Central and Eastern Europe, followed by the collapse of the Soviet Union in 1991, initiated substantial changes in national economic structures, public administration, institutional financing, and educational governance. These transformations did not occur uniformly, nor did they produce identical consequences across different countries. Nevertheless, they provide an important context for examining the restructuring of educational institutions during the subsequent decades.

The same period witnessed broader changes in international economic integration, university financing, educational mobility, and the organization of higher education. In Europe, the Bologna Declaration of 1999 marked a significant stage in efforts to create more comparable higher education qualifications and facilitate academic mobility. Subsequent reforms modified degree structures, accreditation arrangements, and institutional expectations across participating countries. Whether these changes strengthened educational quality, weakened particular forms of disciplinary preparation, or produced different effects across national systems remains a question for empirical examination rather than an assumption adopted in advance.

Although European developments provide an important historical reference point, the systematic empirical comparison is confined to **the United States**, **Canada**, and **England**. Sources concerning other countries may be used for historical, philosophical, or methodological context, but they do not enlarge the comparative sample. The article does not

claim to provide an exhaustive history of education in every country, nor does it presume that national systems experienced a common trajectory of institutional decline.

The terminal date of 2026 identifies the contemporary context of the investigation. Where statistical series or institutional studies do not extend to that year, the analysis will employ the most recent available evidence and identify the actual observation period.

1.2. Subject, Research Questions, and Central Hypothesis

The subject of this article is the transformation of the institutional conditions under which independent scholarly judgment is developed, exercised, verified, and transmitted between generations.

For the purposes of this study, independent scholarly judgment denotes the capacity to formulate research questions, critically examine sources, assess competing explanations, distinguish evidence from assertion, and defend conclusions through reasoned argument. Such independence does not require intellectual isolation or freedom from technological assistance. Scientific knowledge has always depended upon instruments, collaboration, institutional resources, and inherited intellectual traditions. The decisive distinction concerns whether the researcher retains the competence and responsibility necessary to evaluate the work performed with those instruments.

The investigation addresses three interconnected questions. First, how have changes in educational financing, institutional governance, degree structures, and academic incentives affected the conditions of scholarly preparation and research since 1990? Second, what do longitudinal educational assessments, sociological investigations, and studies of scientific activity reveal about the development of relevant intellectual competencies and the institutional production of knowledge? Third, how might the widespread adoption of generative AI interact with these existing conditions, either by supporting independent inquiry or by encouraging the delegation of intellectual tasks without corresponding verification?

The central hypothesis is that the expansion of formal educational provision and scholarly output does not necessarily entail an equivalent strengthening of independent intellectual competence. Where institutional incentives increasingly privilege measurable production over the demonstrable quality of research and judgment, educational systems may become vulnerable to a divergence between the appearance of academic productivity and the substantive reproduction of scholarly knowledge. Generative AI may intensify this divergence by reducing the effort required to produce plausible academic language without necessarily ensuring the understanding, originality, or evidentiary reliability of the resulting work.

This proposition is advanced as a hypothesis requiring critical examination, not as a predetermined conclusion concerning the condition of contemporary universities.

1.3. Research Methodology and Limitations

The article employs an interdisciplinary methodology combining comparative historical-institutional analysis, secondary analysis of educational and sociological evidence, critical examination of scholarship on academic organization and scientific productivity, and

philosophical and jurisprudential analysis of independent judgment and institutional authority.

The historical-economic component compares selected indicators of educational financing, participation, institutional organization, and academic employment around the beginning of the 1990s with the most recent available data. Intermediate developments are considered where necessary to explain significant reforms or changes in institutional conditions. These comparisons will distinguish total expenditure from expenditure per student, public financing from private contributions, and quantitative expansion from qualitative educational outcomes.

The sociological component draws upon established international assessments and research concerning educational competencies, university practices, academic employment, and scientific production. Differences in national coverage, sampling, measurement, and historical comparability will be acknowledged. Educational test results, publication statistics, and institutional performance indicators will not be treated as interchangeable measures of intellectual capacity or scholarly quality.

The philosophical and jurisprudential component examines the relationship between intellectual judgment, the production of meaning, and the authority claimed by institutions that certify knowledge. Particular attention is given to the distinction between the formal recognition of academic achievement and the substantive intellectual competence that such recognition is intended to represent.

The legal-philosophical foundations of the present study were developed in the author's earlier works. In *Artificial Intelligence Is Slave of the Lamp*, the doctrine of *instrumentum vocale* was employed to distinguish technologically sophisticated instruments from legal subjects capable of independent will and responsibility.[1] *LLM Is Not Artificial Intelligence* extended this distinction to the epistemological status of large language models, examining the difference between linguistic performance, understanding, and independent judgment.[2] The subsequent joint study, *Instrumentum Vocale and the Architecture of Responsibility: From Liability to Embedded Accountability*, developed an institutional framework of verification, attribution, and non-delegable human responsibility.[3] The present article applies these foundations to academic knowledge production, scholarly authorship, independent verification, and the institutional certification of intellectual competence.

The analysis does not undertake original survey research, claim to establish a universal causal relationship between artificial intelligence and cognitive decline, or seek to identify individual political or institutional actors as personally responsible for a presumed global collapse of education. Its purpose is to evaluate documented institutional developments,

[1] Kildeev, A. (2025a). Artificial Intelligence Is Slave of the Lamp. Scientific Platform XXI Century, No. 6 (September 2025), 3–31. Published journal version. https://episs.ru/

[2] Kildeev, A. (2025b). LLM Is Not Artificial Intelligence. Scientific Platform XXI Century, No. 7 (October 2025), 3–44. Published journal version. https://episs.ru/

[3] Liebig, T., & Kildeev, A. (2026). Instrumentum Vocale and the Architecture of Responsibility: From Liability to Embedded Accountability. Canadian Review of Artificial Intelligence, Law & Society (CRAIL), 1(1), 11–84. Published journal version. https://crail.ca/

identify plausible mechanisms affecting the reproduction of knowledge, and distinguish demonstrable findings from interpretive conclusions.

The broader question is therefore not whether modern universities produce more publications, award more qualifications, or employ more advanced technologies than their predecessors. It is whether the institutions responsible for certifying knowledge continue to preserve the human intellectual capabilities necessary to justify that certification.

The survival of academia as an institutional system should not be confused, without further examination, with the preservation of scholarship as an independent intellectual practice.

Chapter 2. The University, Scholarly Apprenticeship, and the Continuity of Knowledge

2.1. The Historical Foundations of the University

The contemporary university did not emerge as an institution primarily designed to issue academic qualifications, generate publications, or satisfy the immediate demands of the labour market. Its historical origins lie in the organization of advanced learning, the preservation and interpretation of accumulated knowledge, and the preparation of individuals capable of exercising intellectual and professional judgment.

The European university developed during the Middle Ages through communities of teachers and scholars whose activities gradually acquired institutional recognition and legal privileges. In England, teaching at Oxford can be traced to 1096, while Cambridge emerged as a scholarly community in 1209. These institutions inherited and developed traditions of advanced learning already established in continental Europe. Their educational activities included the liberal arts, theology, law, and medicine.[4]

The medieval university differed fundamentally from its modern counterpart. Its intellectual organization depended heavily upon authoritative texts, lectures, commentaries, and formal disputations. Nevertheless, an important feature of this educational tradition deserves emphasis: academic preparation was not confined to the passive reception of information.

At Cambridge, for example, medieval examinations involved oral disputations in which candidates defended propositions against objections raised by other scholars and their teachers. Academic qualifications were connected to participation in recognized intellectual procedures rather than merely to the completion of a prescribed period of attendance.

Such practices did not guarantee intellectual independence. Medieval universities operated within religious, political, and social hierarchies that frequently constrained inquiry. They were neither universally accessible nor institutions of unrestricted intellectual freedom.

[4] University of Oxford. Our History. https://www.ox.ac.uk/about/the-university/history ; University of Cambridge. The Medieval University. https://www.cam.ac.uk/about-the-university/history/the-medieval-university

Their historical significance lies elsewhere: they established durable organizational mechanisms through which knowledge, disciplinary methods, standards of argumentation, and scholarly qualifications could be transmitted between generations.

During the early modern period, the development of experimental science introduced additional principles of inquiry. The establishment of the Royal Society in England in 1660, followed by the adoption of its motto *Nullius in verba* in 1662, expressed a commitment to examining scientific propositions through observation and experiment rather than accepting them solely upon the authority of established figures.[5]

The Royal Society was not itself a university. Its significance for the present discussion lies in demonstrating that the development of modern scientific inquiry required more than the institutional preservation of received knowledge. It demanded procedures through which established propositions could be questioned, tested, and potentially rejected.

This distinction between the possession of knowledge and the capacity to examine its foundations would become essential to the subsequent development of the research university.

2.2. The University Tradition in North America

The transplantation of European educational institutions to North America produced different national trajectories, but the intellectual influence of English university traditions remained substantial.

Harvard College, founded in Massachusetts in 1636, emerged within the institutional and religious environment of English colonial society. Its early educational purposes reflected the preparation of clergy and educated members of colonial communities. The institution did not begin as a research university in the modern sense. Its later transformation formed part of a much broader evolution of American higher education.[6]

In Canada, British institutional traditions were reflected in the establishment of King's College in Upper Canada by royal charter in 1827. The institution subsequently became the University of Toronto, with its development shaped by disputes concerning religious authority, public education, and the secularization of higher learning.

Canadian university history, however, cannot be understood exclusively through English institutional inheritance. The French-language educational tradition developed through institutions such as the Séminaire de Québec, founded in 1663, from which Université Laval emerged in 1852.[7]

These historical differences are relevant because they demonstrate that the universities of England, the United States, and Canada did not develop according to a single institutional model.

---

[5] The Royal Society. History of the Royal Society. https://royalsociety.org/about-us/who-we-are/history/

[6] Harvard University. The History of Harvard University. https://www.harvard.edu/about/history/

[7] University of Toronto Archives & Records Management Services. The University's Original Charter. https://utarms.library.utoronto.ca/history/university-original-charter ; Université Laval. Origin and History. https://www.ulaval.ca/en/about-us/origin-and-history

The modern research university itself emerged through a combination of intellectual traditions and institutional innovations. During the nineteenth century, the relationship between university instruction and original scientific investigation became increasingly important. Research was no longer understood merely as an activity conducted by individual scholars alongside teaching; it progressively became an institutional responsibility.

The foundation of Johns Hopkins University in Baltimore in 1876 represented an important stage in this development. Its educational philosophy emphasized original investigation, advanced scholarly preparation, and the development of researchers capable of extending existing knowledge.[8]

The significance of this transformation cannot be reduced to the creation of specialized laboratories or graduate programmes. It concerned the intellectual identity of the university. The institution increasingly assumed responsibility not only for transmitting established knowledge but also for sustaining the human and organizational conditions required to produce new knowledge.

By the late twentieth century, universities in England, the United States, and Canada had inherited institutions shaped by centuries of teaching, scholarly argumentation, professional preparation, research, and intellectual exchange. This inheritance was neither homogeneous nor immune to criticism. Nevertheless, it provides the historical context for examining the transformations that occurred during the period addressed by the present article.

2.3. Scientific Schools, Apprenticeship, and the Formation of Scholars

The reproduction of scientific knowledge involves more than the preservation of books, archives, experimental results, and published research.

It also requires the transmission of the intellectual capacities necessary to understand, evaluate, and extend that knowledge. This distinction is particularly important in considering scientific schools and academic apprenticeship.

For the purposes of the present study, a *scientific school* is not understood as a formally designated institution or a group united by unquestioning adherence to the authority of a particular professor. Rather, it denotes a continuing intellectual community in which research problems, methodological standards, disciplinary practices, and forms of scholarly judgment are developed and transmitted across generations.

Such communities may exist within laboratories, departments, research institutes, doctoral programmes, and networks of scholarly collaboration. Their defining feature is intellectual continuity. The preparation of a researcher therefore involves a process fundamentally different from the simple acquisition of information. An aspiring scholar must learn to formulate research questions, distinguish significant problems from trivial ones, identify relevant sources, evaluate competing explanations, and recognize the limitations of available evidence.

[8] Johns Hopkins University. History & Mission. https://www.jhu.edu/about/history/

These capacities cannot be acquired solely through the memorization of established propositions.

Michael Polanyi's conception of *tacit knowledge* provides a philosophical foundation for understanding this process. In *The Tacit Dimension*, Polanyi emphasized that human knowledge includes practical and intellectual capacities that cannot be completely reduced to explicit verbal formulations or codified instructions.[9]

Applied to scholarship, this distinction suggests that familiarity with methodological rules does not necessarily confer the capacity to exercise sound methodological judgment. A researcher may know the formal requirements for scientific citation without understanding the evidentiary significance of a primary source. A doctoral candidate may reproduce accepted theoretical propositions without possessing the competence to challenge their foundations. Similarly, the ability to describe an experimental procedure does not necessarily establish the capacity to evaluate an unexpected result.

Jean Lave and Etienne Wenger's theory of *situated learning* provides a complementary perspective. Their concept of legitimate peripheral participation explains how newcomers acquire knowledge and competence through progressively fuller participation in communities of practice.[10]

Within academic research, such participation may involve seminars, laboratory work, supervised investigations, scholarly correspondence, conferences, critical discussion, and independent publication. The significance of these activities lies not merely in their formal completion, but in the gradual development of intellectual responsibility.

Research supervision, when properly exercised, is therefore not simply an administrative procedure for ensuring that a thesis satisfies institutional requirements. It is a form of intellectual apprenticeship. Its purpose is to develop the researcher's capacity to reach and defend conclusions through independent examination of evidence.

This understanding is supported by the United States National Academies' report *The Science of Effective Mentorship in STEMM* (2019), which identifies mentoring as an important component of research training and professional development. The report also emphasizes that effective mentorship requires institutional attention and cannot be assumed to arise automatically from the existence of academic hierarchies.[11]

It would nevertheless be mistaken to idealize traditional scholarly apprenticeship. The dependence of junior researchers upon established professors may reproduce intellectual conformity, discourage criticism, or reinforce institutional inequalities. The existence of a scientific school does not, by itself, guarantee the quality or independence of its members.

---

[9] Polanyi M. The Tacit Dimension. Garden City, NY: Doubleday, 1966. Reissued by the University of Chicago Press, 2009.

[10] Lave J., Wenger E. Situated Learning: Legitimate Peripheral Participation. Cambridge: Cambridge University Press, 1991. DOI: 10.1017/CBO9780511815355.

[11] National Academies of Sciences, Engineering, and Medicine. The Science of Effective Mentorship in STEMM. Washington, DC: National Academies Press, 2019. DOI: 10.17226/25568.

The decisive criterion is whether the educational relationship develops the intellectual autonomy of the succeeding generation.

A scientific school fulfils its purpose not when its students faithfully reproduce the conclusions of their teachers, but when they acquire the competence to question, improve, and, where necessary, refute them.

2.4. Fundamental Research and the Institutional Reproduction of Knowledge

The relationship between fundamental research and scholarly education constitutes another essential dimension of the university's intellectual function.

Fundamental research is characterized by inquiry directed toward the development of knowledge whose ultimate significance, applications, or practical consequences may not be immediately predictable.

Its value cannot always be measured through short-term economic returns, commercial applications, or predetermined institutional performance targets.

This does not mean that fundamental research should be exempt from evaluation or public accountability. It means that the criteria appropriate to its assessment must reflect the uncertainty, complexity, and potentially long-term significance of scientific discovery.

Universities occupy a distinctive position in this process because they combine research activity with the preparation of future researchers.

The relationship between these functions has repeatedly been emphasized in educational policy. In the United States, the Boyer Commission's 1998 report, *Reinventing Undergraduate Education: A Blueprint for America's Research Universities*, argued that research universities should integrate the process of discovery more directly into undergraduate education. Its recommendations addressed inquiry-based learning, research participation, and the development of students as active participants in intellectual investigation.[12]

Importantly, the report identified institutional shortcomings that already existed during the 1990s. It therefore provides historical evidence that concerns about the relationship between university research and the quality of education long preceded the widespread deployment of generative artificial intelligence.

In Canada, the Natural Sciences and Engineering Research Council's Discovery Grants programme reflects an institutional model in which sustained research activity and the training of highly qualified personnel are treated as interconnected objectives. The programme supports continuing research agendas and recognizes research training as a relevant element in assessing scholarly activity.[13]

---

[12] Boyer Commission on Educating Undergraduates in the Research University. Reinventing Undergraduate Education: A Blueprint for America's Research Universities. Stony Brook, NY: State University of New York at Stony Brook, 1998. https://repo.library.stonybrook.edu/xmlui/handle/11401/57972

[13] Natural Sciences and Engineering Research Council of Canada. Discovery Grants Program (Individual). Official programme documentation. https://nserc.canada.ca/en/funding-opportunity/discovery-grants-program-individual

In England, Research England's quality-related research funding provides institutional support that universities may allocate toward research priorities, infrastructure, and postgraduate research supervision. This arrangement forms part of a broader research-funding structure combining institutional resources with competitively awarded project funding.[14]

These examples do not establish that the three systems provide equivalent levels of financial support or achieve comparable educational outcomes.

Nor do they demonstrate that contemporary research funding necessarily protects independent intellectual inquiry. Their significance is more limited and more precise: they confirm that the institutional relationship between the production of knowledge and the preparation of future researchers remains an explicit objective of university research policy.

The effectiveness of that relationship must be evaluated separately. A university may expand its research budget while reducing the time available for meaningful supervision. It may develop commercially successful research partnerships while placing increasing pressure upon scholars to pursue projects capable of attracting external funding.

Conversely, institutional reforms may strengthen research infrastructure, broaden participation, and create new opportunities for scholarly collaboration. These possibilities require empirical examination rather than assumptions concerning inevitable improvement or decline.

2.5. Intellectual Continuity as an Institutional Responsibility

The historical development of the university reveals a persistent distinction between the accumulation of knowledge and the reproduction of the human competence necessary to sustain it.

Libraries may preserve texts without ensuring that succeeding generations understand their significance. Research repositories may accumulate publications without establishing the originality or reliability of their contents. Universities may confer qualifications without necessarily demonstrating that every recipient has developed the capacity for independent scholarly judgment.

These observations do not constitute evidence of universal academic deterioration.

They identify a structural problem requiring investigation. The survival of scholarly knowledge depends upon the continuity of intellectual practices through which propositions are examined, sources are verified, methodological decisions are justified, and conclusions remain open to criticism.

This continuity cannot be established solely by the number of academic institutions, students, degrees, or published articles. Such indicators measure important dimensions of educational activity. They do not independently measure the quality of scholarly judgment.

[14] UK Research and Innovation / Research England. How We Fund Higher Education Providers. https://www.ukri.org/publications/research-england-how-we-fund-higher-education-providers/how-we-fund-higher-education-providers/

Accordingly, the present investigation distinguishes between three related but analytically separate processes: the expansion of educational institutions, the production of formal academic outputs, and the reproduction of independently competent scholars.

The historical discussion in this chapter establishes the intellectual and institutional background of these processes. It does not extend the empirical research beyond the period 1990–2026 or beyond the three selected educational systems: England, the United States, and Canada.

The quantitative evidence concerning their transformation is presented in Exhibits A and B, while the methodological limitations and boundaries of inference are specified in Exhibit C.

The central question is not whether universities at the beginning of the 1990s were necessarily superior to their contemporary counterparts. It is whether subsequent changes in university organization, financing, research incentives, and educational practice have strengthened or weakened the conditions required for independent scholarly judgment.

The capacity of an institution to produce academic qualifications and publications must not be confused with its capacity to reproduce independent scholars.

This distinction provides the conceptual foundation for the following examination of educational expansion, institutional commercialization, changing research incentives, and the political economy of contemporary academia.

Chapter 3. From Knowledge to Academic Production

3.1. The Transformation of the University: From Intellectual Institution to Measurable Enterprise

The historical development examined in the preceding chapter established the university as an institution whose responsibilities extend beyond the transmission of established information. The preservation of scholarly traditions, the cultivation of independent judgment, the preparation of successive generations of researchers, and the pursuit of fundamental knowledge constitute interdependent elements of its intellectual mission.

