

RESEARCH ARTICLE

Beyond the Traffic-Light Model: Legal Epistemic Traceability, Human- Accountable Reasoning and the Attribution of AI-Assisted Legal Acts

A Contextual Framework for European Union and Spanish Law after the 2026 Digital Omnibus

AUTHOR José Manuel López, BSc, MD, PhD, LLD (C)

ARTICLE TYPE Doctrinal, interdisciplinary and normative research article

FIELDS Artificial intelligence law; procedural law; legal ethics; evidence; judicial governance

JURISDICTIONAL FOCUS European Union and Spain, with selected comparative materials

LAW CURRENT TO Closing date: August 2026

RESEARCH DESIGN Doctrinal analysis + socio-technical analysis + structured normative case testing

CITATION SYSTEM APA 7th edition adapted to legal and institutional sources

Editorial positioning note. This article is not a restatement of professional “allowed / prohibited” guidance. It develops a distinct theory of legal attribution in AI-assisted work, centred on the AI-assisted legal act, Legal Epistemic Traceability (LET), an Epistemic Adoption Test (EAT), and bounded automation for agentic systems.

© 2026 · José M. López · All rights reserved

Abstract

The rapid normalisation of generative and agentic artificial intelligence in legal practice exposes a structural weakness in prevailing regulatory discourse. Practical guidance commonly asks where AI may or may not be used and answers through traffic-light classifications, professional cautions, or system-level risk categories. This article argues that those approaches are insufficient because legal permissibility is not a stable property of an AI tool. It is a contextual property of a concrete AI-assisted legal act. The article formulates an integrated framework for assessing such acts under European Union and Spanish law after Regulation (EU) 2026/1744, with particular attention to the AI Act, the GDPR, professional secrecy, the right of defence, judicial independence, procedural fairness and evidentiary reliability. It advances three original analytical constructs. First, Legal Epistemic Traceability (LET) is defined as the capacity to reconstruct, independently of the AI system itself, the normative and evidential chain through which an AI-assisted proposition becomes attributable to a legally responsible human actor. Secondly, an Epistemic Adoption Test (EAT) specifies when machine output may legitimately be treated as a professional proposition: the responsible actor must know the role played by the AI, understand the material proposition, independently verify it to a degree proportionate to consequence, retain genuine capacity to reject it, and consciously adopt it as part of the actor's own legal reasoning. Thirdly, the article replaces weak human-in-the-loop formulations with human-accountable reasoning and, for agentic AI, with bounded automation based on authority, reversibility and legal consequence. The framework is applied to five legal scenarios, including public-source drafting, confidential RAG, autonomous legal research and filing, judicial drafting, and AI-mediated evidentiary analysis. The article concludes that the next phase of legal AI governance should move from tool classification to professional epistemic governance, where attribution, contestability and reconstructibility become central rule-of-law safeguards.

Keywords: *Artificial Intelligence Act; legal AI; professional responsibility; professional secrecy; legal epistemology; epistemic traceability; evidence; human oversight; agentic AI; judicial AI; automation bias; right of defence.*

Original contribution: The article links AI governance to legal attribution by proposing the AI-assisted legal act as the unit of analysis, Legal Epistemic Traceability as a rule-of-law standard, and the Epistemic Adoption Test as the threshold by which AI output becomes professionally attributable to a human legal actor.

1. Introduction: Why “Where Can AI Be Used?” Is the Wrong Primary Question

Artificial intelligence is becoming embedded in legal work at a speed that makes prohibition-centred analysis increasingly inadequate. Lawyers use generative systems to summarise authorities, compare clauses, review documents, prepare chronologies, generate first drafts and query large collections. Judicial institutions are developing controlled tools for legal research and drafting support. Agentic systems add a further layer: they can plan sub-tasks, call external tools, retrieve documents, interact with databases, send communications and, where permissions allow, execute actions in digital environments. The regulatory question has therefore moved from whether AI is used to how responsibility, knowledge and legal effect are allocated when AI participates in a professional act.

Practical commentary has often responded with a traffic-light model. Uses involving public information and low-consequence drafting are placed in a green zone; confidential information or client work in an amber zone; autonomous decision-making, disclosure of protected material or unverified legal propositions in a red zone. This is useful training material. It is not, however, a sufficient legal theory. A single colour can conflate distinct variables: who is acting, what information enters the system, what the system actually does, how the system is deployed, what legal consequence follows, how the human reviews the result, and whether the result can be reconstructed from authoritative sources.

The inadequacy of tool-level classification becomes obvious when the same model is used in different settings. A general-purpose model may be unobjectionable when asked to reorganise a public statute, professionally unacceptable when supplied with privileged criminal defence strategy through an uncontrolled

service, and potentially subject to a different regulatory regime when integrated into a system intended to assist a judicial authority in interpreting facts and law. Nothing about the underlying language model alone determines the answer.

This article therefore proposes a shift in the unit of legal analysis. The relevant object is the AI-assisted legal act: a concrete professional act in which an AI system materially performs or influences a cognitive or operational function. This move makes it possible to connect the technical reality of modern AI to the normative structure of professional secrecy, data protection, due process, evidence, judicial independence and professional responsibility. The expression “legal act” is used here in a functional sense to denote an act of legal professional practice, and not the civil-law dogmatic category of a juridical act or legal transaction.

Core proposition: Legal permissibility is not a stable property of an AI product. It is a contextual property of the AI-assisted legal act.

2. Research Question, Hypotheses and Claims of Contribution

2.1 Research question

The principal research question is: Can existing concepts of AI risk classification and human oversight adequately govern AI-assisted legal activity, or does generative and agentic AI require an autonomous standard of professional epistemic accountability that determines when machine-mediated propositions and actions become legally attributable to a human professional?

2.2 Working hypotheses

- H1 — Tool-level classifications are underdeterminate: the same AI system can move between lawful, conditionally lawful, professionally incompatible and prohibited use depending on actor, information, function, architecture, legal effect, human control and traceability.
- H2 — Formal human-in-the-loop review is not a sufficient safeguard in consequential legal activity because nominal human presence does not demonstrate independent cognition, resistance to automation bias or capacity to justify the outcome.
- H3 — Professional attribution of AI-assisted legal propositions requires epistemic adoption: awareness of the AI contribution, comprehension of the material proposition, proportionate independent verification, genuine power to reject, and conscious professional endorsement.
- H4 — Where AI contributes materially to evidence, adjudication, legally binding communications or procedural acts, legal accountability requires a reconstructible chain of provenance and verification that is distinct from purely technical logging.
- H5 — Agentic AI requires governance of delegated authority, not merely output quality; the appropriate mechanism is bounded automation, with human authorisation at legally significant transition points.

2.3 Claimed contribution and limits of originality

The article deliberately positions its claims against an emerging 2025–2026 literature. Zinzilieta (2026) argues for a nondelegable duty to think and a contemplation-first approach to generative AI. Webb (2026) analyses epistemic expertise and virtue in AI-assisted legal practice. Pardo (2026) develops the relationship between legal epistemology and proof, while Chin and Roberts (2026) examine the effect of AI on evidence and the process of proof. Cobbe, Lee and Singh (2021) propose reviewability as a socio-technical accountability framework; Green (2022) demonstrates the limits of human oversight policies; and Laux and Ruschmeier (2025) analyse automation bias in the AI Act.

The present contribution does not claim that legal thinking, epistemic competence, traceability or reviewability are new concerns. Its narrower and more defensible claim is that existing strands can be integrated into a specifically legal theory of attribution: Legal Epistemic Traceability (LET) identifies what must be reconstructible, while the Epistemic Adoption Test (EAT) specifies when a machine-originated or machine-transformed proposition can legitimately become the proposition of the responsible professional. The distinctive move is therefore from “duty to think” to a legally operational threshold of attribution, and from technical logging to evidentiary and normative reconstructibility.

