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).