During the period 1990–2026, universities in the United States, Canada, and England experienced substantial changes in their organization, financing, governance, and systems of evaluation.

These transformations occurred under different national conditions. They cannot be reduced to a single policy programme or explained exclusively through the influence of market economics. Nevertheless, they raise a common question: how does an institution devoted to producing and transmitting knowledge preserve its intellectual purpose when its performance is increasingly expressed through measurable administrative and economic indicators?

The distinction is not between universities that operate with financial resources and those that do not. Universities have always required funding, institutional administration, and mechanisms for evaluating educational performance.

Rather, the issue concerns the relationship between institutional means and intellectual ends. Enrolment numbers, degrees awarded, research grants obtained, publications produced, citation indicators, patents, and external partnerships may all provide valuable information about university activity. Yet they describe different aspects of institutional performance and cannot be treated as interchangeable measures of educational or scientific quality.

A university's capacity to attract students does not independently demonstrate the quality of its teaching. Its research income does not establish the originality of its discoveries. Its publication volume does not necessarily reveal whether its scholars have advanced understanding of a particular problem.

The danger arises when indicators intended to inform institutional judgment begin to replace that judgment. This distinction provides the analytical foundation for examining the contemporary movement from knowledge as an intellectual responsibility toward academic production as a system of measurable outputs.

3.2. The Changing Political Economy of Higher Education, 1990–2026

The transformation of university financing represents one of the most significant institutional developments of the period under examination.

In the United States, higher education operates through a complex combination of state universities, private nonprofit institutions, community colleges, and other educational providers. Their revenue structures vary substantially, as do their research activities, governance arrangements, and dependence upon tuition.

Consequently, American higher education cannot be described accurately through a single national funding model.

Nevertheless, the State Higher Education Executive Officers Association's *State Higher Education Finance* reports document the long-term relationship between state appropriations, public university enrolment, and net tuition revenue. The fiscal year 2025 report provides historical financial series extending to 1980, allowing developments within the present study's 1990–2026 period to be examined against a consistent institutional framework.[15]

These records demonstrate the importance of distinguishing aggregate public expenditure from funding available per student, and gross tuition charges from net tuition revenue. Changes in enrolment, financial aid, institutional costs, and state-level policy may alter the interpretation of apparently straightforward funding trends.

The American case therefore illustrates a broader methodological problem: the growth of educational expenditure does not, without additional evidence, establish an equivalent expansion of resources available for independent academic work.

Canada presents a different institutional structure. Universities operate predominantly within systems of provincial responsibility, supplemented by federal research funding, institutional revenues, philanthropy, and other sources. Statistics Canada recorded that student fees represented approximately 12% of Canadian university revenues in 1990/91,

[15] State Higher Education Executive Officers Association. State Higher Education Finance: FY 2025. Published 2026. https://shef.sheeo.org/report/

while total government sources accounted for approximately 69%. By 2002/03, student fees had reached 20.5% of revenues. More recent administrative data reported tuition revenues amounting to 30.1% of total university revenues in 2023/24.[16]

These figures provide evidence of an important long-term change in the composition of university financing. They should not, however, be interpreted as a completely harmonized time series without further examination of historical accounting definitions, institutional coverage, and revenue classifications.

The Canadian experience also reveals substantial provincial variation. Quebec, Ontario, British Columbia, and the Atlantic provinces have followed different financing trajectories, making national averages informative but insufficient for explaining the circumstances of particular universities.

England experienced especially visible legislative and financial restructuring.

The Further and Higher Education Act 1992 reorganized important elements of higher education governance and enabled former polytechnics to obtain university status. This change must be understood as an institutional reclassification and transformation, rather than automatically treated as evidence of the creation of equivalent numbers of new research universities.[17]

The introduction of undergraduate tuition fees in 1998, subsequent changes in 2006, and the increase of the standard upper fee limit to £9,000 in 2012 significantly altered the relationship between students, universities, and public financing.

Later reforms established the Office for Students as the principal regulator of higher education providers in England, reflecting the changed regulatory environment associated with tuition finance, institutional competition, and reduced dependence upon direct teaching grants.[18] The financial consequences remain visible in contemporary data.

A September 2026 House of Commons Library briefing recorded that tuition fees had become a substantially larger component of higher education providers' income, while the relative contribution from direct funding-body grants had declined. The same report documented financial deficits among a significant proportion of publicly funded English universities in 2024/25.[19]

The broader financial series published with the parliamentary briefing include United Kingdom aggregates. Such series must not be presented as exclusively English observations.

[16] Statistics Canada. University Finances, 2002/03. The Daily, 19 August 2004; University Revenues Increase More Than Expenditures in 2023/2024. The Daily, 29 July 2025.

[17] Further and Higher Education Act 1992, c. 13. https://www.legislation.gov.uk/ukpga/1992/13/contents

[18] House of Commons Library. Tuition Fees in England: History, Debates, and International Comparisons. Research Briefing CBP-10155, 23 February 2026; Higher Education and Research Act 2017, c. 29.

[19] Bolton, P., and Lewis, J. Higher Education Finances and Funding in England. House of Commons Library, Research Briefing CBP-10037, 4 September 2026.

The present study distinguishes England-specific evidence from UK-wide contextual statistics.

Across the three countries, the resulting institutional arrangements differ considerably. What connects them is not an identical process of privatization or the disappearance of public funding, but the increasing importance of complex revenue structures, financial sustainability, external accountability, and institutional competition. Whether these changes improved or weakened scholarly competence must be investigated rather than presumed.

3.3. Academic Capitalism and the Commercialization of Knowledge

The transformation of university financing has been accompanied by changes in the relationship between academic institutions and the wider economy.

Sheila Slaughter and Gary Rhoades developed the concept of *academic capitalism* to examine how universities participate in markets for educational services, research products, intellectual property, and commercially valuable knowledge. Their work emphasizes that market-oriented academic activity is not necessarily external to public institutions; it can develop through relationships involving universities, governments, commercial enterprises, and research organizations.[20]

Derek Bok, in *Universities in the Marketplace* (2003), approached the commercialization of higher education through its implications for academic values, research independence, educational priorities, and institutional integrity. His argument was not that universities should be isolated from economic activity, but that financial opportunities create conflicts requiring explicit safeguards.[21]

These contributions identify a fundamental distinction between the economic value of scientific knowledge and the intellectual responsibilities of the institutions producing it.

The commercial application of discoveries may provide substantial public benefits. University research can contribute to medical innovation, technological development, industrial productivity, and economic growth. External partnerships may provide facilities, financial resources, and opportunities that would otherwise be unavailable.

Consequently, commercialization cannot be treated as inherently incompatible with academic freedom or scientific progress. The relevant question is whether commercial objectives remain subordinate to the standards governing independent inquiry.

A research institution may legitimately cooperate with private industry while protecting investigators' freedom to publish inconvenient findings. It may develop intellectual property without allowing proprietary interests to determine the validity of scientific conclusions. Conversely, difficulties arise when the terms of financing, contractual restrictions, reputational interests, or anticipated commercial returns influence the selection, interpretation, or disclosure of research results.

---

[20] Slaughter, S., and Rhoades, G. Academic Capitalism and the New Economy: Markets, State, and Higher Education. Johns Hopkins University Press, 2004.

[21] Bok, D. Universities in the Marketplace: The Commercialization of Higher Education. Princeton University Press, 2003.

These are questions of governance and institutional responsibility, not merely questions of public versus private ownership. The university's scientific authority derives from the quality and independence of its research, not from the commercial value of its discoveries. The distinction becomes particularly important when universities increasingly evaluate departments and researchers through indicators associated with external income, intellectual property, publication activity, and institutional reputation.

3.4. The Academic Production System: Publications, Grants, and Performance Indicators

Scientific publications constitute an indispensable means of communicating research.

They permit findings to be examined, criticized, replicated, extended, and incorporated into continuing scholarly discussion. Without publication, scientific knowledge would lose one of its principal mechanisms of intellectual continuity.

The problem is therefore not publication itself, nor the legitimate expectation that publicly supported researchers should account for their activities. It arises when the numerical production of academic outputs becomes a substitute for the substantive evaluation of research.

A publication count establishes that a certain number of documents have appeared. It does not independently establish their originality, methodological reliability, or importance. Citation counts provide information about patterns of attention and scholarly reference. They may be influenced by disciplinary practices, research-community size, the age of a publication, methodological controversies, or the visibility of particular topics.

Journal reputation may provide contextual information about editorial processes, but it cannot substitute for examination of an individual article. These distinctions have been explicitly recognized within the scientific community.

The *San Francisco Declaration on Research Assessment* (DORA), initiated in 2012, challenged the use of journal-based indicators as proxies for the quality of individual articles or researchers. Its recommendations address hiring, promotion, funding, and institutional research evaluation.[22]

The *Leiden Manifesto for Research Metrics*, published in *Nature* in 2015, similarly emphasized the need for quantitative indicators to support expert assessment rather than replace it. It called attention to disciplinary differences, transparency, verification, and the risks associated with poorly designed evaluation practices.[23]

These initiatives are important for the historical argument of the present article. They demonstrate that concerns about excessive reliance upon academic performance indicators were articulated well before generative artificial intelligence became widely available.

[22] San Francisco Declaration on Research Assessment (DORA), 2012/2013. https://sfdora.org/read/

[23] Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., and Rafols, I. "Bibliometrics: The Leiden Manifesto for Research Metrics." Nature, 520 (2015): 429–431. https://doi.org/10.1038/520429a

The problems identified were therefore not created by language models. They arose within institutional systems already attempting to reconcile research quality, accountability, financial competition, and measurable performance.

England offers a particularly relevant example of formalized research evaluation. The Research Excellence Framework, first conducted in 2014, replaced the earlier Research Assessment Exercise and became a major mechanism for evaluating university research. Its results inform the allocation of public research funding. The framework covers the United Kingdom, although Research England administers the relevant funding arrangements for English institutions.[24]

Such systems serve legitimate public purposes, including accountability, the identification of research strengths, and the distribution of limited resources. Their existence does not demonstrate that British universities have abandoned scholarly judgment.

Nevertheless, any assessment framework may influence institutional priorities, career incentives, the organization of research, and decisions about which activities receive financial or administrative support. The relevant question is whether research assessment preserves the distinction between evidence of academic activity and judgment concerning academic quality.

An institution may be highly successful in generating measurable research outputs without necessarily demonstrating an equivalent capacity for original intellectual discovery. The present investigation therefore treats publication volume, research income, and evaluation indicators as evidence of institutional activity rather than direct measures of scientific competence.

3.5. Academic Labour, Mentorship, and Institutional Time

The transformation of university governance and financing must also be considered in relation to academic employment. Research and teaching depend upon human labour. The preparation of new scholars requires sustained supervision, intellectual discussion, feedback, and opportunities for participation in research.

These activities consume resources and, importantly, time. Changes in academic employment arrangements may therefore have consequences extending beyond employment policy.

In the United States, a 2023 analysis published by the American Association of University Professors, drawing upon National Center for Education Statistics data, documented the extensive use of contingent academic appointments. Its findings showed that approximately 68% of faculty members in American colleges and universities held contingent appointments in 2021.[25]

[24] UK Research and Innovation. Research Excellence Framework. Official documentation, updated 2026. https://www.ukri.org/who-we-are/research-england/research-excellence/research-excellence-framework/

[25] Colby, G. “Data Snapshot: Tenure and Contingency in US Higher Education.” Academe, American Association of University Professors, Spring 2023.

The figure refers to a broad higher education workforce and does not describe the composition of research-university faculty alone. It cannot establish the quality of supervision or research in particular institutions.

Nevertheless, it identifies an institutional development relevant to the present inquiry: a substantial proportion of teaching is performed under employment arrangements differing from the traditional tenured academic appointment.

In Canada, historical data published by Statistics Canada likewise recorded changes in the composition of the university teaching workforce during the 1990s, including an increase in part-time faculty alongside a reduction in full-time faculty between 1990/91 and 1997/98.[26]

These observations do not justify assuming that temporary or part-time academic personnel are less competent than tenured professors.

Employment status is not a measure of intellectual merit. The relevant concern is institutional continuity. Where teaching and research responsibilities are divided among personnel facing different contractual conditions, the university must ensure that supervision, research training, and scholarly mentorship remain adequately supported.

The distinction is particularly significant for scientific schools. Such communities depend upon more than individual intellectual achievement. They require continuity of research programmes, the progressive formation of junior scholars, and institutional opportunities for sustained inquiry.

A university may employ accomplished researchers while failing to provide adequate conditions for intellectual succession. Conversely, stable academic employment does not by itself guarantee effective mentorship. The appropriate question is whether the university's organizational arrangements enable its faculty to perform the intellectual functions for which the institution exists.

3.6. When Institutional Indicators Become Intellectual Objectives

The preceding evidence suggests a distinction between three forms of institutional accountability.

Financial accountability concerns the legitimate management and use of resources.

Administrative accountability concerns compliance with institutional and regulatory obligations.

Epistemic accountability concerns the justification of scholarly claims through appropriate methods, evidence, criticism, and verification.

All three are necessary.

They are not interchangeable.

An institution can satisfy accounting requirements while producing poor research. A publication may meet formal editorial requirements while containing unsupported

[26] Statistics Canada. "Part-time University Faculty, 1990/91 to 1997/98." The Daily, 8 May 2002.

conclusions. A research project may receive substantial external funding without producing significant intellectual advances.

Equally, important scientific findings may emerge from relatively modest projects whose value was not anticipated by their original financial or administrative assessments. The danger is not measurement itself.

It is the transformation of an indicator of scholarly activity into the governing objective of scholarship.

When universities are evaluated principally through what can be counted, their internal priorities may increasingly favour activities that generate measurable evidence of institutional success.

Such an adjustment may be rational from an administrative perspective without necessarily being optimal for the long-term development of scientific knowledge. This is an institutional hypothesis, not a demonstrated universal outcome. Different disciplines, universities, and funding systems may respond differently to similar incentives. Quantitative evaluation can expose previously hidden inequalities, identify ineffective programmes, and improve the accountability of public institutions.

The analytical task is therefore to determine the conditions under which measurement supports scholarly judgment and those under which it displaces it. That distinction becomes particularly important in examining academic publishing. The increasing capacity to produce documents, references, summaries, and apparently sophisticated arguments does not necessarily correspond to an increasing capacity to establish the truth of their contents.

The relevant problem existed before contemporary generative artificial intelligence. Technological developments may nevertheless alter its scale, speed, and institutional consequences. The present chapter does not yet attempt to establish those consequences. It identifies the academic environment into which these technologies were introduced.

3.7. From the Production of Knowledge to the Production of Academic Outputs

Between 1990 and 2026, the universities of the United States, Canada, and England underwent substantial institutional transformations.

The available evidence documents changes in educational financing, student contributions, research evaluation, academic employment, and relationships with commercial and governmental organizations. These changes were not uniform. Nor do they establish that contemporary universities have collectively abandoned their intellectual responsibilities. Some developments expanded educational access, strengthened research infrastructure, promoted public accountability, and created new opportunities for scientific collaboration. Others introduced or intensified questions concerning financial dependence, employment stability, institutional competition, and the relationship between measurable performance and scholarly competence.

The distinction established in Chapter II therefore acquires a new significance. A university may preserve the institutional appearance of scholarship while reorganizing the conditions under which scholarship is performed. It may increase the number of qualifications

awarded, publications produced, or research projects administered without thereby establishing that it has improved its capacity to develop independent scholars.

The central question is whether the university remains primarily an institution in which intellectual judgment is cultivated, tested, and transmitted, or whether those functions increasingly compete with objectives associated with measurable academic production.

The production of academic outputs is not identical to the production of knowledge; and the production of knowledge is not identical to the reproduction of independent scholars.

This threefold distinction will become essential in examining the relationship between academic publishing, institutional responsibility, and generative artificial intelligence. The challenge posed by contemporary technologies is not simply that academic documents can now be produced more rapidly. It is whether universities already organized around measurable outputs possess the institutional capacity to distinguish intellectual production from its simulation. That question forms the basis of the next stage of the investigation.

Chapter 4. The Abdication of Academic Judgment

The transformation of universities into increasingly sophisticated systems of academic production raises a question that cannot be answered by examining publication statistics, research budgets, or institutional rankings alone. What happens to scholarship when the production of its visible results becomes progressively detached from the intellectual processes that once gave those results their meaning? A university may publish more articles, award more degrees, attract greater investment, and expand its digital infrastructure without necessarily strengthening the capacity of its students and researchers to think independently. The difficulty lies in distinguishing the measurable products of academic activity from the less readily measurable qualities that make such activity intellectually valuable.

This distinction has acquired particular importance with the arrival of generative artificial intelligence. Large language models can now produce texts that resemble scholarly arguments, organize extensive bibliographies, formulate hypotheses, compare theoretical positions, and present their conclusions in language sufficiently polished to satisfy many of the formal conventions of academic writing. These capabilities are useful and, in competent hands, can substantially assist research. Yet the sophistication of the resulting document tells us remarkably little about the intellectual contribution of the person who submits it. The outward appearance of scholarship has become easier to reproduce than the independent judgment upon which scholarship ultimately depends.

There is no reason to idealize the academic past. Intellectual mediocrity, professional conformity, questionable research practices, and the pursuit of recognition without corresponding achievement are hardly inventions of the digital age. Universities have never been inhabited exclusively by brilliant scholars, nor have academic titles invariably reflected exceptional intellectual merit. What has changed is the technological environment in which scholarly work is produced and circulated. Generative systems can reduce the time required to assemble a plausible academic manuscript, potentially allowing the weaknesses of an already imperfect publication culture to manifest themselves on a substantially greater scale.

The central concern, therefore, is not whether scholars should use artificial intelligence. The author of this study openly acknowledges the use of such instruments in scholarly preparation, textual analysis, and editorial work. The relevant distinction concerns the intellectual functions that remain genuinely human. A researcher who employs an LLM to improve the clarity of an argument already developed and independently verified is engaged in a very different activity from someone who accepts a generated argument without understanding its assumptions, testing its evidence, or examining its conclusions. Both may submit polished manuscripts. Only the actual process of scholarly work can establish what intellectual contribution stands behind them.

This is where the problem of academic production becomes a problem of academic judgment. Once the appearance of intellectual accomplishment can be produced independently of the intellectual accomplishment itself, the university faces a challenge that extends beyond plagiarism, inaccurate citations, or violations of publication ethics. It must determine whether its established procedures can still distinguish genuine scholarship from its increasingly convincing imitation.

4.1. Cognitive Offloading and the Erosion of Independent Inquiry

Human beings have never developed knowledge without instruments. Writing extended the capacities of memory, libraries preserved the intellectual achievements of previous generations, and printing made knowledge available to audiences unimaginable in antiquity. Calculators relieved mathematicians and scientists of repetitive operations, while computers transformed the possibilities of scientific modelling, information retrieval, and experimental analysis. Each innovation changed the relationship between human thought and the instruments through which that thought was expressed. Yet none of these developments made the intellectual activity of the researcher irrelevant. The value of an instrument depended upon the questions being asked, the methods employed, and the capacity of a human being to interpret and evaluate the resulting information.

The use of external resources to reduce mental effort is commonly described as *cognitive offloading*. In their influential review, Risko and Gilbert examined this phenomenon as a normal feature of human cognition, emphasizing that individuals regularly employ external actions and resources to reduce the cognitive demands of particular tasks. Their analysis also demonstrates the importance of metacognition: people make judgments about their own abilities when deciding whether to perform an operation themselves or entrust it to an external resource. Such decisions may be entirely rational, although inaccurate assessments of one's own competence can also produce inappropriate reliance on external assistance.[27]

There is consequently little justification for treating cognitive offloading as an intellectual deficiency in itself. A scientist who uses statistical software does not thereby abandon statistical reasoning, any more than a historian who consults a digital archive relinquishes historical judgment. What matters is whether the researcher retains the understanding necessary to formulate the problem, select an appropriate method, recognize its limitations, and evaluate the results. The relationship becomes more complicated when the

[27] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

instrument begins to perform functions that previously belonged to the researcher's own reasoning, particularly the formulation of explanations, the organization of arguments, and the evaluation of competing interpretations.