Table 1. Positioning of the proposed framework against adjacent literature

Literature strand	Central concern	What it contributes	Distinct move in this article
Professional ethics / nondelegation	Epistemic atrophy, competence and professional judgment	Zinzilieta (2026); Webb (2026); Remus & Levy (2017)	Defines a legal attribution threshold through epistemic adoption rather than relying only on a general duty to think.
Human oversight	Automation bias and ineffective human review	Green (2022); Laux & Ruschmeier (2025); Banks (2026)	Replaces nominal oversight with human-accountable reasoning and criteria for meaningful adoption.
Algorithmic accountability	Reviewability, process records and institutional governance	Citron (2008); Cobbe et al. (2021)	Distinguishes technical/system reviewability from proposition-level legal epistemic traceability.
Evidence and proof	Authenticity, reliability, proof and AI-mediated evidence	Pardo (2026); Chin & Roberts (2026)	Connects provenance, AI transformation, verification and professional attribution in a single chain.
Agentic AI	Autonomous action and cross-system execution	AEPD (2026); Hacker & Holweg (2026)	Adds bounded automation based on legal consequence, authority and reversibility.

3. Methodology and Scope

The study adopts a doctrinal and interdisciplinary methodology. The doctrinal component analyses primary European Union and Spanish sources, including Regulation (EU) 2024/1689 as amended by Regulation (EU) 2026/1744, Regulation (EU) 2016/679, the Charter of Fundamental Rights of the European Union, the Spanish Constitution, Organic Law 5/2024 on the Right of Defence, Royal Decree 135/2021 approving the General Statute of the Spanish Legal Profession, the Civil Procedure Act, and Instruction 2/2026 of the General Council of the Judiciary. Institutional guidance from the EDPB, AEPD, CCBE, Spanish legal profession and Council of Europe is used to interpret operational expectations without treating soft law as equivalent to legislation.

The interdisciplinary component draws on research in legal epistemology, human-computer interaction, automation bias, AI reliability and algorithmic accountability. That literature is not used to derive legal obligations directly. It performs a different function: it tests the factual assumptions on which legal concepts such as human oversight, competence, explanation and verification depend. Where law assumes that a human can effectively supervise a machine, behavioural evidence about automation bias becomes legally relevant to whether the safeguard is meaningful in practice.

The article also uses structured normative case testing. Five scenarios are analysed through the proposed seven-factor contextual test and the Epistemic Adoption Test. These are not empirical case studies. They are controlled doctrinal hypotheticals designed to reveal when changes in data, architecture, delegated function or legal effect change the legal classification even though the underlying AI model remains constant.

The jurisdictional focus is Spain and the European Union. Comparative cases and professional standards, especially from common-law jurisdictions, are included only when they illuminate recurring problems such as

fabricated citations, authentication of AI-generated evidence or professional verification. The article does not attempt a general comparative survey of AI regulation.

4. Technical Premise: The Model Is Not the System and the Output Is Not the Act

A legal analysis that treats ChatGPT, Claude, Gemini, a legal RAG platform or an enterprise assistant as a single technical object risks missing the components that determine confidentiality, reliability and authority. A modern AI application may combine a foundation model, a system prompt, retrieval components, vector stores, access-control layers, plugins or tools, agent orchestration, logging, moderation, external APIs and human workflows. The legally relevant processing can occur at any of these layers.

This matters particularly for professional secrecy. A provider may truthfully state that customer prompts are not used to train its general-purpose model while still retaining logs, using sub-processors, storing embeddings, allowing diagnostic access, transferring data internationally or permitting tools to act on external systems. The phrase “no training on your data” answers only one question. It is not a synonym for confidentiality, secrecy or GDPR compliance.

It also matters for attribution. The textual output visible to the lawyer may be the final result of retrieval, ranking, model synthesis and tool calls. If the system retrieves an outdated authority, the model inaccurately summarises it and the lawyer adopts the proposition without returning to the source, the relevant failure is not reducible to a hallucination inside the language model. It is a failure across the socio-technical chain.

For the same reason, the professional act must be distinguished from the machine output. An AI-generated draft is not yet a pleading. An AI-generated recommendation is not yet legal advice. A proposed clause is not yet a contract term. The legal act emerges only when a competent actor adopts, communicates, files, signs, executes or otherwise uses the output within a legally relevant process. The central question is therefore how that transition from machine output to legally attributable act should be governed.

5. The Multilayered Normative Architecture of Legal AI

5.1 The AI Act after Regulation (EU) 2026/1744

The AI Act supplies an indispensable but incomplete layer of the analysis. It classifies prohibited practices, high-risk systems, transparency obligations and general-purpose AI responsibilities through a system-oriented regulatory architecture. Legal professional duties operate differently: they attach to actors, relationships, confidential information, procedural positions and professional conduct. It is therefore a category error to infer professional lawfulness solely from the AI Act's risk classification.

Justice-related systems illustrate the point. Annex III, point 8(a), classifies as high-risk certain AI systems intended to be used by or on behalf of judicial authorities to assist in researching and interpreting facts and law and applying law to a concrete set of facts, subject to the Regulation's classification conditions. The high-risk regime contains requirements on risk management, data governance, technical documentation, record-keeping, transparency, human oversight, accuracy and cybersecurity. Articles 12 to 14 are especially relevant to the present argument because they institutionalise traceability, transparency and human oversight at system level.

However, temporal application is essential. Regulation (EU) 2026/1744 amended Article 113 so that Chapter III, Sections 1 to 3, for systems classified as high-risk under Article 6(2) and Annex III, applies from 2 December 2027. As of 26 August 2026, those high-risk obligations are therefore not yet generally applicable to Annex III judicial systems, even though the Regulation's general framework is already partly applicable. A

doctrinal article that simply invokes Article 14 as if all high-risk duties were presently enforceable in the justice context would overstate the law.

The 2026 amendment also changed Article 4. Providers and deployers must take measures to support the development of AI literacy among staff and persons using AI systems on their behalf, taking into account knowledge, experience, training, context and affected persons; they are not required to guarantee a specific individual level of literacy. This matters for professional organisations because literacy is now better understood as an organisational governance obligation than as a promise that each lawyer has achieved a fixed threshold.

Temporal caution: Article 12–14 high-risk obligations are analytically relevant benchmarks for justice AI, but for Annex III systems their general application has been postponed to 2 December 2027 by Regulation (EU) 2026/1744.

5.2 Professional secrecy and the right of defence in Spain

Spanish law provides more immediate constraints. Articles 21 and 22 of Royal Decree 135/2021 define professional secrecy broadly, encompassing facts, communications, data, information, documents and proposals known, issued or received in professional practice. Organic Law 5/2024 strengthens the constitutional and statutory architecture of defence and expressly protects confidential lawyer-client communications and professional secrecy.

The relevance to AI is direct. If privileged or defence-related material is disclosed to a technical environment in circumstances incompatible with secrecy, the problem does not depend on whether the AI Act calls the system high-risk. Nor does the question disappear because a user has pseudonymised names while leaving a fact pattern readily re-identifiable. The professional must assess both the information and the architecture through which the information flows.

The right of defence also changes the significance of errors. Article 24 of the Spanish Constitution protects effective judicial protection, defence, legal assistance, the use of relevant evidence, proceedings with guarantees and the presumption of innocence. Article 47 of the EU Charter guarantees an effective remedy and fair trial; Article 48 protects the presumption of innocence and rights of the defence. Where AI contributes to a legal proposition, evidentiary inference or procedural action capable of affecting those rights, reliability and contestability acquire a constitutional dimension beyond ordinary professional quality control.