Generative artificial intelligence introduces precisely this complication. Unlike conventional search engines or bibliographic databases, large language models do not merely retrieve information that the researcher must subsequently organize. They can produce complete argumentative structures, formulate plausible hypotheses, summarize disagreements, propose methodological approaches, and generate conclusions in polished academic language. Such capabilities can considerably improve research efficiency when used by an intellectually competent investigator. They can also create an environment in which the visible results of research appear before the intellectual work ordinarily associated with producing them has taken place.

Consider a graduate student beginning a dissertation in an unfamiliar field. Traditionally, the initial months of research would involve reading foundational works, discovering competing schools of thought, identifying unresolved questions, and gradually developing a defensible position. The process would almost certainly include misunderstandings, abandoned hypotheses, incomplete bibliographies, and repeated revisions. None of these difficulties would be particularly attractive to a university administrator concerned primarily with measurable productivity. Yet they would constitute an important part of the student's intellectual formation. By encountering and resolving difficulties, the student would gradually acquire the ability to distinguish an established proposition from a disputed interpretation, a promising hypothesis from an unsupported assertion, and an original contribution from the repetition of familiar ideas.

A generative system can now provide the same student with an apparently coherent research proposal, a theoretical framework, a provisional bibliography, and a sequence of conclusions within minutes. The result may be useful, particularly when treated as a collection of suggestions requiring independent examination. But if the student adopts the proposed framework before developing sufficient knowledge to evaluate it, the normal relationship between intellectual development and academic production becomes inverted. Instead of producing a scholarly argument through research, the student begins with an argument whose intellectual foundations remain to be investigated. The completed-looking document may create an illusion of progress even when the underlying understanding has scarcely advanced.

This distinction between receiving an answer and developing an understanding lies at the centre of Alejandro Sinopoli's *EPISTEMÍA: Governing the Human Process of Understanding* (2026).[28] Sinopoli approaches the problem primarily from legal philosophy and epistemology, arguing that the process through which human beings formulate questions, establish relevance, compare alternatives, revise interpretations, and form judgments possesses significance independent of the correctness of the final result.

[28] Sinopoli, A. (2026). EPISTEMÍA: Governing the human process of understanding: A proposal for a new object of legal protection in the age of artificial intelligence. The Journal Scientific Platform XXI Century, No. 5 (14), 40–72. https://episs.ru/wp-content/uploads/2026/09/5-14-ISSUE-2026-English-FINAL.pdf

Understanding, in his account, cannot be reduced to the possession of accurate information. It develops through an active and revisable process in which the individual participates in the construction and evaluation of meaning.[29]

Sinopoli's contribution is particularly valuable because it draws attention to a feature of generative systems that conventional discussions of automation frequently overlook: the temporal character of human understanding. Human reflection requires opportunities for hesitation, disagreement, reconsideration, and the gradual integration of unfamiliar information. Generative systems, by contrast, can produce answers and elaborate arguments at a speed that bears little relationship to the time required for a person to understand and independently assess them. The resulting difficulty is not simply that machines produce too much information. It is that the process of intellectual evaluation may become detached from the process of textual production.[30]

A researcher may receive a twenty-page analysis in a few minutes, but cannot responsibly verify twenty pages of unfamiliar legal, historical, or scientific argument merely because the material has arrived in a convenient and persuasive form. Checking references may require access to original publications; evaluating legal claims may require examination of legislation and judicial decisions; assessing scientific conclusions may demand familiarity with experimental methods, statistical assumptions, and competing interpretations. Some statements can be checked quickly, while others require substantial independent investigation. The apparent efficiency of automated drafting therefore creates a potential asymmetry between the speed at which scholarly claims can be generated and the human resources required to establish their reliability.

This asymmetry should not be mistaken for a universal mathematical relationship or an experimentally demonstrated law of cognitive decline. Its significance depends upon the complexity of the task, the availability of reliable sources, the competence of the reviewer, and the consequences of error. Certain routine checks can themselves be automated, and appropriately designed tools may improve rather than weaken human verification. Nevertheless, increasing the volume of plausible claims does not automatically increase the human capacity to understand them. Where institutions encourage rapid production without providing corresponding resources for independent evaluation, the imbalance can become a significant organizational problem.

Sinopoli develops another concept that deserves particular attention in this context: *cognitive friction*. He uses the expression to describe forms of productive resistance that prevent the premature closure of understanding. Doubt, disagreement, exposure to alternative explanations, the need to reformulate an argument, and the obligation to justify a conclusion can all contribute to the development of intellectual judgment. Such resistance is not valuable merely because it consumes time.

[29] Sinopoli. EPISTEMÍA, pp.40-45.

[30] Sinopoli. EPISTEMÍA, pp.43-48.

Needless administrative procedures, inaccessible information, and defective research tools can obstruct inquiry without improving understanding. The relevant distinction concerns difficulties that stimulate meaningful intellectual engagement rather than those that simply make academic work slower.[31]

The importance of this distinction becomes apparent when we consider the educational function of research. A student who discovers that an apparently convincing argument contradicts a primary source has encountered something more valuable than a corrected paragraph. The experience may transform their understanding of the subject, reveal the limitations of a particular method, or encourage greater caution in evaluating future claims. A professor who identifies an unexpected weakness in a long-established hypothesis may likewise be compelled to reconsider assumptions that have shaped years of work. Intellectual progress frequently depends upon these moments of resistance, when an explanation ceases to be satisfactory and a more demanding investigation becomes necessary.

Generative systems can assist such processes by introducing counterarguments, identifying overlooked literature, or helping researchers compare competing explanations. They may also reduce intellectual friction by delivering immediately acceptable answers that discourage further investigation. Whether assistance becomes substitution depends substantially upon how the instrument is used and upon the intellectual and institutional conditions surrounding its use. The availability of a rapid answer does not compel its acceptance, but an academic environment increasingly organized around speed, measurable output, and shortened deadlines may make independent examination less attractive and more difficult to sustain.

Empirical research provides reasons to take these concerns seriously, although the evidence requires careful interpretation. In *The Impact of Generative AI on Critical Thinking* (2025), Hao-Ping Lee and colleagues surveyed 319 knowledge workers and examined 936 reported examples of GenAI use in professional activities. They found that greater confidence in the generative system was associated with lower reported engagement in critical thinking, while greater confidence in one's own ability was associated with higher engagement. Importantly, the researchers also observed that generative tools can change the nature of critical-thinking activity, shifting attention from gathering information and solving problems toward checking generated material, integrating responses, and supervising the overall task.[32]

These findings do not establish that generative AI necessarily diminishes intelligence or that its users cease to think critically. The research concerns self-reported professional experience, and the observed relationships do not by themselves demonstrate long-term causal deterioration. Indeed, many respondents reported engaging in critical thinking when using such systems. The more significant implication is that intellectual work may be reorganized rather than simply eliminated.

---

[31] Sinopoli. EPISTEMÍA, pp.49-51.

[32] Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (Article 1121, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778

A person who previously developed an answer through investigation may increasingly spend time evaluating an answer prepared by a machine. Whether this reorganization strengthens or weakens professional competence depends upon the quality of that evaluation and the extent to which the individual continues to exercise independent judgment.

A different perspective emerges from the experimental study *Your Brain on ChatGPT* by Nataliya Kosmyna and colleagues (2025). The researchers examined essay-writing tasks under different conditions, including unaided writing, search-engine assistance, and the use of an LLM. Their investigation combined behavioural analysis with electroencephalographic measurements, involving 54 participants in the initial three sessions and 18 in a fourth session in which selected participants changed conditions. The reported differences in patterns of neural connectivity, recall, and engagement suggest that the cognitive experience of producing a text may vary substantially according to the instruments employed.[33]

The study is important, but its limitations are equally important. It is a research preprint based on a comparatively small experimental sample and a particular kind of writing task. Differences in neural activity during essay composition cannot legitimately be translated into claims that LLM users have suffered a general reduction in intelligence, nor can the findings establish the permanent cognitive consequences of sustained AI use across an entire generation. Their relevance to the present inquiry is narrower and more defensible: the method by which a written product is created may matter to the cognitive activity of its author, even when the final product is formally acceptable.

The problem also has a history considerably older than generative artificial intelligence. In *Ironies of Automation* (1983), Lisanne Bainbridge drew attention to the paradoxical position of human operators in increasingly automated systems. Automation may reduce the need for continuous human intervention while leaving people responsible for exceptional situations in which specialized knowledge and effective judgment are most urgently required. Yet the organization of automated work may itself make it harder to maintain the skills needed to respond to those exceptions.[34] Subsequent research by Parasuraman and Manzey (2010) demonstrated that automation-related complacency and bias can affect both inexperienced and expert users, particularly where competing demands for attention, imperfect decision aids, and habitual reliance on automation influence human performance.[35]

[33] Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task (arXiv:2506.08872, version 2). arXiv. Research preprint. https://doi.org/10.48550/arXiv.2506.08872

[34] Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

[35] Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3), 381–410. https://doi.org/10.1177/0018720810376055

Although these studies were not conducted in the contemporary academic environment of generative LLMs, the underlying human-factors problems are directly relevant. A professor may possess extensive knowledge of a discipline yet lack sufficient time to examine every assertion in a lengthy AI-assisted manuscript. An editor may understand the principles of scholarly verification but face a workload that encourages reliance upon superficial indicators of quality. A junior researcher may accept a generated argument because it resembles the academic literature they have been trained to imitate, without recognizing the assumptions that an experienced specialist would immediately question. In each case, the presence of a human participant provides no guarantee that meaningful intellectual evaluation has occurred.

The implications extend beyond the immediate accuracy of published research. Universities are responsible not only for evaluating existing knowledge but also for cultivating the capacities through which future generations will produce and examine it. If students increasingly complete demanding intellectual tasks through systems whose outputs they cannot independently reconstruct, the institution may continue to award degrees and recognize academic achievement without obtaining adequate evidence that the corresponding competencies have developed. This possibility does not imply that such substitution is universal or unavoidable. It identifies a risk that becomes particularly significant when educational assessment rewards the finished product while paying insufficient attention to the intellectual process through which it was obtained.

The same concern applies to established researchers. Expertise cannot be inferred merely from the ability to produce a well-structured document, just as the possession of a substantial bibliography does not demonstrate familiarity with the works it contains. A scholarly argument acquires credibility through the defensibility of its propositions, the quality of its evidence, the appropriateness of its methods, and the author's capacity to respond to criticism. Generative assistance can contribute to each of these activities when directed and evaluated by a competent researcher. It cannot, by the mere production of coherent language, establish that the person submitting the work has exercised the judgment that academic authorship ordinarily presupposes.

This distinction leads naturally to Sinopoli's subsequent proposal of *Cognitive Dignity* (2026),[36] which moves the discussion from the conditions of human understanding toward the possible normative protection of the person as an active participant in the formation of knowledge. His later analysis distinguishes legitimate assistance and voluntary delegation from situations in which technological or institutional arrangements materially displace the cognitive process while continuing to attribute its results to the human participant. The proposal does not establish that a new autonomous human right has already been recognized, nor does it require individuals to perform every intellectual operation without assistance. Its importance for the present study lies in identifying a problem that conventional measures of

[36] Sinopoli, A. (2026). Cognitive dignity: A proposal for an emerging human right: Protecting the human process of sense-making and knowledge formation against technological substitution. Canadian Review of Artificial Intelligence, Law & Society (CRAIL), 1(2), 89–117. https://www.crail.ca/CRAIL_Vol.1-No.2-2026-October.pdf

productivity and procedural compliance may fail to capture: the preservation of meaningful human participation in activities whose legitimacy depends upon human understanding.[37]

For academic institutions, this question is especially demanding. A university cannot reasonably prohibit intellectual tools without undermining the very tradition of scientific innovation it exists to advance. Yet neither can it assume that the increasing sophistication of those tools automatically strengthens the intellectual capacities of their users. The institutional challenge is to preserve opportunities for independent investigation, criticism, and the acquisition of expertise while taking advantage of the genuine benefits of computational assistance. That challenge requires more than instructions to disclose the use of artificial intelligence, important though such disclosure may be. It requires serious consideration of what universities teach, what they evaluate, and what kind of intellectual achievement their qualifications and publications are intended to certify.

The resulting issue reaches beyond cognitive psychology into the foundations of academic authority. If a machine can formulate an argument, a human can approve it without fully understanding it, and an institution can subsequently recognize the document as scholarship, the ordinary signs of academic accomplishment may remain intact while the intellectual process underlying them becomes increasingly difficult to identify. Before examining the institutional consequences of that possibility, however, it is necessary to consider a more fundamental question: how do language, interpretation, and authority acquire the power to transform an unsupported proposition into something that appears to deserve belief?

4.2. The Humpty Dumpty Academy: Meaning, Authority, and Proof

It is difficult to imagine that Lewis Carroll, when introducing Humpty Dumpty to his readers in 1871, anticipated that his extraordinary creation would eventually become a useful guide to some of the intellectual habits of modern academic institutions. Yet the famous conversation between Alice and the self-confident gentleman perched upon a wall has lost remarkably little of its relevance. Indeed, in an age when almost any proposition can be dressed in sophisticated terminology, furnished with references, and presented as an original theoretical contribution, Humpty Dumpty might feel entirely at home in certain university departments.

The episode is familiar, but its philosophical implications deserve attention. Humpty Dumpty informs Alice that he can make a word mean whatever he chooses. When Alice questions whether words can be assigned such arbitrary meanings, he dismisses the objection with a declaration that reveals the real issue: “The question is,” said Humpty Dumpty, “which is to be master—that’s all”.[38] What begins as an amusing disagreement about language becomes a dispute over intellectual authority. Alice assumes that words belong to a shared system of meaning, within which individual speakers may propose definitions but cannot simply abolish the conditions of mutual understanding. Humpty Dumpty proceeds from the

[37] Sinopoli. Cognitive dignity, pp. 102-111.
[38] Carroll, L. (1871). Through the Looking-Glass, and What Alice Found There. London: Macmillan (first published December 1871; title-page dated 1872). Chapter VI, “Humpty Dumpty.” Available at: https://www.gutenberg.org/files/12/12-h/12-h.htm.

opposite assumption. Meaning, for him, is ultimately a privilege of the person who claims the power to determine it.

There is, of course, nothing objectionable about introducing new terminology. Scientific progress frequently requires the refinement of existing concepts or the creation of entirely new ones. The history of mathematics, physics, medicine, and philosophy would be impossible to reconstruct without acknowledging the importance of conceptual innovation. A researcher who proposes a new definition is not necessarily engaging in intellectual arbitrariness. The decisive question is whether that definition is sufficiently precise, logically consistent, distinguishable from existing concepts, and capable of performing a meaningful explanatory function. A genuine theoretical contribution must do more than rename something already familiar. It must help us understand a problem that previous conceptual instruments could not adequately address.

The distinction becomes less comfortable when the language of innovation begins to substitute for innovation itself. A familiar proposition may acquire an imposing new designation, be placed within an elaborate classification, and emerge as the foundation of an allegedly unprecedented theory. Another author may introduce a concept so broad that it accommodates almost every conceivable example while remaining conveniently immune to criticism. The result can be rhetorically impressive without being intellectually productive. What appears to be a conceptual breakthrough may amount to little more than a redistribution of familiar ideas among newly invented categories.

Such practices are not necessarily fraudulent, and their existence long predates generative artificial intelligence. Academic history offers numerous examples of terminological fashions, speculative systems, and ambitious theories whose explanatory value proved less substantial than their authors initially supposed. Indeed, intellectual experimentation requires a measure of freedom, and even mistaken hypotheses may contribute to the development of knowledge. The problem arises when the ordinary distinction between proposing an interpretation and establishing its validity begins to disappear. At that point, the authority of the author may be invoked to protect the concept from precisely the scrutiny that would establish whether it deserves a place in scholarly discourse.

A remarkable historical illustration of this problem emerges from English jurisprudence rather than philosophy. In *Liversidge v Anderson*, decided by the House of Lords on 3 November 1941 and reported in [1942] AC 206, the judges considered the interpretation of Regulation 18B of the Defence (General) Regulations 1939. The regulation authorized the Home Secretary to order detention where he had reasonable cause to believe that a person fell within specified categories associated with threats to public safety or national defence. The dispute concerned the meaning of those apparently straightforward words and, consequently, the extent to which an executive decision could be subjected to judicial scrutiny.[39]

[39] Liversidge v Anderson [1941] UKHL 1; [1942] AC 206. House of Lords, judgment of 3 November 1941. Dissenting opinion of Lord Atkin, especially pp. 225–247, 244–245. Available at: https://uniset.ca/other/cs5/1942AC206.html.

The historical circumstances were exceptional. Britain was engaged in a war that threatened its existence, and the government possessed emergency powers whose justification could not be divorced from the dangers confronting the country. The majority of the House of Lords accepted an interpretation under which the relevant judgment about reasonable cause belonged substantially to the Home Secretary, provided he acted in good faith. Their reasoning reflected concern that judicial examination of sensitive security information and executive assessments would interfere with the exercise of emergency powers. Whatever criticism may subsequently be directed at that approach, it would be historically inaccurate to portray the majority simply as judges indifferent to national survival or incapable of understanding ordinary English.

Lord Atkin nevertheless dissented, and his opinion has acquired a significance extending far beyond the particular wartime dispute. He insisted that the statutory expression concerning reasonable cause could not simply be reduced to the minister's personal conviction that his reasons were sufficient. The existence of reasonable grounds, he argued, was a condition capable of objective examination. His distinction was elementary in grammatical form but profound in legal consequence. To possess reasonable grounds for a belief is not necessarily the same thing as believing that one possesses them. If the two propositions are treated as interchangeable, a condition intended to limit governmental authority risks becoming little more than the government's declaration that it has complied with the condition.

Atkin recognized that the issue was not merely linguistic. The interpretation adopted by the majority had consequences for the liberty of individuals detained without ordinary criminal proceedings. By making the executive's own assessment effectively decisive, the law risked losing an important external safeguard against arbitrary detention. His dissent therefore addressed the relationship between the meaning of legal language, the allocation of institutional authority, and the possibility of independent examination. It was in this context that he introduced his celebrated literary comparison, observing that the proposed interpretation reminded him of the peculiar authority claimed by Humpty Dumpty.[40]

The reference was devastating because it exposed the logical structure of the dispute without requiring a lengthy philosophical excursion. The executive, in Atkin's analysis, could not be permitted to transform a legally prescribed condition into an exclusively subjective judgment merely by asserting its preferred interpretation. His criticism did not establish that every legal expression possesses only one possible meaning, irrespective of context. Nor did it eliminate the ordinary difficulties of judicial interpretation. Its force lay in the insistence that interpretation must remain accountable to language, legal structure, and the purposes of the conditions imposed by law. Otherwise, an apparent limitation upon authority may be converted into an extension of that authority through the very process intended to define it.

There is an important analogy here with the production and assessment of academic knowledge. A researcher who announces that a proposition has been demonstrated must be able to identify the grounds upon which that claim rests. The fact that an argument appears convincing to its author does not establish that its premises are true, its sources authentic, or its conclusions justified.

[40] Liversidge v Anderson, pp.244-245.

Similarly, an editor who declares that a manuscript has undergone adequate scholarly verification cannot establish that fact merely by repeating the declaration. The relevant question concerns what was actually examined, by whom, according to which standards, and with what demonstrable result. The distinction resembles Atkin's opposition between possessing reasonable grounds and merely believing that such grounds exist, although the legal standards and consequences are, of course, different.

This analogy does not turn academic publishing into a system of constitutional adjudication. Courts exercise coercive authority under law, whereas journals operate through professional standards, editorial agreements, and scholarly conventions. The difference is fundamental. Nevertheless, both activities depend upon distinguishing a justified conclusion from the unsupported assertion of a person or institution entitled to pronounce it. Once that distinction weakens, the formal appearance of authority can begin to conceal the absence of the intellectual work from which legitimate authority should arise.

The philosophical foundations of this problem are considerably older than either Carroll's literary invention or Atkin's dissent. In his 1784 essay *Beantwortung der Frage: Was ist Aufklärung?* (*An Answer to the Question: What Is Enlightenment?*), Immanuel Kant described enlightenment through the individual's emergence from dependence upon the guidance of others in the exercise of understanding. His famous invocation of *Sapere aude* expressed the courage required to employ one's own reason rather than surrender its exercise to an external authority.[41]

Kant was addressing the intellectual and political conditions of the Enlightenment, not the operation of digital technologies, and his argument should not be retrospectively transformed into a theory of artificial intelligence. Nevertheless, the distinction he established between possessing the capacity for understanding and being willing to exercise it remains relevant. Intellectual dependence does not always arise because a person lacks the ability to think. It may also persist because established habits, convenient authorities, or institutional arrangements discourage the effort required for independent judgment. In the contemporary university, this problem acquires an additional dimension when external assistance is not provided by a tutor, a religious authority, or an established scholarly tradition, but by an automated system capable of presenting an apparently complete argument almost instantaneously.

One can readily imagine Kant's surprise at discovering that an instrument capable of supplying answers to an extraordinary range of questions might become an occasion for further intellectual dependence rather than intellectual emancipation. Yet such an outcome is neither inevitable nor inherent in the technology. A researcher may use generative AI precisely to challenge accepted interpretations, expose weaknesses in conventional arguments, or expand access to unfamiliar fields of knowledge. The danger lies in confusing the availability of explanations with the possession of understanding. An instrument may greatly enlarge the materials available for independent thought while simultaneously making it tempting to avoid the intellectual effort through which those materials become knowledge.

[41] Kant, I. (1784). Beantwortung der Frage: Was ist Aufklärung? Berlinische Monatsschrift, 4 (December), 481–494. Original German text available at: https://de.wikisource.org/wiki/Beantwortung_der_Frage:_Was_ist_Aufklärung%3F.

This is also where the distinction between intellectual authority and intellectual appearance becomes particularly important. Academic writing has developed a recognizable professional language. It values precision, conceptual consistency, engagement with relevant literature, and the explicit presentation of arguments. These conventions are indispensable when they help communicate complex research. They become problematic when their visible characteristics are mistaken for evidence that the underlying intellectual work has actually been performed.

A manuscript may contain an impressive introduction, elaborate theoretical distinctions, numerous references, and conclusions expressed with perfect confidence. None of these features establishes that the author understands the relevant literature, has examined the primary sources, or can defend the proposed argument against informed criticism. Conversely, a genuinely original investigation may initially be written in awkward prose, contain imperfect transitions, or require substantial editorial assistance before its significance becomes apparent. A serious scholarly evaluation must therefore distinguish literary presentation from intellectual substance without dismissing the importance of either.

Generative language models make this distinction more demanding because they can reproduce the recognizable linguistic conventions of academic discourse with remarkable fluency. They can assemble plausible theoretical frameworks, imitate the tone of scholarly disagreement, produce explanations that appear carefully qualified, and construct elaborate verbal distinctions. These capabilities are valuable when used as instruments under informed human direction. Yet linguistic sophistication does not provide an independent guarantee that the statements produced are well grounded. A coherent argument can be mistaken; a convincing citation can be inaccurate; an elegant classification can be conceptually empty. The difficulty arises when the reader, author, or reviewer begins to treat formal sophistication as a substitute for the verification of substance.

A particularly illuminating philosophical distinction was developed by Harry G. Frankfurt in his essay *On Bullshit*, originally published in 1986. Frankfurt examined a form of discourse whose defining characteristic is not necessarily deliberate falsehood, but a fundamental indifference to the truth of what is being communicated. A liar ordinarily operates with some conception of the truth, which the lie is designed to conceal or distort. The person engaged in what Frankfurt analyzes as bullshit need not maintain such a relationship with the facts. The immediate objective may instead be to create a particular impression, sustain a performance, or communicate an appearance of knowledge without the discipline required to establish its accuracy.[42]

Frankfurt's analysis requires careful application to artificial intelligence. His account concerns human communicative conduct and the intentions associated with it. A language model should not be described as a liar or a producer of bullshit in precisely the same philosophical sense, since the relevant intentional and epistemic conditions cannot simply be attributed to the machine. The analogy concerns the relationship between the resulting discourse and the standards by which claims are assessed. Machine-generated statements may

[42] Frankfurt, H. G. (1986). On Bullshit. Raritan, 6(2), 81–100. Reissued as: Frankfurt, H. G. (2005). On Bullshit. Princeton: Princeton University Press. Available at: https://philpapers.org/rec/FRAOBS.

be accurate, inaccurate, or insufficiently grounded without the system independently assuming responsibility for distinguishing among those possibilities. Whether the text subsequently becomes part of a responsible scholarly argument depends upon the human and institutional processes through which it is examined and adopted.

The more disturbing possibility, therefore, is not that artificial intelligence has invented a new form of intellectual dishonesty. It is that academic institutions may increasingly encounter polished discourse whose relationship to truth has not been adequately established, while the formal procedures of publication continue to confer upon it an appearance of scholarly legitimacy. This situation does not require deliberate deception by every participant. A manuscript may be submitted in good faith, reviewed superficially under severe time pressure, and accepted by an editor who assumes that the reviewer has performed a more comprehensive examination. Each person may believe that the necessary intellectual work has been completed somewhere else in the process. The final publication nevertheless emerges with the institutional marks of scholarly approval.

Here the distinction identified by Lord Atkin returns in a different setting. To assert that a proposition has been verified is not equivalent to having verified it. To describe an article as peer reviewed is not sufficient to demonstrate the competence, independence, or substantive adequacy of the review. Nor does the inclusion of an impressive bibliography establish that the cited works support the propositions for which they have been invoked. These differences may appear self-evident, yet modern academic administration often relies upon simplified indicators that can obscure precisely the distinctions upon which scientific credibility depends.

The sociology of science offers an important counterweight to the tendency to treat such indicators as ends in themselves. In *The Normative Structure of Science*, Robert K. Merton identified organized skepticism as a defining element of the scientific ethos. Scientific claims, on this account, must remain open to critical examination according to empirical and logical criteria, independently of the personal standing of those who advance them. Merton's analysis was normative rather than a claim that every scientific institution consistently behaved according to these principles. Its significance lies in describing the standards against which institutional scientific practice may properly be assessed.[43]

Organized skepticism is not a ceremonial expression of doubt. It requires the possibility of discovering that an influential scholar is mistaken, that a fashionable theory lacks evidence, or that an accepted explanation must be revised. Such criticism may be uncomfortable, particularly in environments where professional advancement depends upon reputation, institutional affiliation, and publication records. Yet the willingness to submit intellectual authority to independent examination remains one of the conditions that distinguish scientific inquiry from the production and circulation of authoritative opinion.

[43] Merton, R. K. (1973). The Normative Structure of Science. In: The Sociology of Science: Theoretical and Empirical Investigations. Chicago: University of Chicago Press, pp. 267–278. Originally published as "Science and Technology in a Democratic Order," Journal of Legal and Political Sociology, 1 (1942), 115–126. Available at: https://sciencepolicy.colorado.edu/students/envs_5110/merton_sociology_science.pdf.

This is precisely why the expansion of AI-assisted academic writing deserves serious attention. The problem cannot be reduced to the frequency of factual errors or the presence of fabricated references. Even a manuscript containing entirely genuine sources may be intellectually deficient if those sources have been misunderstood, selectively interpreted, or assembled into an argument that its author cannot independently defend. More sophisticated generative systems may reduce some forms of technical error while leaving this deeper problem unresolved. Indeed, improvements in fluency and factual reliability could make superficial assessments less capable of identifying the absence of genuine scholarly contribution.

Nor should we assume that extensive publication necessarily indicates either exceptional intellectual achievement or academic misconduct. Productive researchers may legitimately publish frequently, especially when working in collaborative disciplines or developing closely related research projects. Equally, a relatively small publication record does not guarantee originality or depth. The number of papers is an administrative fact; their scientific value requires independent evaluation. Confusing the two would reproduce the very error under examination, substituting a convenient numerical indicator for the more demanding task of scholarly judgment.

The Humpty Dumpty Academy is therefore not an institution in which every professor invents definitions or every manuscript is written by a machine. It is a metaphor for a more subtle institutional disorder: the gradual weakening of the distinction between intellectual authority and the evidence required to justify it. In such an environment, familiar concepts may acquire new names without acquiring greater explanatory power; elaborate arguments may be accepted without sufficient examination of their premises; and the formal procedures designed to certify scholarly quality may become detached from the intellectual activity they were intended to protect.

The image is amusing because its central character is so magnificently certain of his own authority. He does not need to investigate the meaning of a word or persuade Alice that his interpretation is defensible. He simply announces that the word means whatever he wishes. Yet the real danger appears when that attitude ceases to be a literary absurdity and becomes an institutional habit. A university that allows academic standing to substitute for argument, a journal that treats formal compliance as proof of scientific merit, or an author who mistakes the confident output of a generative model for an independently established conclusion has accepted, in different ways, the proposition that authority may determine what ought to be demonstrated.

There remains an essential difference, however, between the world of Carroll's satire and the institutions of contemporary scholarship. Alice was free to question Humpty Dumpty, and his pretensions did not survive the ordinary logic of their conversation. Academic institutions possess more consequential mechanisms of recognition. Publication, professional accreditation, editorial endorsement, and university affiliation can influence which arguments receive attention and which researchers acquire further opportunities. These mechanisms are not inherently defective; scientific communities require procedures for evaluating work and allocating limited resources.

Their legitimacy nevertheless depends upon the quality of the judgments through which recognition is conferred.

The central question is therefore not whether academic authority should exist, but whether its exercise remains connected to independent intellectual examination. This question becomes particularly pressing when automated systems can produce far more convincing scholarly material than existing procedures may be able to evaluate. The challenge is not simply to identify machine-generated language, which may be neither necessary nor sufficient for establishing scientific quality. It is to determine whether the person responsible for a scholarly claim has examined its grounds and whether the institution certifying that claim has exercised a form of judgment worthy of its authority.

At this point, the philosophical and legal discussion returns to the practical organization of academic research. The university cannot preserve its intellectual mission merely by insisting that a human being remains formally responsible for every published manuscript. Responsibility requires the capacity to understand what is being approved, the opportunity to question it, and the willingness to reject it when the evidence proves insufficient. Without these conditions, the forms of scholarly judgment may survive while their substance becomes increasingly uncertain. The consequences of that separation are most visible in the contemporary systems of manuscript production, peer review, and academic recognition, where the distinction between genuine verification and its procedural imitation is no longer a purely theoretical concern.

4.3. Scholarship Without Scholars? The Crisis of Academic Verification

The difficulties examined in the preceding sections become particularly consequential when they enter the institutions responsible for evaluating and transmitting scientific knowledge. An individual researcher may misunderstand a source, misinterpret a statistical result, or accept an argument that proves less convincing upon closer examination. Such mistakes have always accompanied intellectual inquiry. Science does not require its practitioners to be infallible; it requires their claims to remain open to criticism, correction, and independent verification. The more serious problem begins when the institutions established to perform these functions become increasingly unable to distinguish genuine research from its convincing imitation.

Academic publishing occupies a central position in this process. A scientific journal does more than distribute manuscripts. Through editorial selection, peer review, publication, and subsequent correction, it contributes to the construction of a scholarly record upon which other researchers may rely. Publication does not transform a hypothesis into established truth, nor does peer review guarantee that every argument or result has been independently reproduced. Nevertheless, these procedures create a reasonable expectation that the work has undergone some form of competent intellectual examination. When that expectation becomes unreliable, the consequences extend beyond the reputation of an individual author or journal. They affect the conditions under which scientific knowledge accumulates.

The vulnerability of this system did not begin with generative artificial intelligence. Long before ChatGPT became widely available, academic publishing was already confronting organized forms of manuscript production, manipulated peer review, fabricated research data, and commercial arrangements designed to secure publication rather than

advance knowledge. The 2022 joint report of the Committee on Publication Ethics (COPE) and the International Association of Scientific, Technical and Medical Publishers (STM) documented the activities of so-called *paper mills*: organizations offering manufactured manuscripts, authorship positions, and related publication services. The report described the pressures that encourage participation in such arrangements and the difficulties publishers face when attempting to identify fraudulent submissions.[44]

The chronology matters. The existence of a commercial market for questionable academic publications before the widespread adoption of generative LLMs demonstrates that the underlying incentives and institutional vulnerabilities cannot reasonably be attributed to those technologies alone. An academic system that rewards publication counts without adequately evaluating intellectual contribution may create opportunities for the production of low-value or fraudulent research regardless of the available writing tools. Generative AI can make certain parts of that production faster and cheaper, but it did not create the demand for scholarly credentials, the competition for academic recognition, or the temptation to substitute formal achievement for substantive investigation.

The persistence of the problem is illustrated by the international United2Act initiative, launched in 2023 to coordinate responses to paper mills. Its first-phase report, published in September 2025, described the development of guidance, investigative practices, training materials, and collaborative measures intended to strengthen research integrity across the scholarly publishing system. The initiative reflects an important institutional recognition: fraudulent or systematically unreliable publications cannot be addressed adequately by isolated editors acting independently of publishers, universities, research organizations, and other participants in scholarly communication.[45]

There is nevertheless a danger in treating these developments solely as a struggle between honest publishers and dishonest manuscript factories. The market for manufactured academic credentials does not exist in an institutional vacuum. It responds to expectations and rewards established within the academic system itself. When careers depend heavily upon publication records, when research performance is evaluated through numerical indicators, and when institutions seek visible evidence of productivity, the distinction between producing knowledge and producing evidence of having produced knowledge may become increasingly difficult to maintain. Such incentives do not make misconduct inevitable, nor do they render quantitative measures useless. They do, however, create conditions in which the appearance of scholarly accomplishment can acquire an exchange value of its own.

Generative artificial intelligence enters this environment with a capability that differs significantly from earlier instruments of academic production. A manuscript factory once

[44] COPE & STM. (2022). Paper Mills: Research Report from COPE & STM. Committee on Publication Ethics and International Association of Scientific, Technical and Medical Publishers. Version 1, June 2022, 15 pp. https://doi.org/10.24318/jtbG8IHL

[45] United2Act. (2025). United2Act Phase 1: Report and Recommendations. International Association of Scientific, Technical and Medical Publishers (STM). Published September 3, 2025. https://stm-assoc.org/united2act-phase-1-report-and-recommendations/

required a combination of human writers, templates, fabricated or recycled materials, and administrative coordination.

Many of these activities can now be assisted by systems capable of producing technically sophisticated prose, assembling references, summarizing literature, and reorganizing arguments within remarkably short periods. The resulting text may be entirely legitimate when supervised by a competent scholar, but the same technological capabilities can also be used to increase the volume and apparent sophistication of work whose substantive foundations remain weak.

Evidence of the transformation of scholarly writing is already available. In a study published in *Science Advances* in 2025, Dmitry Kobak and colleagues examined changes in vocabulary across more than fifteen million biomedical abstracts indexed in PubMed between 2010 and 2024. Their analysis identified an abrupt increase in words disproportionately associated with LLM-assisted writing, leading the researchers to estimate that at least 13.5 percent of the abstracts published in 2024 had been processed with LLM assistance. The estimated proportion varied considerably across fields, countries, and publication venues.[46]

These findings document a significant change in the production of scientific language, but they must not be confused with evidence that an equivalent proportion of scientific studies were fabricated, intellectually deficient, or written without meaningful human participation. The method identifies statistical features associated with machine-assisted textual production; it does not reconstruct the research process behind each publication. A scientist may employ an LLM to improve grammar or clarity while retaining complete responsibility for the underlying investigation. Another may use the same instrument to generate explanations or conclusions that have never been independently examined. The presence of machine assistance does not, by itself, establish which of these situations occurred.

This distinction is indispensable to any responsible discussion of academic integrity. The use of artificial intelligence in scientific writing is not inherently incompatible with intellectual independence. Indeed, prohibiting such assistance indiscriminately would risk confusing the form of scholarly expression with its substantive origins. Researchers who work in a language other than their native one may benefit substantially from competent linguistic assistance. Others may use computational systems to organize extensive material, identify possible counterarguments, or improve the accessibility of complex findings. These applications can support scholarship without transferring the responsibility for scientific judgment to the instrument.

The real question concerns which intellectual functions remain attributable to the human researcher. Did the author formulate and understand the central problem? Were the cited sources examined independently? Were the methods selected and evaluated by someone competent to assess their suitability? Can the author explain the reasoning connecting evidence to conclusions, respond to informed objections, and identify the limitations of the study?

[46] Kobak, D., González-Márquez, R., Horvát, E.-Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. Science Advances, 11(27), eadt3813. https://doi.org/10.1126/sciadv.adt3813

These questions are not answered by determining whether a particular paragraph was written with an LLM. Nor are they resolved by a general declaration that artificial intelligence was used during manuscript preparation.

The difficulty is recognized, at least in principle, by contemporary scholarly publishing policies. The Committee on Publication Ethics states that AI tools cannot qualify as authors because they cannot assume responsibility for scholarly work, manage conflicts of interest, or undertake the legal obligations associated with authorship. Human authors remain responsible for the submitted manuscript, including its machine-assisted components, and must disclose relevant uses of AI.[47] In August 2026, *Nature Computational Science* similarly emphasized transparency, human accountability, and the preservation of independent scholarly judgment throughout manuscript preparation and peer review.[48]

These policies represent an important normative foundation, but they do not by themselves establish that meaningful verification has occurred. A disclosure statement tells the reader that an instrument was used and, ideally, identifies its role in the research process. It cannot demonstrate that every source was examined, that the analytical framework was independently developed, or that the author possesses sufficient expertise to defend the conclusions. An author may disclose the use of AI with exemplary transparency while nevertheless submitting work of limited intellectual value. Conversely, a manuscript prepared without any generative assistance may contain equally serious methodological weaknesses. Disclosure contributes to accountability, but it is not a substitute for verification.

A closely related problem of non-delegable responsibility has been examined in the author's earlier legal research. *Procedural Liability in the Age of LLMs* addressed the duty of legal professionals to independently verify AI-generated authorities, factual assertions, and procedural submissions.[49] *The Tool Has a Manufacturer: Producer Liability for Legal AI* extended this analysis to the responsibilities of developers and commercial providers whose systems are incorporated into professional decision-making.[50] Although academic publishing and judicial proceedings operate under different legal and institutional regimes, both confront the same underlying problem: the delegation of technical or linguistic operations does not, by itself, transfer responsibility for the substantive validity of the resulting claims.

[47] COPE Council. (2024). Authorship and AI Tools. Committee on Publication Ethics. Position statement, version 1, August 20, 2024. https://doi.org/10.24318/cCVRZBms

[48] Responsible and Transparent Use of AI in Scientific Publishing. (2026). Nature Computational Science, 6(8), 803. Published August 20, 2026. https://doi.org/10.1038/s43588-026-01043-4

[49] Kildeev, A. (2026a). Procedural Liability in the Age of LLMs. Scientific Platform XXI Century, No. 7 (16) (July 2026), pp. 27–51. Published journal version. https://episs.ru/

[50] Kildeev, A. The Tool Has a Manufacturer: Producer Liability for Legal AI. Canadian Review of Artificial Intelligence, Law & Society (CRAIL), 1(2), 118–153. Published journal version. https://www.crail.ca/CRAIL_Vol.1-No.2-2026-October.pdf

Applied to scholarly publishing, this distinction requires attention to the respective duties of authors, reviewers, editors, and technological providers.

Legitimate scholarly collaboration must be distinguished from the abdication of judgment. No single author or reviewer is required to reproduce every specialist contribution personally; what matters is whether material claims remain subject to identifiable, competent and accountable evaluation at the appropriate level. Delegating performance does not, by itself, transfer the duty to assess and justify what an institution ultimately certifies. Related questions of verification and allocated responsibility have been examined in the context of AI-assisted legal practice.