5.3 GDPR: processing governance rather than a ban on personal data

The existence of personal data does not create a general prohibition on AI use. The GDPR requires a lawful and accountable processing operation. Depending on the context, the relevant architecture includes Articles 5 and 6, special-category data under Article 9, criminal-conviction and offence data under Article 10, privacy by design under Article 25, processor arrangements under Article 28, security under Article 32, impact assessment under Article 35, and international transfers under Articles 44 to 49.

Article 22 requires particular care where a decision is based solely on automated processing and produces legal or similarly significant effects. Even where Article 22 does not apply, the broader principles of fairness, transparency, purpose limitation, minimisation and accuracy remain relevant. EDPB Opinion 28/2024 is important because it rejects simplistic assumptions about whether AI models necessarily contain personal data or whether legitimate interests can support AI processing; both issues require contextual assessment.

The AEPD's 2026 guidance on agentic AI adds a further governance dimension. Systems capable of pursuing goals through external tools can continuously access and combine personal data, expand context

and create downstream effects. In such settings, governance cannot stop at the input prompt. Permission scopes, tool access, data flows and action logs become part of the legal processing operation.

Table 2. Normative layers relevant to an AI-assisted legal act

Normative layer	Primary object	Illustrative sources	Key question for AI-assisted legal work
AI regulation	System/provider/deployer and intended purpose	AI Act; Reg. (EU) 2026/1744	What regulatory category and lifecycle obligations apply to the system and use?
Data protection	Processing of personal data	GDPR; EDPB; AEPD	Is processing lawful, proportionate, secure and properly governed?
Professional law	Professional actor and client relationship	RD 135/2021; LO 5/2024; professional codes	Are secrecy, competence, independence, diligence and loyalty preserved?
Procedural/fundamental rights	Legal process and affected persons	CE art. 24; EU Charter arts. 47–48; procedural laws	Can the AI-assisted act be challenged, verified and fairly used within proceedings?
Evidence	Sources, authenticity, transformation and proof	LEC arts. 299, 382–384; evidentiary doctrine	Is the evidential chain sufficiently authentic, reliable and reconstructible?
Institutional judicial governance	Judicial use of AI	CGPJ Instruction 2/2026	Does AI remain assistance rather than substitution of adjudicative judgment?

6. The AI-Assisted Legal Act as the Proper Unit of Analysis

An AI-assisted legal act is a professional act in which an AI system materially performs, transforms or influences a cognitive or operational function that contributes to a legally relevant output or action. The definition is intentionally broader than automated decision-making. It covers research and drafting, evidence analysis, strategic recommendations, correspondence, contract formation, procedural filings and judicial assistance.

The concept has four advantages. First, it directs attention to the actor. A lawyer, judge, prosecutor, public authority, arbitrator and in-house counsel occupy different institutional positions. Secondly, it captures the information used, including public law, confidential business information, personal data, privileged communications, criminal data and judicial records. Thirdly, it identifies the delegated function: retrieval, classification, summarisation, synthesis, recommendation, prediction, decision support or execution. Fourthly, it captures legal effect and reversibility.

This final element is crucial. A hallucinated paragraph in a private brainstorming note has a different legal significance from the same paragraph filed in court. A wrong date in an internal chronology may be corrected; an autonomously transmitted procedural submission may trigger deadlines or admissions. The same error therefore requires different controls depending on where it sits in the chain of legal effect.

The AI-assisted legal act also clarifies why professional legality cannot be reduced to provider assurances. Even a technically excellent, secure and compliant system can be used negligently by a professional who fails to verify its legal output. Conversely, a professional may use a general-purpose model safely for low-risk public-source tasks if no protected information enters the system and the output is independently verified before consequential use.

7. Why Human-in-the-Loop Is Too Weak a Legal Standard

Human oversight has become one of the most repeated safeguards in AI governance. Article 14 of the AI Act requires high-risk systems to be designed so they can be effectively overseen by natural persons. The difficulty is that human presence is not equivalent to human control. A nominal reviewer may approve an

output without independently interrogating it, especially where the system is fluent, fast and institutionally endorsed.

The behavioural literature provides reasons for caution. Automation bias can lead humans to over-rely on algorithmic outputs and neglect contradictory information. Green (2022) shows that policies requiring human oversight often assume capabilities that empirical research does not support. Laux and Ruschmeier (2025) analyse the AI Act's express concern with automation bias and the difficulty of translating awareness into effective legal safeguards. Banks (2026), focusing on judicial decision support, similarly questions whether simply placing a judge in the loop guarantees meaningful oversight.

Generative AI adds an interaction problem. Its conversational interface, fluency and apparent responsiveness can produce authority effects that differ from conventional software. Webb (2026) emphasises the importance of epistemic expertise, integrity and practical wisdom; Zinzilieta (2026) warns of epistemic atrophy. These concerns show why the legal safeguard cannot be defined by topology alone — human before machine, human after machine, or human inside a workflow. It must be defined by the quality of the professional cognition that accompanies the use.

This article therefore proposes human-accountable reasoning. The standard asks whether the human actor can genuinely explain and defend the final proposition or action as the actor's own. The relevant questions are whether the actor understood the task delegated to AI, had access to the necessary source material, had sufficient competence and time to challenge the system, independently checked material propositions where risk justified it, and retained actual authority to reject or modify the output.

Human-accountable reasoning is satisfied by substantive cognitive responsibility, not by a final signature or approval click.

8. The Principle of Non-Transferable Professional Judgment

Lawyers routinely delegate operations. Legal research, document review, calculations, due-diligence extraction and factual chronologies have long been distributed among teams and technological systems. The relevant distinction is therefore not between delegation and non-delegation. It is between delegating operations and transferring the professional judgment that the legal order attributes to the responsible actor.

The distinction is visible in judicial activity. A judge can receive assistance in identifying authorities, organising a file or preparing a draft. The judicial function nevertheless requires the judge to decide, to assess the legally material facts and to provide reasons. Spain's Instruction 2/2026 makes that boundary explicit by requiring real, conscious and effective human control, prohibiting AI from autonomously taking judicial decisions or substituting the judge in the assessment of facts and evidence or in interpretation and application of law, and requiring personal, complete and critical review of AI-generated draft decisions.

For lawyers, the constitutional position is different, but an analogous professional logic applies. A legal professional may adopt machine-generated analysis only if the adoption is genuine. If the lawyer cannot explain why the cited authority supports the proposition, cannot identify the factual assumptions on which the conclusion depends, or lacks the capacity to reject the model's recommendation, the lawyer has not merely delegated research; the lawyer has allowed the machine to occupy the place of professional judgment.

The principle of non-transferable professional judgment should therefore be understood functionally. The closer the delegated function is to determining rights, evaluating evidence, choosing legal strategy, making representations to a court or creating legally binding external effects, the stronger the requirement that the responsible human independently own the reasoning.

Principle of Non-Transferable Professional Judgment: A professional may delegate cognitive operations to AI, but cannot transfer the judgment for which law or professional norms make that human actor responsible and recover it merely through formal approval.

9. Legal Epistemic Traceability (LET)

9.1 From technical traceability to legal epistemic traceability

Traceability already appears in AI governance, but usually in a technical or lifecycle sense. Article 12 of the AI Act requires high-risk systems to support automatic recording of events so that system operation can be traced for risk identification, monitoring and other regulatory functions. Technical logs can be indispensable. They do not, however, answer the lawyer's central epistemic question: why is the legal proposition that appears in the final act justified?

Legal reasoning depends on sources and evidence. A proposition that Article X requires Y must be traceable to an authoritative legal source that is in force and relevant. A proposition that a defendant was at location Z must be traceable to admissible or otherwise reliable evidence. Generative systems can sever or obscure those relationships by producing fluent propositions whose internal generation process does not preserve a reliable chain of provenance.