The distinction becomes even more significant when the scholarly manuscript passes from its author to the reviewer. Peer review is frequently described as the principal safeguard of scientific publishing, but its effectiveness depends upon conditions that cannot be guaranteed merely by establishing a formal procedure. The reviewer must possess relevant expertise, have sufficient time to examine the work, retain independence from improper influences, and understand the limits of the evaluation being performed. Even under favourable circumstances, a reviewer cannot necessarily reproduce an entire experimental investigation, inspect every underlying dataset, or independently verify every historical and legal proposition. A responsible review therefore requires both substantive examination and a clear understanding of what remains unverified.[51]

The growing capacity for automated manuscript production may place additional pressure upon this already imperfect arrangement. Where the number and complexity of submitted papers increase faster than the availability of qualified reviewers, editorial offices face a difficult choice. They may extend review periods, reject more manuscripts before external evaluation, distribute work among additional specialists, or gradually reduce the depth of examination expected from each reviewer. None of these consequences follows automatically from the use of LLMs; journals differ substantially in resources, submission practices, and editorial standards. Nevertheless, the underlying asymmetry identified earlier in this chapter remains relevant. Producing an apparently complete argument can be considerably faster than independently establishing whether the argument is correct.

A survey of legal academics provides a useful illustration of the professional distinctions involved. In *AI Use in Legal Scholarship*, Wayne Unger reported preliminary findings from 268 responses concerning the acceptable uses of generative AI across different stages of scholarly writing. According to the author's publicly available research memorandum, 78 percent of respondents considered AI assistance with grammar acceptable, while only 11 percent considered AI drafting analytical sections containing novel arguments acceptable. The survey also reported strong support for developing norms governing scholarly AI use, alongside relatively limited reported adoption of formal institutional policies.[52]

[51] Kildeev A., Cited works.

[52] Unger, W. (2026). AI Use in Legal Scholarship. SSRN Working Paper. Preliminary research memorandum, abstract and appendices, 15 pp. Posted August 17, 2026; last revised August 25, 2026. https://doi.org/10.2139/ssrn.7290238

These findings are instructive, but their evidentiary status requires qualification. The publicly available SSRN document is a fifteen-page work-in-progress memorandum containing an abstract, questionnaire, and preliminary results rather than the completed research article. The available materials do not yet establish all the methodological details necessary to assess representativeness, including a sufficiently documented sampling frame, response rate, and subgroup analysis. The reported attitudes therefore describe the respondents rather than proving a consensus across the entire legal academy. Even with these limitations, the contrast between acceptance of linguistic assistance and concern about machine-generated original arguments is directly relevant to the distinction developed in the present study.

The problem is not that every scholarly argument must originate in complete intellectual isolation. Research has always involved consultation, collaboration, disagreement, and the development of ideas through engagement with other minds. Scientific authorship is rarely an exercise in solitary invention. What matters is the relationship between the contribution attributed to an author and the intellectual work that the author actually performed. A researcher may legitimately adopt an argument proposed by a colleague, provided it is examined, understood, appropriately acknowledged, and incorporated into a defensible scholarly position. The same standards of understanding and responsibility should apply when an argument originates from a computational instrument.

Yet the formal attribution of responsibility becomes increasingly fragile when the person expected to exercise judgment cannot reconstruct the reasoning upon which a conclusion depends. This is the problem identified in Alejandro Sinopoli's discussion of *Cognitive Dignity*: a person may remain formally present within a process while the cognitively significant functions attributed to that person have been materially displaced. The distinction between assistance, voluntary delegation, and substantive substitution becomes particularly important when institutions treat formal approval as evidence of independent understanding.[53]

An academic reviewer who receives an AI-generated manuscript may understand the relevant discipline but lack the time necessary to verify its numerous claims. Another reviewer may possess impressive professional credentials but insufficient competence in the particular subject under examination. A third may accept the text because its terminology, references, and conclusions resemble familiar forms of academic argument. These situations are not equivalent. The first concerns the practical limits of qualified verification under increasing workloads; the second concerns the inappropriate allocation of expertise; the third involves the danger of mistaking formal familiarity for substantive reliability. In each case, however, the existence of a nominal human approval stage may conceal the absence of meaningful evaluation.

At this point, the familiar expression *human-in-the-loop* deserves a moment of reconsideration. The phrase suggests that the presence of a human participant provides an essential safeguard against the limitations of automated systems. But the effectiveness of that safeguard depends upon the person's competence, independence, attention, and genuine opportunity to reject the machine's conclusions. Without these conditions, one might imagine

[53] Sinopoli. Cognitive Dignity, pp.102-108.

Homer Simpson sitting contentedly before a control panel, pressing the appropriate button while remaining cheerfully unaware of what the machinery is actually doing. The procedure has retained its human operator, but not necessarily the judgment for which that operator was supposed to be present. What was intended as ***human-in-the-loop*** has quietly become ***Homer-in-the-loop***.

The humour is familiar; the institutional problem is not. A reviewer who lacks the competence to assess a manuscript does not become qualified merely by occupying an editorial position. Equally, a highly competent reviewer cannot perform unlimited verification simply because an administrative system assigns additional manuscripts. The distinction matters because the corresponding remedies differ. Inadequate expertise requires better reviewer selection and clearer disciplinary boundaries. Excessive workloads require realistic allocation of time, resources, and expectations. Neither problem can be solved by attaching another automated approval stage to a process whose underlying deficiencies remain unchanged.

Nor would it be sensible to dismiss machine assistance in editorial work altogether. Automated reference checks, plagiarism screening, manuscript-format validation, and identification of potentially problematic images can provide useful preliminary safeguards. Computational tools may help reviewers locate relevant sources, compare versions of a document, or identify inconsistencies requiring further examination. Their contribution is valuable precisely because it allows human attention to be directed toward questions that require interpretation and judgment. What these instruments cannot establish merely through their operation is that the research has been independently understood and found intellectually defensible.

The distinction between technical screening and scholarly evaluation becomes clearer when one considers the difficulty of assessing original theoretical work. A citation can often be checked against an identifiable publication, but verifying its existence does not determine whether it supports the proposition for which it is cited. A mathematical formula can be syntactically correct without being relevant to the phenomenon under investigation. A philosophical concept may be expressed with great precision while contributing little beyond the restatement of established arguments. In such cases, evaluating the work requires knowledge of the field, intellectual independence, and a willingness to challenge the author's assumptions rather than merely inspect the formal properties of the document.

The evaluation of scholarly originality presents a further difficulty that no automated screening procedure can fully resolve. A theoretical argument may be elegantly constructed, extensively referenced, and expressed in the language of an established discipline while contributing little to its development. Conversely, an unfamiliar or unconventional proposition may challenge accepted assumptions and deserve serious consideration precisely because it does not conform to prevailing academic expectations. The responsibility of a competent reviewer is to distinguish these possibilities through careful examination of the argument, its sources, and its actual contribution to knowledge. Such evaluation requires intellectual independence as well as disciplinary expertise. Without them, peer review risks becoming an exercise in recognizing familiar academic forms rather than assessing the substance of scholarly work.

A related danger concerns the institutional treatment of preprints and working papers. Repositories perform a valuable function by allowing researchers to circulate preliminary results, establish public records of their work, and invite criticism before formal publication. Yet repository availability should not be confused with independent scholarly endorsement. A paper may be publicly accessible, assigned a persistent identifier, and indexed by search engines without having undergone substantive external peer review. Such dissemination is entirely legitimate, provided its status is clear. The problem arises when the visibility and formal presentation of a manuscript encourage readers or institutions to attribute to it a level of scientific validation that has not actually occurred.

The same distinction applies to the growing use of AI-generated research summaries, automated literature reviews, and academic discovery systems. These tools may help readers navigate the expanding scientific literature, but they can also reproduce the appearance of scholarly consensus without adequately distinguishing original research, preliminary hypotheses, disputed interpretations, and unverified claims. The resulting risk is cumulative. A speculative proposition may be summarized as a finding, cited by another manuscript as established background, and eventually appear in later literature stripped of the qualifications that originally accompanied it. Such processes existed long before generative AI, but automated production and redistribution can alter their speed and scale.

The possibility of recursively incorporating machine-generated material into subsequent computational systems introduces a further concern, although it must be distinguished from the direct deterioration of human scholarly competence. Shumailov and colleagues demonstrated that repeated training on recursively generated data can produce forms of model collapse under the conditions examined in their study. This result concerns the statistical properties of machine-learning systems and should not be presented as proof that scientific knowledge itself must inevitably collapse when scholars use generative AI.[54] Its relevance is more limited: the quality and provenance of information entering subsequent processes matter, whether those processes are computational or institutional. The analogy does not eliminate important differences between human scientific inquiry and machine learning.

At the level of scholarly publishing, the corresponding institutional question concerns the preservation of reliable distinctions within the scientific record. A journal must be able to distinguish original findings from commentary, verified evidence from inference, preliminary research from established results, and editorial acceptance from independent reproduction. These distinctions do not guarantee that scientific knowledge will progress without error, but they provide the conditions under which errors can be identified and corrected. When publication systems lose the ability or willingness to maintain them, the accumulation of documents may increasingly diverge from the accumulation of dependable knowledge.

It would nevertheless be a mistake to describe contemporary academia as a uniformly collapsing institution. Strong universities, demanding editorial boards, independent research groups, and exceptional scholars continue to make substantial contributions across numerous disciplines.

[54] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. Nature, 631, 755–759. https://doi.org/10.1038/s41586-024-07566-y

Developments in medicine, physics, mathematics, engineering, and other fields provide ample evidence that scientific inquiry remains vigorous and productive. The available evidence does not establish an irreversible or universal decline in academic competence. It does, however, justify concern about institutional practices that may weaken the conditions necessary for maintaining that competence over time.

The question is therefore not whether scientific knowledge has ceased to advance, but whether the institutions responsible for its reproduction can sustain the intellectual standards upon which future advances depend. A university may preserve excellent research laboratories while simultaneously encouraging superficial forms of academic assessment. A journal may publish outstanding articles while struggling to identify questionable submissions elsewhere in its portfolio. A researcher may possess genuine expertise in one field while accepting unreliable automated summaries in another. These possibilities are not mutually exclusive. Institutional strength and institutional vulnerability can coexist within the same academic system.

This coexistence makes the task of reform more demanding than either technological enthusiasm or generalized cultural pessimism would suggest. Simply prohibiting generative AI would not eliminate publication incentives, commercial manuscript factories, inadequate reviewer selection, or the misuse of academic credentials. Equally, unrestricted adoption accompanied only by statements about human responsibility would not ensure that the relevant judgment is actually exercised. Responsible integration requires an institutional environment in which authors remain answerable for their arguments, reviewers receive tasks appropriate to their competence, editors exercise independent authority, and scholarly claims remain open to examination after publication.

Such an environment also requires intellectual honesty about the limits of verification. No editor can guarantee the truth of every published proposition, and no reviewer should be expected to possess universal disciplinary competence. The purpose of scholarly evaluation is not to create an illusion of infallibility, but to establish credible, transparent, and corrigible procedures through which intellectual claims may be examined. A responsible journal should be able to explain the kind of review a manuscript received, identify the standards applied, manage conflicts of interest, and correct the scholarly record when significant errors emerge. These duties are not innovations demanded by artificial intelligence; they are established responsibilities whose importance becomes more visible as automated production expands.

There is a final irony in the contemporary debate. Generative AI has frequently been presented as a technology capable of democratizing access to intellectual production. In significant respects, this promise is real. Researchers without substantial institutional resources can obtain linguistic assistance, explore unfamiliar literature, and prepare work that might previously have been difficult to communicate effectively. Yet wider access to the means of producing scholarly text does not automatically provide the education, expertise, or critical judgment necessary to produce scholarship. Democratizing the tools of intellectual expression and cultivating the capacity for independent thought are related but distinct objectives. The former can be achieved through technological distribution; the latter requires sustained educational and institutional effort.

The historical achievement of the university was never merely its ability to preserve books, organize examinations, or confer academic titles. Its deeper contribution lay in cultivating communities capable of understanding, criticizing, extending, and transmitting knowledge. Scientific publications were among the visible results of that activity, but they were not its ultimate purpose. When the production of those results becomes increasingly independent of the intellectual capacities traditionally required to produce them, the university faces a question that cannot be resolved through publication metrics or improvements in automated writing alone.

The challenge extends beyond the accuracy of today's scholarly record. Universities reproduce themselves through the training of future researchers, the transmission of disciplinary standards, and the continuity of intellectual traditions. If successive generations become increasingly dependent upon systems whose conclusions they cannot independently examine, the consequences may eventually affect not only the quality of individual publications but also the capacity of academic institutions to preserve and renew their own intellectual foundations. This possibility requires empirical examination before its wider institutional consequences, including the weakening of intellectual continuity, can be assessed.

Chapter 5. Empirical Evidence and the Limits of Cognitive Delegation

5.1. Cognitive Offloading and the Question of Independent Judgment

The preceding analysis of the abdication of academic judgment raises a question that legal philosophy cannot resolve through conceptual argument alone. If scholarly institutions increasingly accept intellectual products generated with the assistance of large language models, is there any empirical reason to question whether the people responsible for those products will retain the knowledge and evaluative competence necessary to assess them independently? This chapter addresses that limited question. It does not diagnose cognitive decline among students or university teachers, assume that generative artificial intelligence causes a general deterioration of intelligence, or attempt an independent investigation in psychology or educational science. Its purpose is to determine whether relevant findings in those disciplines identify a sufficiently credible risk to warrant institutional attention.

An appropriate starting point is the established concept of cognitive offloading. Risko and Gilbert describe how people use external actions and resources to reduce the demands placed on internal cognitive processes.[55] Such practices long predate generative AI: writing, calculation, indexing, and electronic reminders all redistribute cognitive work. Offloading is neither inherently deficient nor inherently harmful. It may release attention for more demanding intellectual activity. Its consequences depend upon the task, the individual's existing competence, and the activity being delegated. This distinction matters because the mere use of an external instrument cannot itself establish the erosion of judgment.

[55] Risko, E. F. Cognitive Offloading / E. F. Risko, S. J. Gilbert // Trends in Cognitive Sciences. 2016. Vol. 20, no. 9. pp. 676–688. DOI: 10.1016/j.tics.2016.07.002.

Generative systems nevertheless introduce a particular institutional problem. They can supply not only facts or computational assistance but apparently complete explanations, arguments, evaluations, and written conclusions. Where the purpose of an educational or scholarly exercise is to develop the very operation being outsourced, the immediate quality of the resulting product may cease to be a reliable measure of the person's acquired competence. A student may submit an accurate answer without being able to reproduce its reasoning; a researcher may receive a plausible synthesis without independently testing its sources. These are possibilities, not findings about all users. They identify the question against which the empirical studies must be read: what happens when assistance improves the performance of a task but may reduce the intellectual work through which independent mastery is formed?

5.2. Experimental Evidence: Performance, Retention, and the Conditions of Assistance

The study designs, principal measurements, and inferential limits are summarized in Exhibit D. One of the clearest experimental illustrations comes from Bastani and colleagues, who studied generative-AI tutoring in mathematics with nearly one thousand secondary-school students.[56] They distinguished between a relatively unrestricted GPT-4-based interface and a tutor designed to support the student's reasoning rather than simply provide solutions. Access to both systems improved performance during assisted practice. The critical distinction emerged when access was withdrawn: students assigned to the less restricted interface performed approximately 17 per cent worse than those who had never received AI assistance on the subsequent unassisted assessment. The experiment therefore separates two outcomes often conflated in public discussion—successful completion of assisted work and learning that remains available without the tool. Its scope must, however, be stated precisely. This was a school mathematics experiment, not an investigation of university faculty or a demonstration of long-term cognitive decline. The result is relevant because it establishes that the design of assistance can change the relationship between immediate performance and subsequent independent performance.

Barcaui examined a related question in higher education through a randomized study of undergraduate learning and later knowledge retention.[57] Initially, 120 students were assigned to ChatGPT-assisted or traditional study conditions. At a surprise assessment 45 days later, the mean score among the 85 participants who completed follow-up was 57.5 per cent in the assisted condition and 68.5 per cent in the comparison condition. The observed difference was statistically significant. Unlike an immediate test of productivity, this outcome concerns retention after the initial learning activity. Yet the study does not justify a claim about durable cognitive impairment. Its particular instructional setting limits generalization, and the loss of 35 participants—29.2 per cent of the original sample—creates a substantial attrition concern, notwithstanding the reported balance of follow-up completion across conditions. The result

[56] Bastani, H. Generative AI without guardrails can harm learning: Evidence from high school mathematics / H. Bastani, O. Bastani, A. Sungu [et al.] // Proceedings of the National Academy of Sciences. 2025. Vol. 122, no. 26. Article e2422633122. DOI: 10.1073/pnas.2422633122.

[57] Barcaui, A. ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention / A. Barcaui // Social Sciences & Humanities Open. 2025. Vol. 12. Article 102287. DOI: 10.1016/j.ssaho.2025.102287.

is best treated as evidence of a possible learning cost under specified circumstances, not as a general law of AI-supported study.

Any assessment confined to adverse findings would itself fail the standard of independent judgment that this article defends. Makransky and colleagues reported two controlled experiments involving 175 university students and 234 secondary-school students.[58] They evaluated a theory-informed conversational tutor that encouraged learners to explain material and actively construct understanding, comparing it with a generic AI interface and relevant control conditions. The structured tutor yielded stronger conceptual-knowledge outcomes than generic ChatGPT in both immediate assessments and in a delayed assessment in one of the studies; comparisons with the non-AI controls were more variable. These findings neither prove that generative AI always improves learning nor erase the adverse results reported elsewhere. They demonstrate something more analytically useful: the educational consequences of the same broad technological category cannot be assessed without examining whether its design supports or substitutes the learner's own cognitive activity.

The issue is not limited to retention. Yan's randomized, scenario-based experiment enrolled 240 preservice teachers at a Chinese university; 218 were retained in the principal cleaned analysis.[59] Participants faced pedagogical cases containing both accurate and misleading generative-AI recommendations. Under misleading advice, those who recorded an independent judgment before seeing the recommendation achieved higher expert-keyed accuracy than those who saw the recommendation before recording their first judgment (53.5 per cent versus 37.3 per cent on the relevant items). The controlled result concerns susceptibility to erroneous recommendations rather than any demonstrated decline of intelligence. Moreover, the participants were teachers in training, the tasks were simulated, and transfer to genuine classroom decisions remains unproven. The experiment nevertheless identifies a practically significant variable: the sequence in which a person first reasons and then consults an algorithmic recommendation may affect the integrity of the final judgment.

Taken together, these experiments support a cautious but important proposition. Generative-AI assistance is not a unitary educational intervention. It may facilitate understanding when designed and used to sustain active explanation, while forms of assistance that replace reasoning may produce gains in visible performance without comparable gains in independent competence. The empirical claim is conditional. It does not follow that every use of an LLM weakens cognition, still less that an entire generation of students or scholars is undergoing cognitive deterioration.

[58] Makransky, G. Beyond the "Wow" Factor: Using Generative AI for Increasing Generative Sense-Making / G. Makransky, B. M. Shiwalia, T. Herlau, S. Blurton // Educational Psychology Review. 2025. Vol. 37. Article 60. DOI: 10.1007/s10648-025-10039-x.

[59] Yan, D. Independent judgment before viewing generative AI recommendations reduces susceptibility to misleading advice: A randomized scenario-based experiment with preservice teachers / D. Yan // Frontiers in Psychology. 2026. Vol. 17. Article 1936522. DOI: 10.3389/fpsyg.2026.1936522.

5.3. From Limited Findings to Academic Responsibility

Evidence concerning university teachers must be handled with particular restraint. Qiang and Wang analyzed multi-source accounts from 25 university teachers recruited through an AI-focused professional learning community and associated networks in China.[60] Their qualitative analysis addresses teachers' reported experiences of cognitive delegation, professional judgment, and strategies for checking generated material. Such accounts are valuable for identifying how practitioners describe the shifting boundaries of their work, but they are not measurements of diminished intellectual ability. The sample is small and selectively recruited; the materials are uneven across cases; and self-reports cannot establish the effects of AI use on actual classroom performance. It would therefore be unjustified to convert this study into evidence that university teaching as a profession has cognitively deteriorated. Its narrower significance lies in documenting that the preservation of independent assessment has become an explicit concern within teachers' accounts of their own practice.

A broader methodological perspective is supplied by the systematic review and meta-analysis of Boolzen and colleagues, published in 2026.[61] The authors identified 85 eligible studies of generative AI and cognitive learning outcomes in STEM education and included 49 studies, representing 59 effect sizes, in the quantitative synthesis. Although a conventional random-effects model indicated an overall positive effect, the findings displayed extreme heterogeneity. An analysis designed to account for possible publication bias substantially reduced the estimated average benefit. The authors also reported that none of the included randomized or non-randomized intervention studies was assessed as having a low overall risk of bias. These observations preclude simple declarations either that AI reliably damages learning or that it reliably improves it. Of particular relevance is the review's distinction between AI used to augment learners' cognitive activity and AI used to substitute for it: this distinction helps account for variation in outcomes, while leaving considerable uncertainty unresolved.

Several limits of inference follow. First, a short-term difference in test performance or knowledge retention is not equivalent to a persistent change in general cognitive capacity. Secondly, experimental findings involving school pupils, undergraduates, and preservice teachers cannot be transferred without qualification to experienced scholars and university professors. Thirdly, heterogeneous interventions—ordinary chat interfaces, specially designed tutors, different tasks, and different degrees of human verification—cannot be treated as one uniform exposure. Finally, the existing literature does not establish that changes attributed to AI are irreversible, widespread across academic systems, or caused by the technology independently of prior weaknesses in teaching, assessment, research incentives,

[60] Qiang, Z. Between Human and Machine: University Teachers' Accounts of Judgment, Metacognitive Defense, and Cognitive Offloading in Generative AI Use / Z. Qiang, Z. Wang // Journal of Intelligence. 2026. Vol. 14, no. 10. Article 254. DOI: 10.3390/jintelligence14100254.

[61] Boolzen, C. Evidence of impact and interpretational limits of generative AI in STEM education: A systematic review and meta-analysis on cognitive learning outcomes / C. Boolzen, J. Kuhn, S. Flegr [et al.] // Artificial Intelligence Review. 2026. Vol. 59. Article 223. DOI: 10.1007/s10462-026-11665-9.

or institutional governance. The cited findings support a reason for caution; they do not provide a scientific warrant for a diagnosis of academia as cognitively degraded.

That conclusion does not deprive the evidence of institutional significance. Academic institutions routinely assign responsibility to identifiable persons: the student must demonstrate what has been learned, the author must substantiate claims, the reviewer must examine the submission, and the editor must make an accountable decision. These duties presuppose more than a human name attached to a finished document. They presuppose an effective capacity to inspect arguments, recognize uncertainty, evaluate sources, and refuse an unsupported conclusion. The studies considered here do not show that this capacity has generally disappeared. They do, however, reveal circumstances in which successful use of a cognitive instrument may be difficult to distinguish from the independent competence that academic procedures are intended to cultivate and verify.

The resulting issue is juridical and institutional rather than clinical. A journal may accept AI-assisted drafting while still requiring authors to verify every material assertion and citation. A university may permit AI-assisted learning while preserving forms of assessment that test knowledge without the tool. Reviewers and editors may use automated assistance without treating the resulting text as an independent act of scholarly judgment. None of these safeguards demands hostility to technology. Their common purpose is to ensure that the allocation of responsibility remains connected to a demonstrable human capacity to exercise it.

In the terminology developed in this article, the empirical literature does not establish *abdicatio iudicii* as an accomplished, universal condition. It identifies conditions under which the risk of such abdication becomes intelligible. The doctrinal concern begins when institutional procedures continue to attribute judgment to a human actor while allowing the actual formation and verification of that judgment to be displaced by generated outputs. Whether, how often, and with what consequences this occurs must remain open to further empirical investigation. The legal and philosophical obligation is already clear: apparent fluency or assisted productivity must not be mistaken for independent verification, and procedural responsibility must not be reduced to the ceremonial presence of a human signatory. This distinction provides the bridge to the following examination of institutional memory and the economics of intellectual discontinuity.

Chapter 6. Institutional Amnesia and the Economics of Intellectual Discontinuity

Universities are commonly understood as institutions devoted to the preservation and transmission of knowledge. Their libraries, laboratories, archives, and increasingly sophisticated digital repositories give this mission an impressive material form. A modern university can preserve millions of publications, provide access to scientific databases, and maintain an institutional record extending across generations. Yet the continuity of scientific knowledge depends upon something more demanding than the survival of documents. A scientific tradition persists only when successive generations acquire the intellectual capacities necessary to understand, question, reproduce, and extend the work of their predecessors.

This distinction becomes particularly important when considering the possibility of institutional amnesia. The expression does not necessarily imply that universities have forgotten their histories or abandoned their libraries. Scientific knowledge can remain extensively documented while the practical competence required to interpret and develop it becomes increasingly difficult to reproduce. An archive may preserve the results of a research programme without preserving every methodological judgment, experimental adjustment, unsuccessful hypothesis, or intellectual disagreement through which those results were obtained. The published record remains available, but its effective continuation depends upon the education and experience of those who inherit it.

The transmission of scientific knowledge has therefore always involved relationships between experienced researchers and those entering their disciplines. Laboratory practices, professional judgment, methodological caution, and the ability to recognize significant problems are developed through sustained participation in research communities. Formal instruction provides indispensable foundations, but it cannot fully replace the experience of working alongside scholars who possess the accumulated knowledge of a field. The importance of these relationships is supported by the National Academies of Sciences, Engineering, and Medicine in *The Science of Effective Mentorship in STEMM* (2019), which identifies mentorship as a significant element of scientific training and emphasizes the responsibility of institutions to support its quality and effectiveness.[62]

The same institutional concern appeared in the National Academies' earlier report *Graduate STEM Education for the 21st Century* (2018). That study acknowledged the substantial achievements of American graduate education while calling for improvements in professional preparation, mentoring, transparency, and the development of competencies suitable for different research careers.[63] These recommendations are particularly instructive because they were formulated before the widespread adoption of generative language models. The challenge of reproducing scientific expertise was already recognized as an institutional problem, independent of the technologies now transforming academic writing.

The conditions under which this transmission occurs have nevertheless been changing. Data published by the American Association of University Professors show that approximately 47 percent of faculty members in American higher education held contingent appointments outside the tenure system in 1987, compared with approximately 68 percent in 2023. Over the same period, the proportion holding full-time tenured or tenure-track appointments declined from approximately 53 percent to 32 percent. These figures describe substantial changes in the organization of academic employment, although the proportions differ considerably among institutional categories and should not be interpreted as a uniform transformation of every research university.[64]

---

[62] National Academies of Sciences, Engineering, and Medicine. (2019). The Science of Effective Mentorship in STEMM. Washington, DC: The National Academies Press. https://doi.org/10.17226/25568

[63] National Academies of Sciences, Engineering, and Medicine. (2018). Graduate STEM Education for the 21st Century. Washington, DC: The National Academies Press. https://doi.org/10.17226/25038

[64] American Association of University Professors. (2025). Data Snapshot: Tenure and Contingency in US Higher Education, Fall 2023. Academe. https://www.aaup.org/academe/issues/spring-2025/data-snapshot-tenure-and-contingency-us-higher-education-fall-2023, see also Exhibit B.

The significance of these developments requires careful interpretation. A temporary or part-time appointment does not establish intellectual inferiority, insufficient professional commitment, or poor teaching. Many accomplished scholars work outside traditional tenure arrangements, while permanent academic appointments provide no automatic guarantee of scientific originality. Nor do aggregate employment figures directly measure the quality of research mentorship. Nevertheless, employment arrangements influence the organizational conditions within which scholars teach, supervise students, conduct long-term investigations, and participate in departmental life. Where continuity of employment is uncertain or professional responsibilities are fragmented among several institutions, maintaining sustained research relationships may become more difficult. Whether these risks materialize depends upon institutional practices and cannot be inferred solely from appointment categories.

Changes in the destinations of doctoral graduates provide another perspective on the reproduction of the academic workforce. According to the U.S. National Center for Science and Engineering Statistics, the proportion of doctorate recipients with definite non-postdoctoral employment commitments in the United States who expected to work in academia declined from 56 percent in 2004 to 40 percent in 2024. In the latter year, industry and business accounted for a comparable proportion of those employment commitments.[65] These percentages concern a specifically defined group of graduates reporting definite employment commitments; they are not percentages of all doctorate recipients and should not be treated as measures of permanent career outcomes.

The movement of highly educated researchers into industry is not necessarily evidence of institutional weakness. Research-intensive companies, technological enterprises, medical organizations, and other employers offer important opportunities for scientific work. The ability of universities to prepare researchers for such careers may itself be regarded as a substantial educational achievement. The institutional question is different: how does the university maintain the professional capacity required to train future researchers when the distribution of scientific employment changes? A successful research system may benefit from extensive movement between universities and industry, but the continuity of academic research communities still requires sufficient opportunities for experienced scholars to supervise new generations.

England introduces a third institutional configuration, one which should not be treated as interchangeable with either the United States or Canada. Its universities operate within a distinctive arrangement of tuition fees, student finance, public funding bodies and research grants. England-specific statistics also have to be distinguished from figures reported for the United Kingdom as a whole. This matters particularly for employment and participation data, where superficially similar categories may refer to different populations or appointment regimes. The relevant comparison is consequently institutional rather than a forced equivalence between American tenure, Canadian academic contracts and British employment classifications.

---

[65] National Center for Science and Engineering Statistics. (2026). Doctorate Recipients from U.S. Universities: 2024. NSF 26-315. https://ncses.nsf.gov/pubs/nsf26315/report/postgraduation-trends, see also Exhibit B.

These developments acquire particular significance when considered alongside the financial expansion of research activity. The American National Center for Science and Engineering Statistics reported that higher-education research and development expenditures reached $117.7 billion in fiscal year 2024, representing an 8.1 percent nominal increase over the preceding year. Between 2014 and 2024, university research expenditure grew at an average annual rate of approximately 3 percent after adjustment for inflation.[66] This is substantial evidence of the continuing scale and financial strength of American academic research. It would be misleading to portray the entire university research sector as an institution undergoing straightforward contraction.

Canadian statistics likewise demonstrate considerable research activity. Statistics Canada reported higher-education research and development expenditure of C$19.0 billion in 2023/24, an annual increase of 7.2 percent.[67] Although the American and Canadian measures differ in currency, institutional coverage, and reporting methodology, they independently establish that significant financial resources continue to support university-related research in both countries. These expenditures cannot, however, be converted into direct measurements of scholarly originality, methodological competence, or intellectual independence.

The English financial picture requires a different indicator. According to the House of Commons Library's September 2026 briefing, higher-education providers in England received £46.1 billion in total income in 2024/25. This represents institutional income from multiple sources, not research-and-development expenditure, and therefore must not be placed in a three-country ranking beside the American and Canadian R&D figures. The same parliamentary analysis traces a pronounced alteration in England's funding structure: tuition fees accounted for 24 percent of provider income in 2005/06 but exceeded one half from 2019/20, while funding-body grants declined from 39 percent in 2005/06 to 11 percent in 2024/25. These observations describe a transformation in the financing of English higher education, without by themselves demonstrating changes in the quality of scholarship.

The distinction between expenditure and institutional capacity deserves closer examination. Research budgets finance equipment, laboratories, technical personnel, scientific projects, and numerous activities essential to discovery. An increase in expenditure may support valuable new investigations or compensate for rising costs without producing a proportionate change in the number of researchers, the quality of training, or the continuity of particular research programmes. Conversely, an institution may maintain highly productive research groups despite financial constraints. Aggregate expenditure therefore provides necessary economic context, but it cannot independently establish whether the conditions for reproducing scientific expertise are improving.

[66] National Center for Science and Engineering Statistics. (2026). Universities Report 8.1% Growth in R&D Expenditures in FY 2024, Reaching Over $117 Billion. NSF 26-305. https://ncses.nsf.gov/pubs/nsf26305, see also Exhibit B.

[67] Statistics Canada. (2025). Spending on Research and Development in the Higher Education Sector, 2023/2024. The Daily, 17 October. https://www150.statcan.gc.ca/n1/daily-quotidien/251017/dq251017e-eng.htm, see also Exhibit B.

The Canadian university financing system illustrates the complexity of this relationship. According to Statistics Canada, total university revenues reached C$52.4 billion in 2023/24, expressed in constant 2024 dollars. Provincial funding represented 32.4 percent of revenues, while tuition accounted for 30.1 percent. Over the preceding decade, the tuition share had increased by approximately 5.4 percentage points.[68] These figures reveal changes in financial composition, not a direct deterioration of educational quality. Universities differ in their dependence upon particular revenue sources, institutional mandates, and allocation of funds. Moreover, apparent financial surpluses may include resources restricted to purposes other than ordinary academic operations.

English institutions exhibit a further tension between substantial aggregate revenues and uneven financial resilience. The House of Commons Library records that 48 publicly funded universities in England, representing 40 percent of that defined group, reported expenditure exceeding income in 2024/25.[69] In a separate regulatory population, the Office for Students reported deficits at 35.8 percent of higher-education providers in 2024/25, while estimating that 42.7 percent could report deficits in 2025/26 without mitigating action. The OfS also recorded £218.2 million in restructuring costs in 2024/25.[70] The percentages must not be blended: the parliamentary university subgroup and the OfS provider population have different denominators, and the 2025/26 figure is a conditional forecast rather than an observed result. Moreover, the regulator noted an improvement in aggregate financial performance for 2024/25. The simultaneous presence of improvement and significant institutional pressure is central to a balanced interpretation.

England is also instructive because the financial structure distinguishes direct institutional support for research from revenue derived from student fees. The parliamentary briefing reports that direct funding through Research England, which supports research capacity and infrastructure rather than only specific projects, broadly maintained its real value from 2010 to 2024 before falling in real terms during the following two years.[71] This does not establish an inevitable decline in scientific capacity. It does, however, bring the economics of continuity into focus: research infrastructure and established scholarly communities may be maintained by funding streams whose purposes and time horizons differ from those of student-facing operations.

Financial composition matters because universities must make decisions about competing institutional priorities. Resources devoted to facilities, technology, administrative systems, immediate educational demand, and research infrastructure may all be necessary. The difficulty is that their economic benefits become visible over different periods. A completed building, a new laboratory instrument, or a digital platform can be identified in an

[68] Statistics Canada. (2025). University Revenues Increase More Than Expenditures in 2023/2024. The Daily, 29 July. https://www150.statcan.gc.ca/n1/daily-quotidien/250729/dq250729b-eng.htm, see also Exhibit B.

[69] Bolton, P., and Lewis, J. (2026). Higher Education Finances and Funding in England. House of Commons Library, Research Briefing CBP-10037, 4 September. https://commonslibrary.parliament.uk/research-briefings/cbp-10037/

[70] Office for Students. (2026). Financial Sustainability of Higher Education Providers in England: 2026. 14 May 2026, updated report. https://www.officeforstudents.org.uk/publications/financial-sustainability-of-higher-education-providers-in-england-2026/, see also Exhibit B.

[71] Bolton, P., and Lewis, J., Cited article.

annual budget. The development of a researcher capable of independently formulating significant scientific problems may require many years, and its eventual benefits may emerge outside the institution that financed the original training.

This difference in time horizons creates a potentially important economic problem. The immediate costs of scientific education are concentrated and measurable, while some of its most valuable returns are uncertain, delayed, and distributed across institutions, industries, and generations. A university that supports demanding research training may benefit society through discoveries made decades later by graduates working elsewhere. Such benefits are difficult to capture through annual institutional performance indicators. The resulting incentive problem does not establish that universities systematically neglect fundamental research or mentorship, but it helps explain why these activities require deliberate institutional protection.

Canadian federal advisory research has addressed related concerns. In its 2023 *Report of the Advisory Panel on the Federal Research Support System*, an independent expert panel emphasized the importance of investigator-initiated research, sustainable support for graduate students and postdoctoral researchers, and the long-term development of scientific talent. The panel warned that inadequate support for research trainees and insufficient investment in fundamental research could weaken Canada's capacity to attract and retain researchers.[72] These observations represent a documented institutional assessment at the time of the report, rather than proof that all identified problems remained unchanged in 2026.

The importance of investigator-initiated research extends beyond the allocation of grants. Scientific progress cannot always be organized around problems whose practical significance has already been established. Many discoveries emerge from investigations that initially appear remote from immediate economic or administrative priorities. Yet a balanced research system must also accommodate public priorities, technological development, and practical applications. The relevant economic question concerns the institutional conditions under which short-term objectives and long-term scientific inquiry can coexist without one progressively displacing the other.

It is equally necessary to distinguish concerns about academic continuity from broader claims of educational decline. The available evidence is heterogeneous. In the OECD's 2023 Survey of Adult Skills, Canadian adults achieved average scores of 271 in literacy and numeracy, both above the OECD averages. Compared with the preceding survey cycle, literacy performance remained broadly stable while numeracy improved.[73] These findings do not measure original scientific judgment or university research competence, but they caution against narratives portraying an uninterrupted and universal deterioration in intellectual capacities.

[72] Advisory Panel on the Federal Research Support System. (2023). Report of the Advisory Panel on the Federal Research Support System. Innovation, Science and Economic Development Canada. https://ised-isde.canada.ca/site/panel-federal-research-support/en/report-advisory-panel-federal-research-support-system

[73] OECD. (2024). Survey of Adult Skills 2023: Canada, Country Note. https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/canada_5ecab9d9-en.html, see also Exhibit A.

The English results reinforce the need for such caution. In the same OECD adult-skills programme, adults aged 16–65 in England obtained mean scores of 272 in literacy, 268 in numeracy and 259 in adaptive problem solving. The OECD reported broadly stable literacy and improved numeracy relative to the 2011/12 survey cycle.[74] School-based PISA observations address a different age group and cannot be combined numerically with adult-skills results; England-specific PISA 2022 reporting also contains explicit sampling qualifications.[75] None of these assessments measures independent academic scholarship directly. Their value is methodological: improvements and stability must remain in view alongside financial and institutional vulnerabilities.

What emerges from the three cases is a set of institutional tensions rather than a single statistical trajectory. Universities in the United States, Canada and England continue to support substantial scientific investigation, educate researchers, and preserve extensive intellectual resources. At the same time, differing employment structures, doctoral career opportunities, public funding arrangements and financial pressures raise questions about the sustainability of the conditions under which scholarly expertise is reproduced. The available indicators do not establish a uniform degree of institutional decline, nor do they justify assuming a single causal mechanism across three distinct university systems.

Generative artificial intelligence adds a new dimension to these existing challenges. Digital systems can make scientific literature more accessible, assist the preservation of technical documentation, facilitate collaboration, and reduce the time required for routine research operations. Properly integrated into education, they may strengthen the ability of researchers to engage with difficult problems. Yet the preservation of information should not be confused with the preservation of competence. A university may acquire increasingly powerful systems for retrieving and generating scientific explanations without ensuring that successive generations retain the ability to examine their assumptions, reproduce their methods, or identify their limitations.

The economic attraction of such systems is understandable. Where institutions face pressure to increase measurable output while controlling expenditure, automation offers opportunities to accelerate routine processes and reduce particular categories of work. These benefits deserve serious evaluation. Their long-term value, however, depends upon whether the resulting efficiencies strengthen intellectual training or merely increase production. If automation reduces the time required for scholarly administration and allows researchers to devote greater attention to difficult questions, its contribution may be substantial. If it instead becomes an excuse to weaken the institutional conditions through which scientific expertise is developed, apparent efficiency may conceal costs that become visible only later.

There is no reliable basis for calculating a universal economic cost of such intellectual discontinuity, still less for predicting the inevitable disappearance of scientific schools. The empirical material considered here does not establish that generative AI has already caused a

[74] OECD. (2024). Survey of Adult Skills 2023: England (United Kingdom), Country Note. https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/united-kingdom_02bc78e4-en.html, see also Exhibit A.

[75] Department for Education. (2023). PISA 2022: National Report for England. Research Report RR1392. https://www.gov.uk/government/publications/pisa-2022-national-report-for-england

measurable decline in research competence in any of the three cases. Much of the institutional transformation discussed in this chapter began decades before contemporary LLMs entered academic practice. The technology must therefore be understood within existing economic and organizational environments rather than treated as their original cause.

The deeper question concerns the relationship between the preservation of intellectual products and the preservation of the human capacities required to create them. A scientific institution may accumulate publications without adequately supporting the formation of new investigators; it may acquire advanced instruments while neglecting the skills necessary to evaluate their results; and it may celebrate research productivity without sufficiently examining the conditions under which future discoveries will become possible. These outcomes are neither universal nor inevitable. They are institutional risks whose seriousness depends upon decisions made by universities, public authorities, funding bodies, and scholarly communities.