The resulting risk is epistemic discontinuity: the professional output appears justified, but the path from authoritative source or evidence to the final proposition cannot be reconstructed without returning to the machine as if the machine itself were an authority. This is unacceptable in legal contexts where the professional must be able to justify the act independently of the tool.

Legal Epistemic Traceability (LET): the capacity to reconstruct, independently of the AI system itself, the evidential and normative chain through which an AI-assisted proposition becomes attributable to a legally responsible human actor.

9.2 The LET chain

LET can be represented as a sequence of legally relevant transitions rather than as a record of model internals:

AUTHORITATIVE SOURCE / EVIDENCE → RETRIEVAL OR INPUT → AI TRANSFORMATION → HUMAN VERIFICATION → EPISTEMIC ADOPTION → LEGALLY ATTRIBUTABLE ACT

Each transition can fail. Retrieval may omit controlling authority. AI transformation may fabricate, distort or overgeneralise. Human verification may rely on the model's own citations rather than the primary source. Adoption may be formal rather than cognitive. The final act may therefore appear human while remaining epistemically dependent on the machine.

Table 3. LET chain and principal failure modes

Stage	Required property	Typical failure	Legal/professional consequence
Source / evidence	Authority, authenticity, relevance, currency	Repealed statute; inadmissible or manipulated material; wrong version	Invalid premise; evidentiary challenge; negligent advice
Retrieval / input	Completeness and faithful selection	Controlling authority omitted; biased corpus; truncated file	Selection bias; incomplete legal frame
AI transformation	Faithfulness to source and task	Fabrication; distortion; conflation; unsupported inference	Hallucinated or misleading proposition
Human verification	Independent and proportionate checking	Reviewer checks only the AI answer or lacks primary-source access	Automation bias; false assurance
Epistemic adoption	Understanding, rejection capacity, conscious endorsement	Rubber-stamp approval; no ability to explain reasons	Attribution deficit; professional negligence risk
Legal act	Correct communication, filing or execution	Wrong version sent; autonomous action beyond authority	Procedural, contractual or rights-based consequence

9.3 LET is not a demand for impossible model explainability

LET should not be confused with a requirement that lawyers reconstruct the internal mathematical operation of a large language model. That demand would often be impossible, irrelevant and disproportionate. A professional need not explain billions of model parameters to justify a legal submission.

The relevant object is the chain of justification, not the latent state of the model. Wachter, Mittelstadt and Russell's work on counterfactual explanations illustrates a broader point: accountability can sometimes be improved without fully opening the technical black box. In legal practice, the professional must be able to open the normative box. The actor must be able to identify the source, understand the proposition, verify the relation between source and proposition, and explain why the proposition was adopted.

This distinction makes LET compatible with trade secrets and complex models. It also aligns with reviewability approaches that focus on the wider socio-technical process rather than only on model explanation. Cobbe et al. (2021) argue that accountable automated decision-making requires records that allow review of the system and its context. LET narrows that insight to the legal epistemic chain by asking what must be reconstructible for a particular professional proposition to be defended and contested.

10. The Epistemic Adoption Test (EAT): When Does AI Output Become the Lawyer's or Judge's Proposition?

The most difficult attribution problem arises between verification and the final legal act. A machine may generate the words, but law commonly attributes the act to the human who signs, files, advises, decides or communicates. Formal attribution is easy; epistemic attribution is not. The Epistemic Adoption Test is proposed to distinguish genuine professional adoption from formal ratification.

Epistemic Adoption Test (EAT): AI-assisted content should be treated as genuinely adopted by the responsible professional only where five cumulative conditions are satisfied — Awareness, Comprehension, Verification, Rejection Capacity and Reasoned Endorsement.

Table 4. Epistemic Adoption Test

Element	Question	Minimum evidence of satisfaction	Failure indicator
1. Awareness	Does the actor know where AI materially contributed?	Actor can identify the AI-assisted task and material transformations.	Professional cannot distinguish machine-derived analysis from verified source material.
2. Comprehension	Does the actor understand the material proposition and its assumptions?	Actor can explain the rule, factual premise and inferential step in own words.	Actor relies on fluency or model confidence without understanding.
3. Verification	Was the proposition checked independently to a degree proportionate to consequence?	Primary authority, original record or reliable evidence checked where material.	Only the AI output or its generated citations were checked.
4. Rejection capacity	Could the actor realistically reject or modify the output?	Time, competence, authority and source access make disagreement practical.	Workflow or organisational pressure makes approval effectively automatic.
5. Reasoned endorsement	Did the actor consciously adopt the proposition as part of own reasoning?	Final act can be justified without treating the AI as an authority.	Signature or click is the only evidence of human involvement.

The EAT is intentionally cumulative for materially consequential propositions. The verification intensity is proportional rather than absolute. A grammatical reformulation of a public quotation does not require the same checking as an AI-generated assertion that a Supreme Court judgment establishes a dispositive rule or that digital evidence supports a criminal inference.

The test also avoids the fiction that all professional adoption can be inferred from signature. Signature remains legally important, but it should not be confused with cognitive adoption. In sanction cases involving fabricated authorities, the core failure is precisely that a professional representation reached the court without sufficient epistemic adoption. The human formally owned the document while failing to own the proposition in the sense required by professional competence and candour.

EAT also provides a bridge between ethics and procedure. If a party challenges an AI-assisted proposition that materially affected a filing, expert analysis or decision, the relevant inquiry is not simply whether AI was used. It is whether the responsible actor can reconstruct the source and verification path and demonstrate meaningful adoption. In that sense, LET supplies the chain and EAT supplies the attribution threshold.

11. AI, Evidence and the Process of Proof

11.1 Four different roles of AI in evidence

Discussion of “AI evidence” often collapses different problems. At least four categories should be distinguished:

- AI-generated or AI-altered material offered as evidence, including synthetic text, images, audio or video.
- Evidence produced or measured through an AI system, such as classification, detection, scoring or expert-support outputs.
- Ordinary evidence selected, ranked, summarised or transformed by AI during litigation preparation.
- AI used by a decision-maker to evaluate or draw inferences from admitted evidence.

Each category raises different questions of authenticity, reliability, disclosure, expert evidence, contestability and attribution. Chin and Roberts (2026) correctly emphasise that AI affects not only the admissibility of AI-generated material but the broader process of proof, including how evidence is elicited, analysed and presented. Pardo (2026) likewise stresses that legal proof is an explanatory and adversarial process rather than a simple aggregation of information.

11.2 Spanish civil procedure: an existing but incomplete architecture

The Spanish Civil Procedure Act already contains technologically open evidentiary provisions. Article 299(2) recognises means of reproducing words, sound and images and instruments capable of storing, knowing or reproducing words, data, figures and operations relevant to the process. Articles 382 and 384 regulate reproduction and data-storage instruments, allow parties to challenge authenticity and accuracy through expert or instrumental evidence, and require that the opposing parties be placed in a position to know and contest the material. Article 383 emphasises identification and preservation so that relevant materials do not suffer alteration.

These provisions were not designed for generative AI, but they contain three principles that are highly relevant to AI-mediated proof: identification, preservation and adversarial intelligibility. They support an argument that when AI has materially transformed evidential content or generated an inferential product that a party asks the court to rely upon, the opposing party must receive enough information to challenge the authenticity, methodology, transformation and reliability of that product.

It would be premature to claim that every AI output automatically falls within Article 384 or that Spanish evidence law already supplies a complete AI authentication regime. The stronger and more defensible claim is functional: the existing procedural emphasis on equal knowledge, challenge and preservation provides doctrinal support for extending traceability requirements where AI materially mediates evidence.

11.3 Criminal proceedings and the right of defence

The stakes are higher in criminal proceedings because Article 24 of the Spanish Constitution, Article 48 of the EU Charter and Organic Law 5/2024 protect defence rights and the presumption of innocence. If AI materially selects evidence, generates investigative leads, ranks suspects, interprets digital material or assists an expert inference, opacity can impair the defence's ability to test the evidential basis of the accusation.

LET does not imply automatic disclosure of source code or proprietary model weights. It implies that the proposition used against an accused person must be contestable through a reconstructible evidentiary chain. Depending on the system, that may require information about source data, version, purpose, relevant parameters or thresholds, human interventions, error rates, tool outputs and verification procedures. The proportionality of disclosure must be balanced against security, privacy and legitimate confidentiality, but those interests should not convert technologically mediated evidence into an unchallengeable assertion.

This is where epistemic traceability becomes more than professional best practice. In a rights-determining context, it can operate as a procedural condition of meaningful contradiction. If the accused cannot identify the evidential premise, understand the machine's legally material role or challenge the transformation that generated the incriminating proposition, the formal existence of a human prosecutor or judge does not necessarily cure the deficit.

11.4 Authenticity, deepfakes and the inversion of evidentiary presumptions

Generative AI also creates a different problem: synthetic evidence that is visually or aurally persuasive. The traditional evidentiary challenge was often to prove that a digital file had not been altered. Generative systems make fabrication inexpensive and scalable, which can increase both false positives and false negatives. Courts may wrongly accept synthetic material, but they may also become excessively sceptical of authentic recordings — the so-called liar's dividend.

Recent US judicial and evidence-rule work, including proposed amendments addressing AI-generated evidence, illustrates the direction of travel: authenticity and reliability may require greater attention to

provenance, system validity and case-specific performance. These developments are comparative rather than binding in Spain, but they reinforce the core insight that the evidentiary system must increasingly ask not only what a digital object displays, but how it was created, transformed, preserved and verified.

For European proceedings, LET offers a technology-neutral formulation. The central demand is not that every digital object carry a cryptographic pedigree, but that the evidential proposition be capable of adversarial reconstruction to a degree proportionate to its significance.

12. Judicial AI in Spain: From Assistance to Constitutionally Attributable Adjudication

Spain's Instruction 2/2026 of the General Council of the Judiciary is one of the most important operational sources for the present framework. It is already effective and specifically governs the use of AI in jurisdictional activity. It requires real, conscious and effective human control, preserves full judicial responsibility and limits external systems where judicial data or legally sensitive functions are involved.

Its doctrinal significance exceeds information security. The Instruction reflects a theory of adjudicative authorship: certain functions cannot be outsourced without changing who is actually deciding. Autonomous assessment of facts and evidence, autonomous interpretation and application of law, and substitution of judicial decision-making are incompatible with the requirement that judgment remain the act of the judge.

LET and EAT help specify what that principle should mean when AI assistance is substantial. A judge who receives an AI-generated draft should be able to identify the controlling authorities, verify material factual propositions against the record, understand the inferential path, reject the draft where appropriate and provide reasons independently. The fact that the final judgment is signed by a judge is necessary but not conceptually sufficient if the cognitive path remains unreconstructed.

This should not be misread as an argument against judicial AI. Controlled retrieval, summarisation, consistency checks and drafting assistance may improve efficiency and even reduce some human errors. The argument is that judicial legitimacy depends on attributable reasons. AI may assist the production of those reasons; it should not become an unreviewable source of reasons.

13. Agentic AI and the Need for Bounded Automation

Generative assistants principally create content. Agentic systems can act. They may decompose objectives, select tools, query repositories, modify documents, communicate with third parties and execute workflows across systems. This changes the legal risk from incorrect content to incorrect action.

The AEPD's 2026 guidance recognises this change from a data-protection perspective. Hacker and Holweg (2026) similarly argue that agentic AI requires attention to orchestration and autonomous task execution. For legal work, the core issue is delegated authority. A system that drafts a notice but cannot send it poses a different risk from a system authorised to select the recipient, send the notice and update the matter-management system without a human checkpoint.

The concept of bounded automation is proposed to govern that authority. The degree of autonomous action should decrease as legal consequence increases and reversibility decreases. Permissions should therefore be task-specific, revocable and technically enforced. Human approval should be inserted at legally significant transition points — filing, acceptance, waiver, settlement communication, external disclosure, deletion, irreversible data modification or rights-determining decision — rather than after every harmless computational step.

This model is preferable to a generic human-in-the-loop requirement because it focuses oversight where law changes. An agent can autonomously search public authorities for twenty minutes without meaningful legal consequence; the critical boundary is when it transforms that analysis into a representation made to a client, court or counterparty.

Table 5. Bounded-automation model for agentic legal systems

Action class	Illustrative task	Consequence / reversibility	Authority design
Internal, public-source	Search, cluster and summarise public decisions	Low consequence; high reversibility	Automation permitted within approved system; sampled quality control.
Internal, confidential	Summarise privileged file; build chronology	High confidentiality; usually reversible	Secure approved environment; least privilege; access logging; human review proportionate to downstream use.
External but non-binding	Draft client update or negotiation options	Reputational and advisory effect; reversible with cost	Human approval before transmission; verify material legal propositions.
Procedural / legally operative	File pleading; send formal notice; accept term; waive right	Potentially high consequence; limited reversibility	Mandatory human authorisation at action boundary; auditable authority and version control.
Adjudicative / coercive	Determine rights, assess evidence, impose sanction	Fundamental-rights effect; low reversibility	No autonomous substitution; accountable human decision-maker with reconstructible reasons.

14. A Seven-Factor Contextual Legality Test

The preceding analysis can be operationalised through seven variables. The test is not a mathematical compliance score. It is a disciplined sequence for preventing a tool label from substituting for legal analysis.

Table 6. Seven-factor contextual legality test

Factor	Question	Why it matters	Examples of red flags
1. Actor	Who deploys or relies on the system?	Institutional role determines duties, authority and constitutional constraints.	Judicial authority, prosecutor or lawyer uses a tool outside authorised conditions.
2. Information	What information enters or is exposed to the system?	Secrecy, GDPR, criminal data, judicial records and privilege may apply.	Full privileged case file sent to an uncontrolled external service.
3. Function	What cognitive or operational task is delegated?	Retrieval, drafting, recommendation, evidence assessment and execution have different normative weight.	AI chooses litigation strategy or evaluates evidence without meaningful professional review.
4. Legal effect	What changes if the system is wrong?	Consequence and reversibility should determine control intensity.	Error enters a pleading, limitation calculation, waiver, sanction or judicial decision.
5. Technical environment	Where and how does processing occur?	Retention, sub-processing, data transfer, RAG, logging and tool permissions shape risk.	Unknown retention, uncontrolled sub-processors, excessive agent permissions.
6. Human accountability	Can the human genuinely understand, challenge and reject?	Formal presence is not meaningful oversight.	Reviewer lacks source access, competence, time or practical authority to disagree.
7. Epistemic traceability	Can material propositions be reconstructed independently?	Legal accountability and contestability depend on provenance and verification.	No source chain, fabricated citation, unverifiable transformation or unexplained inferential leap.

Professional legality of an AI-assisted legal act = f(Actor, Information, Function, Legal Effect, Technical Environment, Human Accountability, Epistemic Traceability)

15. Applying the Framework: Five Legal Scenarios

15.1 Scenario A — Public-source drafting by a lawyer

A lawyer asks a general-purpose model to produce a first outline on a legal issue using only legislation and judgments that are publicly available. No client information is entered. The lawyer then checks every material authority in the official source and rewrites the analysis before giving advice.

The risk profile is relatively low. The information factor is benign, confidentiality risk is limited and the AI function is auxiliary. The decisive safeguards are verification and attribution. If the lawyer returns to the official authorities, corrects the model's synthesis and can explain the analysis independently, the EAT can be satisfied. The same model would move into a different category if the lawyer copied the output directly into a filing without checking whether the cases exist or support the stated propositions.