Institutional amnesia, in this sense, begins when an organization loses sight of the processes that allow knowledge to remain a living intellectual practice. Its consequences need not be immediate. Scientific institutions may continue to function successfully for considerable periods while depending upon competencies developed through earlier arrangements. The task of responsible academic governance is to ensure that the conditions required to renew those competencies remain available to future generations.

The evidence from the United States, Canada and England does not justify announcing the collapse of the university. It does justify examining how different research systems allocate resources, sustain scientific careers, support mentorship, and distinguish intellectual achievement from its measurable appearances. The relevant uncertainty concerns the future, but institutional choices must be made in the present. The concluding chapter therefore turns from the diagnosis of possible intellectual discontinuity to the principles through which universities may preserve the independence of human judgment while continuing to benefit from technological progress.

Chapter 7. Preserving the Human Academy

The central finding of this study is neither the imminent collapse of the university nor the inevitable degradation of human intelligence through artificial intelligence. The evidence examined across the United States, Canada, and England supports neither conclusion. Universities continue to generate significant scientific discoveries, sustain extensive research programmes, and educate new generations of scholars. Their financial structures, employment practices, and educational outcomes reveal substantial differences and, in some instances, contradictory tendencies. What emerges from this investigation is a more specific institutional problem: the growing possibility that the visible production of academic knowledge may become detached from the human capacities required to understand, verify, and reproduce it.

This distinction provides the conceptual basis for the present inquiry. Academic knowledge has always involved an interaction between intellectual activity and its material expression. A scientific publication records an investigation; a degree recognizes an educational achievement; a research institution preserves and develops disciplinary expertise.

These relationships have never been perfect. An incompetent scholar could publish before the invention of the computer, just as a distinguished university could confer recognition upon work whose significance later proved negligible. Generative artificial intelligence did not create these weaknesses. It has, however, introduced an instrument capable of producing many of the external characteristics of scholarly accomplishment without requiring a corresponding intellectual achievement from the person who employs it.

The resulting danger extends beyond the familiar problem of inaccurate machine-generated information. A fabricated citation can be identified, a statistical error corrected, and a defective argument rejected. More difficult to recognize is the situation in which the document appears satisfactory while the intellectual competence normally presumed behind it is absent or insufficient. An article may contain authentic references, plausible conclusions, and grammatically impeccable explanations without establishing that its nominal author understands the argument. An editor may accept the article after completing the required procedural stages without establishing that its central claims have received competent scrutiny. A university may subsequently recognize the publication as evidence of scholarly achievement. At every stage, the formal requirements may appear satisfied, while the substantive basis of that recognition remains uncertain.

This is the mechanism of academic abdication examined in the preceding chapters. It occurs when responsibility for intellectual judgment remains formally assigned to a human being, but the relevant person either cannot exercise that judgment or is not provided with the conditions necessary to exercise it meaningfully. Its causes may include inadequate expertise, excessive workloads, inappropriate institutional incentives, or habitual reliance upon automated systems. The presence of artificial intelligence is neither a necessary nor a sufficient condition. What makes contemporary generative technologies significant is their capacity to widen the distance between the speed of intellectual production and the resources required for independent evaluation.

The first consequence concerns the meaning of scholarly authorship. Authorship cannot reasonably require a researcher to perform every calculation personally, compose every sentence without assistance, or possess exhaustive knowledge of every discipline relevant to a collaborative investigation. Modern science depends upon instruments, specialization, and intellectual cooperation. Nevertheless, the attribution of a scholarly contribution presupposes a defensible relationship between the author and the claims presented under that author’s name. A researcher must be capable of explaining the central reasoning, identifying the evidentiary foundation, understanding the relevant methodological limitations, and responding responsibly to substantive criticism. Where specialist contributions are involved, their provenance and responsibility must remain identifiable. A polished manuscript cannot substitute for these conditions.

Disclosure of artificial intelligence assistance is an important element of transparency, but it cannot establish intellectual competence. The statement that a manuscript was prepared with the assistance of an LLM tells the reader something about its production.

It does not demonstrate that its sources were examined or its arguments understood. Equally, the absence of such a disclosure cannot establish scholarly merit. Scientific publishing must therefore avoid confusing the identification of a technological instrument with the evaluation of the intellectual work performed through it. The central question is not simply who or what produced the words, but whether the person responsible for the scientific claims can justify them.

The second consequence concerns peer review. A review procedure acquires scholarly legitimacy through the quality of the judgment exercised within it. Editorial appointments, institutional affiliations, submission platforms, and standardized evaluation forms cannot create expertise where it is absent. Nor can they guarantee that a qualified reviewer has sufficient time to examine a difficult manuscript. Effective verification requires the appropriate allocation of disciplinary competence, reasonable opportunities for independent examination, access to relevant evidence, and procedures through which uncertainty, disagreement, and error can be recorded and addressed. These conditions impose costs, but eliminating them does not necessarily eliminate the costs of scientific evaluation. It may merely transfer those costs to future researchers, institutions, and readers who must confront inadequately examined claims after publication.

This does not mean that every reviewer must reproduce an entire research project or personally verify every proposition within a multidisciplinary paper. Such an expectation would be unrealistic and could make scientific publishing impossible. The appropriate standard is proportionate, competent, and transparent evaluation. Reviewers should examine the claims they are qualified to assess, identify material questions requiring additional expertise, and avoid representing unperformed verification as completed work. Editors must exercise judgment in selecting reviewers, interpreting their reports, and deciding whether unresolved deficiencies justify rejection, revision, or further investigation. Accountability requires an identifiable human decision and a reconstructible basis for that decision. A signature, an approval button, or a completed checklist is not itself evidence that the corresponding intellectual examination occurred.

The third consequence concerns the educational mission of universities. Academic qualifications have meaning because they are understood to recognize capabilities developed through study and research. If students can satisfy increasingly demanding formal requirements without acquiring the corresponding competence, the relationship between academic certification and intellectual achievement becomes weaker. This possibility cannot be resolved solely through prohibitions on generative technologies or attempts to identify machine-produced prose. Universities must retain credible methods of establishing whether students understand their subjects, can defend their reasoning, recognize competing interpretations, and respond when evidence contradicts their initial conclusions. Substantive supervision, methodological examinations, discussion of research decisions, and the critical assessment of intellectual development may be more informative than reliance upon the finished document alone.

Such measures must be designed with respect for academic freedom and the legitimate use of technological assistance. The objective is not to compel researchers to demonstrate every moment of independent thought or expose their private intellectual processes to

institutional surveillance. It is to ensure that the achievement recognized by an academic qualification corresponds to an actual human capability. The use of computational instruments is compatible with that objective when the researcher retains the competence necessary to direct, interpret, and evaluate their contribution.

The fourth consequence is institutional and economic. Scientific knowledge is reproduced through people, relationships, and organizations whose development frequently exceeds the horizon of annual budgets. Research schools require continuity; mentorship requires time; fundamental investigation may yield its most important results only after prolonged uncertainty. The statistical material examined in this study documents changes in academic employment, doctoral career pathways, research expenditure, university revenues, and financial conditions across three national systems. These observations do not establish a common trajectory of intellectual decline. They do, however, demonstrate why the sustainability of academic expertise cannot be evaluated through expenditure totals or publication volumes alone.

A university may legitimately seek to reduce administrative costs, improve operational efficiency, and employ automation to support research. The relevant economic question is what happens to the resources thereby released. If technological efficiencies allow experienced scholars to devote more time to teaching, mentorship, experimental work, and difficult intellectual problems, automation may strengthen the very capacities whose preservation concerns us. If the same efficiencies are treated exclusively as opportunities to increase measurable output while reducing support for the development of expertise, the institution risks weakening its long-term intellectual foundations. The possibility of such a trade-off is a reason for systematic evaluation, not a finding that every university has already made the wrong choice.

Here the limitations of the present research must remain explicit. Neither university budgets nor standardized educational assessments directly measure the independent judgment of scholars. Employment data do not establish the quality of mentorship, and changes in research expenditure do not reveal whether individual scientific traditions are becoming more or less sustainable. The evidence assembled for the United States, Canada, and England identifies institutional conditions, historical transformations, and questions requiring investigation. It does not establish that generative artificial intelligence has caused a measurable decline in scientific competence or that such a decline is inevitable. Those propositions would require longitudinal research capable of distinguishing technological effects from pre-existing institutional developments.

Such research is both possible and necessary. Universities could examine whether students retain the ability to explain and reconstruct arguments developed with computational assistance, whether different approaches to AI-supported education affect independent problem-solving over time, and whether the growth of automated manuscript production changes the accuracy, duration, and substantive depth of peer review. Longitudinal studies of research training could investigate the relationship between employment continuity, mentorship, and the development of independent scientific expertise. These investigations would require carefully defined populations, appropriate comparison groups, and a willingness to accept results that contradict expectations. The concern defended here is a

researchable institutional hypothesis; it should not become an ideology protected from empirical criticism.

The practical implications are nevertheless sufficiently clear to justify action without awaiting a definitive historical verdict. Universities and scholarly publishers already possess responsibilities that do not depend upon resolving the future capabilities of artificial intelligence. They must distinguish formal compliance from substantive evaluation, require accountable authorship, support competent independent review, and maintain conditions under which scientific expertise can be transmitted. They must also preserve the capacity to acknowledge institutional errors and correct the scholarly record. These obligations arise from the purpose of scientific inquiry itself, rather than from a particular technological controversy.

The deeper issue is the legitimacy of academic authority. A university possesses the power to confer qualifications, distribute recognition, and influence the development of knowledge. A scientific journal may determine which arguments enter the published record under its editorial endorsement. These forms of authority are necessary, but their exercise must remain answerable to standards external to the preferences of those who hold them. Neither an academic title nor an institutional seal can transform an unsupported proposition into a demonstrated conclusion. A persistent identifier establishes a durable reference to a publication; it does not certify the truth of its contents. Peer review may justify serious attention to a manuscript, but it cannot exempt that manuscript from subsequent criticism.

This is why the philosophical and legal examples considered earlier are more than literary ornament. Kant's insistence upon the independent exercise of understanding and Lord Atkin's refusal to allow institutional authority to determine the meaning of its own legal limitations illuminate a common intellectual principle. A claim to authority cannot replace the grounds upon which that authority is justified. Humpty Dumpty's extraordinary confidence becomes genuinely dangerous only when the power to declare is mistaken for the power to establish what is true. Scholarship survives through the opposite discipline: the willingness to submit even confident and authoritative propositions to independent examination.

The preservation of the human academy is therefore neither a programme of technological resistance nor an appeal to an idealized past. It is an institutional commitment to maintaining the relationship between knowledge, understanding, judgment, and responsibility. That commitment does not exclude independent researchers working beyond traditional universities, collaborative research organizations, or increasingly sophisticated computational instruments. It requires that scholarly authority remain connected to demonstrable intellectual competence wherever research is undertaken and wherever its results are evaluated.

The title *Academia Without Scholars* should consequently be understood as a warning about a possible institutional condition, not as a declaration that such a condition has already become universal. A university without scholars need not resemble an abandoned institution. It may possess impressive buildings, extensive budgets, sophisticated digital systems, flourishing publication statistics, and an international reputation. What it would lack is something less immediately visible and more difficult to replace: a sufficiently independent

community of people capable of understanding the knowledge it preserves, judging the claims it circulates, and creating knowledge that does not yet exist.

We cannot predict precisely what tomorrow's universities or artificial intelligence systems will become. That uncertainty provides no reason to suspend the standards by which scholarship is distinguished from its imitation. The responsibility to preserve independent inquiry belongs to the institutions and people who possess the authority to shape academic life today.

The decisive question is no longer whether machines can write like scholars. It is whether universities will continue to require that someone understands, verifies, and takes responsibility for what is written.

ACADEMIA WITHOUT SCHOLARS: ARTIFICIAL INTELLIGENCE AND THE EROSION OF INDEPENDENT JUDGMENT

EXHIBITS A–D | UNITED STATES • CANADA • ENGLAND

Empirical documentation supporting Chapter V, Exhibits A-C relate to Chapter 6, Exhibit D relates to Chapter 5. Reference date: 15 October 2026. Historical anchor: approximately 1990; contemporary observations reflect the latest compatible published data identified for each measure.

Reading rule. Enrolments, test results, academic employment, finances and R&D expenditure describe different phenomena. None directly measures the originality of scholarship, independent judgment or the causal effects of generative AI. The data include improvements and stability as well as adverse changes.

EXHIBIT A — Educational Scale and Measured Skills

A.1. Higher-education participation: observations, not forced growth rates

System / population	Historical observation	Recent observation	Comparison status
United States: enrolments at degree-granting institutions	13,818,600 (1990)	19,659,600 (2024)	NCES published series; historic institutional coverage has changed [A01]
Canada: university registrations / enrolments	841,328 (1990)	1,439,850 (2024/25)	Historical full/part-time categories differ from current domicile-based counts [A02–A03]
England: UK-domiciled higher-education students in England	800,600 (1990/91)	—	No definition-matched contemporary comparison established [A04]
England: English-domiciled higher-level enrolments in England	—	2,040,275 (2024/25)	DfE category includes apprenticeships and other higher-level learning; distinct from historical HE count [A05]

Canada: 1990 = 532,131 full-time + 309,197 part-time. In 2024/25, 1,172,976 domestic and 266,874 international university enrolments were recorded. These are different collection frameworks; no percentage growth is calculated. England: the two listed populations differ in domicile, programme coverage and institutional universe. The UK-wide HESA total of 2,863,180 in 2024/25 is not an England-only comparator [A06]. The 2024/25 figures represent programme enrolments at the autumn reference date, not necessarily unique persons [A03].

A.2. School assessments: longitudinal observations within each programme

Assessment	Baseline	Recent	Difference / qualification
United States: NAEP grade-4 mathematics	213 (1990)	237 (2024)	+24 scale points [A07]
United States: NAEP grade-4 reading	217 (1992)	215 (2024)	−2; not statistically significant [A08]
Canada: PISA reading, age 15	534 (2000)	507 (2022)	−27 points [A09]
Canada: PISA mathematics, age 15	532 (2003)	497 (2022)	−35 points [A09]
England: PISA reading, age 15	496 (2006)	496 (2022)	0 observed change; 2022 sampling caveats [A10–A11]
England: PISA mathematics, age 15	495 (2006)	492 (2022)	−3 observed points; 2022 sampling caveats [A10–A11]

NAEP, PISA and PIAAC involve distinct ages, scales and populations; their raw scores must not be pooled or read as measures of university research quality. The NAEP reading series above is the national all-schools result, not the public-schools-only estimate. England's 2022 PISA sampling qualifications are disclosed by the Department for Education.

A.3. Adult skills: evidence against an undifferentiated decline thesis

OECD Survey of Adult Skills, 2022/23	Literacy	Numeracy	Adaptive problem solving
Canada, adults 16–65	271	271	259
England, adults 16–65	272	268	259

OECD country notes [A12–A13]. Canadian and English literacy was broadly stable against the preceding survey cycle, while numeracy improved. Adaptive problem solving is presented as a contemporary observation. These are adult population skills, not measures of independently produced scholarship.

EXHIBIT B — Academic Workforce and Research Economics

B.1. United States: structure of academic appointments

Appointment category	1987	2023	Interpretation
Contingent appointments, outside tenure line	~47%	~68%	Employment structure [B01]
Full-time tenured or tenure-track	~53%	~32%	Complementary category [B01]
Part-time faculty	~33%	~49%	Subset overlapping contingent appointments [B01]

AAUP analysis of federal academic-workforce datasets. Categories may overlap: part-time is not an additional population to add to contingent. Employment type alone is not evidence of competence, teaching quality, disciplinary expertise, or mentorship performance. The AAUP population comprises appointments of nonmedical instructional staff, with or without formal faculty status, at degree-granting nonprofit institutions; graduate student employees are reported separately [B01].

B.2. Doctorate recipients: definite non-postdoctoral employment in the United States

Destination measure	2004	2024	Denominator
Academic employment sector	56%	40%	Recipients with definite U.S. non-postdoc employment commitments [B02]
Industry / business sector	—	~40%	Same 2024 group; not all doctorate recipients [B02]

In the 2024 definite-commitment group, 8,098 recipients identified academia and 8,266 industry/business among 20,469 recipients. Movements toward industry may signify opportunities outside universities, not a decline in scholarly ability. These sector shares do not, by themselves, establish a decline in the absolute number of academic positions [B02].

B.3. University research expenditure: distinct national series

Indicator	Latest observed	Reported change	Measurement boundary
United States: higher-education R&D	US$117.7bn (FY 2024)	+8.1% nominal year-on-year	US institutional HERD; +~3% annual real growth 2014–24 [B03]
Canada: higher-education R&D	C$19.0bn (2023/24)	+7.2% year-on-year	Canadian HERD sector; not an exchange-rate-adjusted comparator [B04]

U.S. and Canadian R&D totals cannot be ranked against English higher-education-provider income. Research expenditure can rise without an identical change in scholarly quality, mentorship or continuity.

B.4. University finances: Canada and England have different reporting frameworks

Population / measure	Reported observation	Scope and status
Canada: university revenue, 2023/24	C$52.4bn	Inflation-adjusted 2024 Canadian dollars [B05]
Canada: provincial grants / tuition share	32.4% / 30.1%	Share of Canadian university revenues; tuition share up 5.4 pp over ten years [B05]
England: HE-provider income, 2024/25	£46.1bn	Total England provider revenue, not R&D spending [B06]
England: tuition / funding-body grants	Tuition >50%; grants 11%	2024/25 provider income composition [B06]
England: deficits — publicly funded universities	48 institutions; 40%	Parliamentary/HESA university subgroup, 2024/25 [B06]
England: deficits — OfS-regulated providers	35.8%	Different OfS denominator, 2024/25 [B07]
England: predicted deficits, 2025/26	42.7%	OfS forward-looking provider forecasts; not observed outcomes [B07]
England: restructuring costs, 2024/25	£218.2m	OfS sector costs; 20.7% above 2023/24 [B07]

England: 48 universities/40% [B06] and OfS 35.8% [B07] concern different institutional populations and must not be combined. The OfS describes 42.7% as a prospective deficit exposure derived from submitted forecasts and its assessment of potential financial outcomes, including the role of mitigating actions. Research England's direct support for research capacity is a separate series from total provider income [B06].

EXHIBIT C — Primary Sources, Provenance and Limits of Inference

The identifiers below apply only to these Exhibits; the article-wide References remain separately numbered. Links identify official statistical or institutional sources. Dates refer to the relevant observation or publication, not to an inferred common census year.

C.1. Source register

[A01] NCES, Digest of Education Statistics, Table 306.10 (1990 and 2024 enrolments). https://nces.ed.gov/programs/digest/d25/tables/dt25_306.10.asp

[A02] Senate of Canada, Subcommittee on Post-Secondary Education, final report (1997), historical enrolment counts. https://sencanada.ca/en/content/sen/committee/361/post/rep/repfinaldec97part2-e

[A03] Statistics Canada, Canadian postsecondary enrolments and graduates 2024/25, The Daily (1 Oct 2026). https://www150.statcan.gc.ca/n1/daily-quotidien/261001/dq261001a-eng.htm

[A04] House of Commons, Hansard (20 June 1994), UK-domiciled higher education students studying in England, 1990/91. https://api.parliament.uk/historic-hansard/written-answers/1994/jun/20/student-numbers

[A05] Department for Education, Higher-Level Learners in England, 2024/25 (8 Oct 2026); English-domiciled learners. https://explore-education-statistics.service.gov.uk/find-statistics/higher-level-learners-in-england/2024-25

[A06] HESA, Higher Education Student Statistics: UK, 2024/25; UK-wide only as context. https://www.hesa.ac.uk/data-and-analysis/sb273/figure-7

[A07] NCSES / NSF, STEM Talent, NAEP mathematics, Table STAL-15, all grade-4 students, 1990–2024. https://ncses.nsf.gov/pubs/nsb20261/table/STAL-15

[A08] NCES, The Nation's Report Card: Reading 2024, grade 4, national trends. https://www.nationsreportcard.gov/reports/reading/2024/g4_8/national-trends/

[A09] OECD, PISA 2022 Results, Vol. I, Canadian country trends and historical assessments. https://doi.org/10.1787/53f23881-en

[A10] Department for Education, PISA 2018: National Report for England (historical trend tables). https://www.gov.uk/government/publications/pisa-2018-national-report-for-england

[A11] Department for Education, PISA 2022: National Report for England, RR1392, incl. sampling caveats. https://www.gov.uk/government/publications/pisa-2022-national-report-for-england

[A12] OECD, Survey of Adult Skills 2023: Canada, country note (2024). https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/canada_5ecab9d9-en.html

[A13] OECD, Survey of Adult Skills 2023: England (United Kingdom), country note (2024). https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/united-kingdom_02bc78e4-en.html

[B01] AAUP, Data Snapshot: Tenure and Contingency in U.S. Higher Education, Fall 2023 (2025). https://www.aaup.org/academe/issues/spring-2025/data-snapshot-tenure-and-contingency-us-higher-education-fall-2023

[B02] NCSES, Doctorate Recipients from U.S. Universities: 2024, postgraduation trends. https://ncses.nsf.gov/pubs/nsf26315/report/postgraduation-trends

[B03] NCSES, University R&D expenditure in FY 2024; NSF 26-305 (2026). https://ncses.nsf.gov/pubs/nsf26305

[B04] Statistics Canada, Higher Education R&D 2023/24, The Daily (17 Oct 2025). https://www150.statcan.gc.ca/n1/daily-quotidien/251017/dq251017e-eng.htm

[B05] Statistics Canada, University Revenues 2023/24, The Daily (29 July 2025). https://www150.statcan.gc.ca/n1/daily-quotidien/250729/dq250729b-eng.htm

[B06] Bolton & Lewis, Higher Education Finances and Funding in England, Commons Library CBP-10037 (4 Sep 2026). https://commonslibrary.parliament.uk/research-briefings/cbp-10037/

[B07] Office for Students, Financial Sustainability of Higher Education Providers in England: 2026 (updated June). https://www.officeforstudents.org.uk/publications/financial-sustainability-of-higher-education-providers-in-england-2026/

C.2. Inference boundaries

- Geography: only the United States, Canada and England are substantive cases. UK-wide data do not replace England-specific observations.
- Categories: Canada 1990/2024–25 student counts and England 1990/91/2024–25 higher-learning counts are non-matched observations; no percentage growth is reported for those pairs.
- Instruments: NAEP and PISA are school measures; OECD PIAAC is an adult-skills measure. None is a direct test of research judgment or of scientific originality.
- Employment: AAUP part-time and contingent categories overlap. The doctorate employment figures refer to graduates with definite U.S. non-postdoc job commitments, not all doctoral graduates.
- Finances: U.S. and Canadian HERD, Canadian university revenue and England HE-provider income have different institutional scopes, currencies, accounting bases and purposes; no ranking or conversion is justified.
- Forecasts: the OfS 2025/26 deficit statistic is forward-looking, not an observed loss rate. Its provider population is distinct from the publicly funded university group reported in Parliament.
- Causality: pre-LLM institutional trends cannot be ascribed to generative AI. Descriptive endpoints do not establish a common causal path or a general collapse of academic competence.
- Balance: increases in research resources, stable literacy and improved adult numeracy are retained alongside documented financial and institutional pressures.

Editorial scope statement. The selected indicators support inquiry into the conditions for reproducing scientific expertise; they are not measurements of scholarly merit. The series have been sourced to official records, while the excluded like-for-like comparisons are intentionally withheld rather than supplied with unsupported estimates.

EXHIBIT D — Selected Evidence on Cognitive Delegation and Independent Judgment

Methodological matrix for Chapter 5 · selected published studies · reference numbers 1-7.

Study / source	Research design and population	Principal observation	Scope and inferential limits
Risko & Gilbert (2016) [1] · 10.1016/j.tics.2016.07.002	Conceptual review of cognitive offloading; no experiment or participant sample.	External aids redistribute demands on memory and cognition. Offloading may support rather than impair thinking.	General cognitive framework, not LLM-specific evidence. Neither cognitive decline nor causation is tested.
Bastani et al. (2025) [2] · 10.1073/pnas.2422633122	Field experiment in secondary-school mathematics; approximately 1,000 pupils; unrestricted GPT-4 assistance compared with a guided tutor and controls.	Assisted exercise performance improved, yet unrestricted assistance was associated with about 17% worse unassisted assessment performance than the no-AI control.	School mathematics and particular tutoring interfaces; a learning-performance gap, not evidence of permanent impairment or university-wide effects.
Barcaui (2025) [3] · 10.1016/j.ssaho.2025.102287	Randomized undergraduate study; 120 initially allocated; 85 completed a surprise retention assessment after 45 days.	Mean retained-knowledge scores: 57.5% with ChatGPT assistance versus 68.5% in the comparison condition (reported p = 0.002).	35/120 participants absent at follow-up (29.2% attrition); one teaching setting. Does not establish generalized cognitive deterioration.
Makransky et al. (2025) [4] · 10.1007/s10648-025-10039-x	Two controlled experiments: 175 university students and 234 school students; theory-guided conversational tutoring versus generic assistance and comparators.	Structured prompting to explain and construct understanding produced stronger conceptual learning outcomes than generic ChatGPT in the tested settings.	Effects depend on instructional design and outcome measure. Material counterevidence against treating all AI use as harmful.
Yan (2026) [5] · 10.3389/fpsyg.2026.1936522	Randomized scenario experiment with 240 preservice teachers (218 in principal analysis); accurate and deliberately misleading AI advice.	For misleading items, independent judgment before AI advice was associated with higher expert-keyed accuracy (53.5% versus 37.3%).	Single training context and simulated pedagogical decisions; examines advice susceptibility, not longitudinal decline.
Qiang & Wang (2026) [6] · 10.3390/jintelligence14100254	Qualitative accounts and accompanying materials from 25 university teachers using generative AI.	Teachers described cognitive delegation, professional judgment and deliberate verification or "metacognitive defense" strategies.	Small purposive/self-selecting sample; narratives do not quantify competence loss or establish causality.
Boolzen et al. (2026) [7] · 10.1007/s10462-026-11665-9	Systematic review: 85 eligible studies; meta-analysis: 49 studies and 59 effect sizes in STEM learning.	Average positive effects in conventional synthesis were strongly qualified by very high heterogeneity and sensitivity to publication-bias adjustment.	No included study had a low overall risk-of-bias rating. Neither universal learning gains nor generalized harm is established.

Interpretive boundary. These studies address cognitive offloading, assisted performance, retention, advice reliance and scholarly practice. They do not establish generalized or irreversible cognitive deterioration among students or university faculty.

Note. Bibliographic details and DOI are provided in Chapter 5, reference numbers 1-7. Evidence types and denominators are deliberately kept separate.

References

1. Kildeev, A. (2025a). Artificial Intelligence Is Slave of the Lamp. Scientific Platform XXI Century, No. 6 (September 2025), 3–31. Published journal version. https://episs.ru/
2. Kildeev, A. (2025b). LLM Is Not Artificial Intelligence. Scientific Platform XXI Century, No. 7 (October 2025), 3–44. Published journal version. https://episs.ru/
3. Liebig, T., & Kildeev, A. (2026). Instrumentum Vocale and the Architecture of Responsibility: From Liability to Embedded Accountability. Canadian Review of Artificial Intelligence, Law & Society (CRAIL), 1(1), 11–84. Published journal version. https://crail.ca/
4. University of Oxford. Our History. https://www.ox.ac.uk/about/the-university/history; University of Cambridge. The Medieval University. https://www.cam.ac.uk/about-the-university/history/the-medieval-university

5. The Royal Society. History of the Royal Society. https://royalsociety.org/about-us/who-we-are/history/
6. Harvard University. The History of Harvard University. https://www.harvard.edu/about/history/
7. University of Toronto Archives & Records Management Services. The University's Original Charter. https://utarms.library.utoronto.ca/history/university-original-charter ; Université Laval. Origin and History. https://www.ulaval.ca/en/about-us/origin-and-history
8. Johns Hopkins University. History & Mission. https://www.jhu.edu/about/history/
9. Polanyi M. The Tacit Dimension. Garden City, NY: Doubleday, 1966. Reissued by the University of Chicago Press, 2009.
10. Lave J., Wenger E. Situated Learning: Legitimate Peripheral Participation. Cambridge: Cambridge University Press, 1991. DOI: 10.1017/CBO9780511815355.
11. National Academies of Sciences, Engineering, and Medicine. The Science of Effective Mentorship in STEMM. Washington, DC: National Academies Press, 2019. DOI: 10.17226/25568.
12. Boyer Commission on Educating Undergraduates in the Research University. Reinventing Undergraduate Education: A Blueprint for America's Research Universities. Stony Brook, NY: State University of New York at Stony Brook, 1998. https://repo.library.stonybrook.edu/xmlui/handle/11401/57972
13. Natural Sciences and Engineering Research Council of Canada. Discovery Grants Program (Individual). Official programme documentation. https://nserc.canada.ca/en/funding-opportunity/discovery-grants-program-individual
14. UK Research and Innovation / Research England. How We Fund Higher Education Providers. https://www.ukri.org/publications/research-england-how-we-fund-higher-education-providers/how-we-fund-higher-education-providers/
15. State Higher Education Executive Officers Association. State Higher Education Finance: FY 2025. Published 2026. https://shef.sheeo.org/report/
16. Statistics Canada. University Finances, 2002/03. The Daily, 19 August 2004; University Revenues Increase More Than Expenditures in 2023/2024. The Daily, 29 July 2025.
17. Further and Higher Education Act 1992, c. 13. https://www.legislation.gov.uk/ukpga/1992/13/contents
18. House of Commons Library. Tuition Fees in England: History, Debates, and International Comparisons. Research Briefing CBP-10155, 23 February 2026; Higher Education and Research Act 2017, c. 29.
19. Bolton, P., and Lewis, J. Higher Education Finances and Funding in England. House of Commons Library, Research Briefing CBP-10037, 4 September 2026.
20. Slaughter, S., and Rhoades, G. Academic Capitalism and the New Economy: Markets, State, and Higher Education. Johns Hopkins University Press, 2004.

21. Bok, D. Universities in the Marketplace: The Commercialization of Higher Education. Princeton University Press, 2003.
22. San Francisco Declaration on Research Assessment (DORA), 2012/2013. https://sfdora.org/read/
23. Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., and Rafols, I. "Bibliometrics: The Leiden Manifesto for Research Metrics." Nature, 520 (2015): 429–431. https://doi.org/10.1038/520429a
24. UK Research and Innovation. Research Excellence Framework. Official documentation, updated 2026. https://www.ukri.org/who-we-are/research-england/research-excellence/research-excellence-framework/
25. Colby, G. "Data Snapshot: Tenure and Contingency in US Higher Education." Academe, American Association of University Professors, Spring 2023.
26. Statistics Canada. "Part-time University Faculty, 1990/91 to 1997/98." The Daily, 8 May 2002.
27. Sinopoli, A. (2026). EPISTEMÍA: Governing the human process of understanding: A proposal for a new object of legal protection in the age of artificial intelligence. The Journal Scientific Platform XXI Century, No. 5 (14), 40–72. https://episs.ru/wp-content/uploads/2026/09/5-14-ISSUE-2026-English-FINAL.pdf
28. Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (Article 1121, pp. 1–22). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713778
29. Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task (arXiv:2506.08872, version 2). arXiv. Research preprint. https://doi.org/10.48550/arXiv.2506.08872
30. Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8
31. Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3), 381–410. https://doi.org/10.1177/0018720810376055
32. Sinopoli, A. (2026). Cognitive dignity: A proposal for an emerging human right: Protecting the human process of sense-making and knowledge formation against technological substitution. Canadian Review of Artificial Intelligence, Law & Society (CRAIL), 1(2), 89–117. https://www.crail.ca/CRAIL_Vol.1-No.2-2026-October.pdf
33. Carroll, L. (1871). Through the Looking-Glass, and What Alice Found There. London: Macmillan (first published December 1871; title-page dated 1872). Chapter VI, "Humpty Dumpty." Available at: https://www.gutenberg.org/files/12/12-h/12-h.htm.

34. Liversidge v Anderson [1941] UKHL 1; [1942] AC 206. House of Lords, judgment of 3 November 1941. Dissenting opinion of Lord Atkin, especially pp. 225–247, 244–245. Available at: https://uniset.ca/other/cs5/1942AC206.html.
35. Kant, I. (1784). Beantwortung der Frage: Was ist Aufklärung? Berlinische Monatsschrift, 4 (December), 481–494. Original German text available at: https://de.wikisource.org/wiki/Beantwortung_der_Frage:_Was_ist_Aufklärung%3F.
36. Frankfurt, H. G. (1986). On Bullshit. Raritan, 6(2), 81–100. Reissued as: Frankfurt, H. G. (2005). On Bullshit. Princeton: Princeton University Press. Available at: https://philpapers.org/rec/FRAOBS.
37. Merton, R. K. (1973). The Normative Structure of Science. In: The Sociology of Science: Theoretical and Empirical Investigations. Chicago: University of Chicago Press, pp. 267–278. Originally published as "Science and Technology in a Democratic Order," Journal of Legal and Political Sociology, 1 (1942), 115–126. Available at: https://sciencepolicy.colorado.edu/students/envs_5110/merton_sociology_science.pdf.
38. COPE & STM. (2022). Paper Mills: Research Report from COPE & STM. Committee on Publication Ethics and International Association of Scientific, Technical and Medical Publishers. Version 1, June 2022, 15 pp. https://doi.org/10.24318/jtbG8IHL
39. United2Act. (2025). United2Act Phase 1: Report and Recommendations. International Association of Scientific, Technical and Medical Publishers (STM). Published September 3, 2025. https://stm-assoc.org/united2act-phase-1-report-and-recommendations/
40. Kobak, D., González-Márquez, R., Horvát, E.-Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. Science Advances, 11(27), eadt3813. https://doi.org/10.1126/sciadv.adt3813
41. COPE Council. (2024). Authorship and AI Tools. Committee on Publication Ethics. Position statement, version 1, August 20, 2024. https://doi.org/10.24318/cCVRZBms
42. Responsible and Transparent Use of AI in Scientific Publishing. (2026). Nature Computational Science, 6(8), 803. Published August 20, 2026. https://doi.org/10.1038/s43588-026-01043-4
43. Kildeev, A. Procedural Liability in the Age of LLMs. Scientific Platform XXI Century, No. 7 (16) (July 2026), 27–51. Published journal version. https://episs.ru/
44. Kildeev, A. The Tool Has a Manufacturer: Producer Liability for Legal AI. Canadian Review of Artificial Intelligence, Law & Society (CRAIL), 1(2), 118–153. Published journal version. https://www.crail.ca/CRAIL_Vol.1-No.2-2026-October.pdf
45. Unger, W. (2026). AI Use in Legal Scholarship. SSRN Working Paper. Preliminary research memorandum, abstract and appendices, 15 pp. Posted

August 17, 2026; last revised August 25, 2026. https://doi.org/10.2139/ssrn.7290238
46. Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. Nature, 631, 755–759. https://doi.org/10.1038/s41586-024-07566-y
47. Risko, E. F. Cognitive Offloading / E. F. Risko, S. J. Gilbert // Trends in Cognitive Sciences. 2016. Vol. 20, no. 9. pp. 676–688. DOI: 10.1016/j.tics.2016.07.002.
48. Bastani, H. Generative AI without guardrails can harm learning: Evidence from high school mathematics / H. Bastani, O. Bastani, A. Sungu [et al.] // Proceedings of the National Academy of Sciences. 2025. Vol. 122, no. 26. Article e2422633122. DOI: 10.1073/pnas.2422633122.
49. Barcaui, A. ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention / A. Barcaui // Social Sciences & Humanities Open. 2025. Vol. 12. Article 102287. DOI: 10.1016/j.ssaho.2025.102287.
50. Makransky, G. Beyond the “Wow” Factor: Using Generative AI for Increasing Generative Sense-Making / G. Makransky, B. M. Shiwalia, T. Herlau, S. Blurton // Educational Psychology Review. 2025. Vol. 37. Article 60. DOI: 10.1007/s10648-025-10039-x.
51. Yan, D. Independent judgment before viewing generative AI recommendations reduces susceptibility to misleading advice: A randomized scenario-based experiment with preservice teachers / D. Yan // Frontiers in Psychology. 2026. Vol. 17. Article 1936522. DOI: 10.3389/fpsyg.2026.1936522.
52. Qiang, Z. Between Human and Machine: University Teachers’ Accounts of Judgment, Metacognitive Defense, and Cognitive Offloading in Generative AI Use / Z. Qiang, Z. Wang // Journal of Intelligence. 2026. Vol. 14, no. 10. Article 254. DOI: 10.3390/jintelligence14100254.
53. Boolzen, C. Evidence of impact and interpretational limits of generative AI in STEM education: A systematic review and meta-analysis on cognitive learning outcomes / C. Boolzen, J. Kuhn, S. Flegr [et al.] // Artificial Intelligence Review. 2026. Vol. 59. Article 223. DOI: 10.1007/s10462-026-11665-9.
54. National Center for Science and Engineering Statistics. (2026). Doctorate Recipients from U.S. Universities: 2024. NSF 26-315. https://ncses.nsf.gov/pubs/nsf26315/report/postgraduation-trends
55. National Center for Science and Engineering Statistics. (2026). Universities Report 8.1% Growth in R&D Expenditures in FY 2024, Reaching Over $117 Billion. NSF 26-305. https://ncses.nsf.gov/pubs/nsf26305
56. Statistics Canada. (2025). Spending on Research and Development in the Higher Education Sector, 2023/2024. The Daily, 17 October. https://www150.statcan.gc.ca/n1/daily-quotidien/251017/dq251017e-eng.htm
57. Statistics Canada. (2025). University Revenues Increase More Than Expenditures in 2023/2024. The Daily, 29 July. https://www150.statcan.gc.ca/n1/daily-quotidien/250729/dq250729b-eng.htm

58. Advisory Panel on the Federal Research Support System. (2023). Report of the Advisory Panel on the Federal Research Support System. Innovation, Science and Economic Development Canada. https://ised-isde.canada.ca/site/panel-federal-research-support/en/report-advisory-panel-federal-research-support-system
59. OECD. (2024). Survey of Adult Skills 2023: Canada, Country Note. https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/canada_5ecab9d9-en.html
60. Bolton, P., and Lewis, J. (2026). Higher Education Finances and Funding in England. House of Commons Library, Research Briefing CBP-10037, 4 September. https://commonslibrary.parliament.uk/research-briefings/cbp-10037/
61. Office for Students. (2026). Financial Sustainability of Higher Education Providers in England: 2026. 14 May 2026, updated report. https://www.officeforstudents.org.uk/publications/financial-sustainability-of-higher-education-providers-in-england-2026/
62. OECD. (2024). Survey of Adult Skills 2023: England (United Kingdom), Country Note. https://www.oecd.org/en/publications/survey-of-adults-skills-2023-country-notes_ab4f6b8c-en/united-kingdom_02bc78e4-en.html
63. Department for Education. (2023). PISA 2022: National Report for England. Research Report RR1392. https://www.gov.uk/government/publications/pisa-2022-national-report-for-england
64. Colby, G. (2025). Data Snapshot: Tenure and Contingency in US Higher Education, Fall 2023. Academe, 111 (Spring). American Association of University Professors. https://www.aaup.org/article/data-snapshot-tenure-and-contingency-us-higher-education-fall-2023
65. National Academies of Sciences, Engineering, and Medicine. (2018). Graduate STEM Education for the 21st Century. Washington, DC: The National Academies Press. https://doi.org/10.17226/25038



CRAIL

Canadian Review of Artificial Intelligence, Law & Society

DIGITUS NON EST IUDICIUM

Vol. 1 | No. 3 | Fall 2026, Part III

ISSN 3111-458X

crail.ca



CRAIL

CANADIAN REVIEW OF ARTIFICIAL INTELLIGENCE, LAW & SOCIETY

Vol. 1 | No. 3 | Fall 2026, Part III

NOVEMBER 2026

DIGITUS NON EST IUDICIUM

crail.ca