15.2 Scenario B — Confidential RAG over a client matter

A law firm deploys an enterprise RAG system over a repository containing privileged correspondence, pleadings and internal strategy. The service is contractually configured not to use client data for model training, operates under a data-processing agreement, uses access controls, documented retention and sub-processor governance, and limits external tool access.

This use cannot be classified simply as “amber because confidential data are involved”. It may be professionally acceptable if secrecy, GDPR and security are properly governed. Yet the architecture changes the verification problem: a RAG answer can faithfully quote the wrong document, retrieve an incomplete subset or generate an inference that the source does not contain. LET therefore requires the lawyer to distinguish retrieved evidence from model synthesis and to verify material propositions against the original file before consequential use.

15.3 Scenario C — Agentic legal research followed by autonomous filing

An agent is authorised to research authorities, draft a procedural submission and file it through the court platform once a confidence threshold is met. A lawyer receives a notification after filing.

The principal problem is not the quality of the draft but the delegated authority. Filing is a legally significant transition point. Even an accurate submission can create procedural effects, representations to the court and professional responsibility. Bounded automation therefore requires human authorisation before filing. Post hoc review is insufficient because the action may be difficult or impossible to reverse without consequence. The workflow fails the EAT at the rejection-capacity stage: the responsible lawyer had no effective opportunity to reject before legal effect occurred.

15.4 Scenario D — Judicial AI draft with human signature

A judge uses an authorised judicial AI system to generate a draft decision based on the electronic record. The judge reads the draft, makes limited stylistic changes and signs it without independently checking several factual inferences or authorities identified by the system.

Formal human-in-the-loop is present, but human-accountable reasoning may not be. Under the logic of Instruction 2/2026, the relevant question is whether review was personal, complete and critical and whether the judicial decision remained the product of the judge's intellectual and volitional activity. If material factual assessment or legal interpretation was effectively accepted from the AI without independent engagement, the use risks crossing the boundary from assistance to substitution. Signature alone does not satisfy epistemic adoption.

15.5 Scenario E — AI-mediated analysis of evidence in criminal proceedings

A prosecution team uses an AI system to analyse a large set of communications and produce a network of allegedly relevant interactions. The system's output identifies several messages as highly probative. The defence receives the selected messages and the final analytical chart, but not sufficient information to understand the selection criteria, relevant model version, excluded material or human validation process.

Here the primary issue is contestability. The AI is not itself the legal decision-maker, yet it structures what humans see and may shape the evidentiary narrative. If the defence cannot test whether the selection omitted exculpatory context or produced false associations, the human decision-makers operate on a machine-curated evidential field that is not meaningfully challengeable. LET supports disclosure of sufficient provenance and process information to reconstruct the material inferential path, subject to proportionate protection of legitimate security and third-party interests.

Table 7. Results of the structured case analysis

Scenario	Primary risk	EAT status	LET requirement	Normative result
A. Public-source drafting	Hallucinated or misapplied authority	Satisfied if primary-source verification and independent reasoning occur	Source → AI synthesis → checked authority → professional adoption	Ordinarily permissible with normal professional controls.
B. Confidential RAG	Secrecy, data governance, incomplete retrieval	Potentially satisfied	Architecture + original file + retrieval provenance + verification	Permissible with enhanced technical and professional safeguards.
C. Autonomous filing	Delegated authority and irreversibility	Fails before filing because no rejection capacity	Action log alone insufficient; human approval required at filing boundary	Presumptively incompatible unless workflow redesigned.
D. Judicial draft	Substitution of adjudicative reasoning	Fails if review is formal rather than substantive	Record + authorities + material AI contribution + judicial verification	Compatible only with complete, critical and independently reasoned judicial adoption.
E. AI evidence analysis	Opaque selection and impaired contradiction	Not solely professional; procedural attribution problem	Sufficient provenance, model role, validation and excluded-context information	Requires enhanced disclosure/contestability proportional to evidentiary significance.

16. Organisational Governance for Law Firms, Legal Departments and Courts

The proposed framework has organisational implications. Individual competence cannot compensate for a law firm's failure to govern systems, data and agent permissions. Conversely, a technically approved system does not relieve the individual professional of verifying consequential legal propositions. Effective governance therefore requires both institutional controls and professional epistemic duties.

The revised Article 4 AI Act reinforces this organisational perspective by requiring providers and deployers to support the development of AI literacy in a context-sensitive way. The Spanish legal profession's 2026 principles similarly emphasise confidentiality, security, transparency, explainability, proportionality, bias prevention and continuous training. Those principles should be converted into operational rules.

- Approved-system inventory: identify permitted systems, providers, deployment modes and prohibited consumer configurations.
- Data-classification matrix: distinguish public, internal, confidential, privileged, special-category, criminal and judicial information.
- Function map: classify AI tasks as clerical, retrieval, summarisation, synthesis, recommendation, prediction, evidence analysis or external execution.

- LET protocol: define which material propositions require primary-source or original-record verification and what provenance must be retained.
- EAT responsibility: assign a named professional responsible for adoption of consequential AI-assisted output.
- Agent permission design: apply least privilege and require human approval for filing, sending, accepting, waiving, deleting or other legally operative actions.
- Version and source control: preserve the final sources actually relied upon, not merely the generated narrative.
- Incident handling: establish procedures for fabricated authorities, confidentiality leakage, model failure, prompt injection and unintended agent actions.
- Client transparency: disclose AI assistance where required by professional rules, client expectations, contractual arrangements or the materiality of the AI role.
- Proportionate record-keeping: preserve enough provenance for review without indiscriminately storing privileged prompts or unnecessary personal data.

This last point is important. Epistemic traceability is not synonymous with maximal logging. Permanent retention of prompts can create confidentiality, minimisation and security risks. The record should be designed around the legal need to reconstruct the material proposition: system and version where relevant, task, source set, responsible reviewer, verification step and final adopted reasoning. In many contexts that is more useful than a verbatim archive of every conversational turn.

17. Responsibility and Liability: Non-Transferable Accountability, Potentially Distributed Liability

Professional accountability and legal liability should be separated analytically. A lawyer who signs and files a pleading normally remains responsible for the representations made in it. Reliance on an AI system does not transform the system into counsel of record. The sanction cases involving fabricated citations are a concrete reminder that professional duties attach to the human actor who places the proposition before the court.

That does not mean the provider is legally irrelevant. A defective system, misleading product representation, insecure processing environment or unlawful data practice may generate contractual, regulatory, data-protection or product-liability consequences. Directive (EU) 2024/2853 modernises product liability and includes software within the concept of product, subject to its material and temporal conditions. The AI Act also distributes lifecycle duties among providers, deployers and other actors.

The appropriate proposition is therefore asymmetric. Professional accountability for the final legal act is ordinarily non-transferable; liability across the technical and organisational chain may nonetheless be concurrent. LET assists allocation by identifying where the chain failed: source selection, retrieval, system transformation, organisational governance, human verification, adoption or external execution.

18. From Explainability to Contestability: A Rule-of-Law Perspective

Legal AI governance should ultimately be evaluated by its effect on the rule of law. A technically explainable system can still produce an outcome that is difficult to challenge. Conversely, a complex model whose internal parameters cannot be fully explained may be used consistently with procedural fairness if the legally material proposition can be reconstructed, tested and contested through independent sources and records.

Citron's concept of technological due process identified long before generative AI that automated systems can undermine notice, hearing and accountability when their operation is difficult to challenge. Cobbe et al.'s

reviewability framework extends this insight to socio-technical systems. The emerging literature on contestability likewise emphasises that explanation is valuable only insofar as it enables meaningful challenge and correction.

LET should therefore be understood as a rule-of-law mechanism rather than merely an audit technique. It supports three capacities: attribution — identifying whose professional proposition or action the law is dealing with; contestability — allowing an affected person to challenge the evidential or normative basis; and reviewability — enabling a court, regulator, professional body or organisation to reconstruct whether the act satisfied the applicable standard.

This triad is particularly important where AI influences public power. The closer a system comes to adjudication, coercion, sanction or rights determination, the less acceptable it is for a material inferential step to be defended by saying that the model produced it.

Rule-of-law triad: ATTRIBUTION + CONTESTABILITY + REVIEWABILITY. Legal Epistemic Traceability provides the connective infrastructure between the three.

19. Discussion: What the Framework Changes

The framework changes the object of legal inquiry. Instead of asking whether ChatGPT, a legal assistant or an agent is “allowed”, the decision-maker asks what AI-assisted legal act is proposed and evaluates its variables. This prevents vendor branding from becoming legal classification.

It changes the meaning of human oversight. Human presence is treated as evidence only when accompanied by cognitive and institutional conditions that make challenge real. This is important because automation bias can coexist with formal review. Human-accountable reasoning and the EAT therefore provide a more demanding standard than a signature or approval step.

It changes the concept of traceability. Technical logs remain important, particularly under the AI Act's high-risk architecture, but legal accountability turns on the chain between source, transformation, verification and adoption. A perfect system log that cannot show why a legal proposition is justified does not satisfy the professional epistemic problem.

It changes agentic governance. The central safeguard is not to require a human after every sub-task, but to design authority boundaries around legally significant transitions. This allows useful automation while preserving professional control where external legal consequences arise.

Finally, it creates a bridge between AI regulation and procedural law. Where AI materially participates in the production, selection or evaluation of evidence, LET can inform questions of disclosure, authentication, expert challenge and adversarial equality. That bridge is especially promising for future research because evidence law supplies mature concepts — relevance, authenticity, reliability, contradiction, burden, sufficiency and preservation — that can discipline otherwise abstract AI-governance debates.

20. Implications for Further Doctoral Research

The framework developed here is suitable for expansion into a broader doctoral research programme. The first line would examine AI-mediated evidence and the point at which lack of traceability should affect admissibility, evidentiary weight, disclosure obligations or the validity of a resulting decision. This requires separate treatment across civil, criminal, administrative and constitutional procedure.

The second line concerns attribution. Future work could test whether epistemic adoption can be recognised as a general legal concept across professional negligence, judicial reasoning, expert evidence and public administration. The central question would be whether formal authorship should remain sufficient when the substantive reasons originate in a machine-mediated process that the human cannot reconstruct.

The third line is empirical. Human-accountable reasoning and the EAT should be tested in controlled legal tasks. Researchers could compare conventional human review, AI-assisted review with source verification, and review using explicit anti-automation-bias protocols. Measurable outcomes would include error detection, citation accuracy, confidence calibration, time and ability to explain reasons independently.

The fourth line concerns technical design. LET could be translated into system requirements for legal RAG and agentic platforms: source locking, provenance display, citation verification, model/version identification, retrieval completeness indicators, action-boundary approval, reversible execution and protected audit trails. The resulting work would connect doctrinal requirements to legal technology design without treating technical standards as substitutes for law.

21. Limitations

This article proposes a doctrinal framework rather than a settled rule of positive law. LET and EAT are analytical constructs and should not be presented as existing statutory obligations unless and until incorporated into legislation, judicial doctrine, professional codes or contractual standards. Their value lies in organising requirements already dispersed across professional responsibility, evidence, procedural rights and AI governance.

The article also concentrates on legal professional and judicial uses in the European Union and Spain. Different jurisdictions allocate professional secrecy, discovery, evidence, legal privilege and judicial responsibilities differently. Comparative validation is therefore necessary before the framework is generalised.

Finally, technical architectures evolve rapidly. Some current reliability problems may diminish, while agentic capabilities will likely increase. The framework is intentionally technology-neutral: it focuses on authority, provenance, verification and legal effect rather than on any particular model family. Its durability will depend on maintaining that functional orientation.

22. Conclusion: Towards Professional Epistemic Governance

The central legal problem created by AI in law is not that machines generate text. It is that machine-generated or machine-transformed propositions can enter professional and judicial acts with an appearance of authority that is disconnected from an independently reconstructible source of legal or evidentiary justification.

Traffic-light guidance remains useful for training, but it cannot resolve that problem. The lawfulness of AI use depends on the AI-assisted legal act: the actor, information, function, technical environment, legal effect, human accountability and traceability. The same model can therefore be harmless, conditionally acceptable or professionally incompatible depending on context.

The article has proposed Legal Epistemic Traceability as the requirement that material AI-assisted propositions remain reconstructible from authoritative or otherwise verifiable sources independently of the AI system itself. It has paired LET with the Epistemic Adoption Test, under which awareness, comprehension, independent verification, rejection capacity and reasoned endorsement determine whether a machine-mediated proposition can genuinely be attributed to the responsible professional.

Those concepts strengthen the weak formula of human-in-the-loop. Human-accountable reasoning asks whether the professional really owns the reasons. Bounded automation asks whether an agent's authority is constrained before legally significant action occurs. Together, they convert AI governance from a question of tool permission into a question of professional epistemic governance.

As AI moves from assistance to agency, the rule-of-law requirement should remain constant: when a legal system attributes a proposition, decision or action to a human professional, that actor must be able to understand, challenge, verify and justify the material path by which the proposition became legally their own.

Final thesis: The future boundary between lawful AI assistance and unlawful or professionally defective substitution will be determined less by whether a human is present than by whether the human can reconstruct and own the legally material reasons and actions.

References

- Agencia Española de Protección de Datos. (2026). *Inteligencia Artificial Agéntica desde la perspectiva de Protección de Datos (v1.2, February 2026)*.
- American Bar Association Standing Committee on Ethics and Professional Responsibility. (2024). *Formal Opinion 512: Generative artificial intelligence tools*.
- Armour, J., Parnham, R., & Sako, M. (2022). *Augmented lawyering*. *University of Illinois Law Review*, 2022(1), 71–138.
- Armour, J., & Sako, M. (2020). *AI-enabled business models in legal services: From traditional law firms to next-generation law companies?* *Journal of Professions and Organization*, 7(1), 27–46. <https://doi.org/10.1093/jpo/joaa001>
- Banks, I. (2026). *Judges-in-the-loop? Judicial involvement in human oversight of high-risk decision support systems under the EU AI Act*. *International Journal of Law and Information Technology*, 34, eaag001. <https://doi.org/10.1093/ijlit/eaag001>
- Blair-Stanek, A., Holzenberger, N., & Van Durme, B. (2023). *Can GPT-3 perform statutory reasoning?* *arXiv:2302.06100*.
- Caserta, S. (2022). *New technologies and law firms—an uneasy relationship: A European perspective*. *Law, Technology and Humans*, 4(2), 183–196. <https://doi.org/10.5204/lthj.2246>
- Chin, J., & Roberts, A. (2026). *AI, the law of evidence, and the process of proof*. SSRN. <https://doi.org/10.2139/ssrn.6820460>
- Chiodo, S. (2020). *The greatest epistemological externalisation: Reflecting on the puzzling direction we are heading to through algorithmic automatisisation*. *AI & Society*, 35, 431–440. <https://doi.org/10.1007/s00146-019-00905-y>
- Choi, J. H., Hickman, K. E., Monahan, A. B., & Schwarcz, D. (2022). *ChatGPT goes to law school*. *Journal of Legal Education*, 71(3), 387–400.
- Citron, D. K. (2008). *Technological due process*. *Washington University Law Review*, 85(6), 1249–1313.
- Cobbe, J., Lee, M. S. A., & Singh, J. (2021). *Reviewable automated decision-making: A framework for accountable algorithmic systems*. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 598–609. <https://doi.org/10.1145/3442188.3445921>
- Consejo de la Abogacía Española, Consejo Asesor sobre Innovación y Abogacía. (2026). *Decálogo de principios para el uso seguro y responsable de sistemas de IA por la abogacía española*.
- Consejo General del Poder Judicial. (2026). *Acuerdo de 28 de enero de 2026 por el que se aprueba la Instrucción 2/2026 sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional*. *Boletín Oficial del Estado*, 27, 14480–14487.
- Council of Bars and Law Societies of Europe. (2025). *CCBE guide on the use of generative AI by lawyers*.
- Council of Bars and Law Societies of Europe. (2026). *CCBE technical guide on the use of AI tools and models by lawyers*.

Council of Europe. (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225).

Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024). Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis*, 16(1), 64–93. <https://doi.org/10.1093/jla/laae003>

Delacroix, S. (2018). Computer systems fit for the legal profession? *Legal Ethics*, 21(2), 119–135. <https://doi.org/10.1080/1460728X.2018.1551702>

European Commission for the Efficiency of Justice. (2025). *Assessment Tool for the Operationalisation of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (CEPEJ(2023)16rev).

European Data Protection Board. (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*.

European Parliament & Council of the European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. *Official Journal of the European Union*, L 119.

European Parliament & Council of the European Union. (2024a). *Directive (EU) 2024/2853 on liability for defective products*. *Official Journal of the European Union*.

European Parliament & Council of the European Union. (2024b). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*.

European Parliament & Council of the European Union. (2026). *Regulation (EU) 2026/1744 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI)*. *Official Journal of the European Union*, L 2026/1744.

Faulconbridge, J. R., Sarwar, A., & Spring, M. (2024). Accommodating machine learning algorithms in professional service firms. *Organization Studies*, 45(7), 1009–1037. <https://doi.org/10.1177/01708406241252930>

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology*, 36, Article 15. <https://doi.org/10.1007/s13347-023-00621-y>

Francesconi, E. (2022). The winter, the summer and the summer dream of artificial intelligence in law. *Artificial Intelligence and Law*, 30, 147–161. <https://doi.org/10.1007/s10506-022-09309-8>

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45, 105681. <https://doi.org/10.1016/j.clsr.2022.105681>

Guha, N., Nyarko, J., Ho, D. E., Ré, C., Chilton, A., Chohlas-Wood, A., et al. (2023). LegalBench: A collaboratively built benchmark for measuring legal reasoning in large language models. *Advances in Neural Information Processing Systems*, 36, 44123–44279.

Hacker, P., & Holweg, M. (2026). *A pragmatic approach to regulating AI agents*. arXiv:2604.22819.

Laux, J. M., & Ruschmeier, H. (2025). Automation bias in the AI Act: On the legal implications of attempting to de-bias human oversight of AI. *European Journal of Risk Regulation*, 16(4), 1519–1534. <https://doi.org/10.1017/err.2025.10033>

Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. *Journal of Empirical Legal Studies*, 22(2), 216–242. <https://doi.org/10.1111/jels.12413>

Martinez, E. (2025). Re-evaluating GPT-4's bar exam performance. *Artificial Intelligence and Law*, 33, 581–604. <https://doi.org/10.1007/s10506-024-09396-9>

McCarthy, L. T. (1977). Reflections on TAXMAN: An experiment in artificial intelligence and legal reasoning. *Harvard Law Review*, 90(5), 837–893.

McCarthy, L. T. (1990). Artificial intelligence and law: How to get there from here. *Ratio Juris*, 3(2), 189–200. <https://doi.org/10.1111/j.1467-9337.1990.tb00057.x>

McGinnis, J. O., & Pearce, R. G. (2014). The great disruption: How machine intelligence will transform the role of lawyers in the delivery of legal services. *Fordham Law Review*, 82, 3041–3066.

Niblett, A., & Yoon, A. (2024). AI and the nature of disagreement. *Philosophical Transactions of the Royal Society A*, 382, 20230162. <https://doi.org/10.1098/rsta.2023.0162>

- Paliwala, A. (2010). A history of legal informatics: An introduction to the special issue. *European Journal of Law and Technology*, 1(1).
- Pardo, M. S. (2026). Legal epistemology and legal proof. *Quaestio Facti*, 10, 187–219. https://doi.org/10.33115/udg_bib/qf.i10.23222
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Pasquale, F. (2020). *New laws of robotics: Defending human expertise in the age of AI*. Belknap Press.
- Peoples, L. F. (2025). Artificial intelligence and legal analysis: Implications for legal education and the profession. SSRN. <https://doi.org/10.2139/ssrn.5123122>
- Pintó Sala, A. (2026, August 26). ¿Dónde puede —y dónde no— utilizarse la inteligencia artificial en la actividad jurídica? *Economist & Jurist*.
- Remus, D., & Levy, F. S. (2017). Can robots be lawyers? *Computers, lawyers, and the practice of law*. *Georgetown Journal of Legal Ethics*, 30, 501–558.
- Schneider, J. (2024). Explainable generative AI (GenXAI): A survey, conceptualization, and research agenda. *Artificial Intelligence Review*, 57, Article 289. <https://doi.org/10.1007/s10462-024-10916-x>
- Schwarcz, D., & Choi, J. H. (2023). AI tools for lawyers: A practical guide. *Minnesota Law Review Headnotes*, 108, 1–39.
- Susskind, R. (1986). Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning. *Modern Law Review*, 49(2), 168–194. <https://doi.org/10.1111/j.1468-2230.1986.tb01683.x>
- Susskind, R., & Susskind, D. (2022). *The future of the professions (Updated ed.)*. Oxford University Press.
- Tan, J., Westermann, H., & Benyekhlef, K. (2023). ChatGPT as an artificial lawyer? *Proceedings of the ICAIL Workshop on Artificial Intelligence for Access to Justice (AI4AJ 2023)*.
- Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy & Technology*, 28, 107–124.
- Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.
- Walters, E. (2019). The model rules of autonomous conduct: Ethical responsibilities of lawyers and artificial intelligence. *Georgia State University Law Review*, 35, 1073–1092.
- Webb, J. (2026). Generative artificial intelligence in legal practice: Epistemic expertise, unethical AI, and the role of virtue. *Minds and Machines*, 36, Article 19. <https://doi.org/10.1007/s11023-026-09771-2>
- Webb, J., & Paterson, J. M. (2025). The evolution of legal knowledge work in an age of brilliant(?) technologies: From robo-lawyer to digital law clerk. *ANU Journal of Law and Technology*, 6(1), 51–84.
- Webley, L., Flood, J., Webb, J., Bartlett, F., Galloway, K., & Tranter, K. (2019). The profession's engagements with lawtech: Narratives and archetypes of future law. *Law, Technology and Humans*, 1(1), 6–26. <https://doi.org/10.5204/lthj.v1.i1.1314>
- Zinzilieta, A. (2026). The Lawyer's Swan Lake: Generative AI, legal epistemology, and the nondelegable duty to think. *Business, Entrepreneurship & Tax Law Review*, 10, 73–104.
- Constitución Española. (1978). *Boletín Oficial del Estado*, 311, 29 December 1978.
- Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil. *Boletín Oficial del Estado*, 7, 8 January 2000.
- Ley Orgánica 5/2024, de 11 de noviembre, del Derecho de Defensa. *Boletín Oficial del Estado*, 275, 13 November 2024.
- Real Decreto 135/2021, de 2 de marzo, por el que se aprueba el Estatuto General de la Abogacía Española. *Boletín Oficial del Estado*, 71, 24 March 2021.
- R (Ayinde) v London Borough of Haringey; Al-Haroun v Qatar National Bank QPSC & Anor [2025] EWHC 1383 (Admin).