ALM | LAW.COM
LAW JOURNAL NEWSLETTERS

Legal AI Is Not a Singular Category ... and Cannot Be Governed As One Solution

Before asking whether AI use is permissible, defensible or reliable, practitioners have to identify what function the system is supporting in the legal workflow. Treating all legal AI as a single category obscures the actual source of professional risk: not the existence of AI itself but the role the system plays in the provision of legal services.

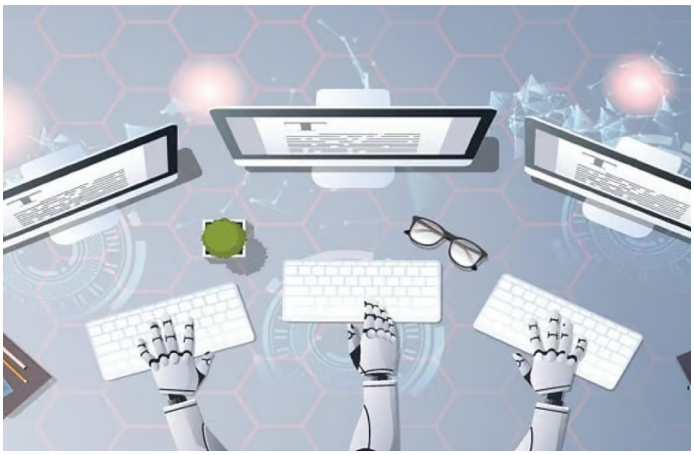
By Esther Birnbaum

June 30, 2026

Legal AI has become shorthand for a broad collection of technologies that do very different things. Some systems support the operational infrastructure around legal services. Others are directly used in the exercise of legal judgment. These are not variants of the same activity. They involve different risks, different supervision obligations and different standards for evaluating reliability. The reality is that they cannot be governed as a singular solution because professional obligations depend on the role the system plays in legal work.

This matters because the legal profession is actively trying to govern AI in its work.

Several federal courts have issued standing orders or guidance addressing the use of generative AI in legal filings. Bar associations and professional organizations have published ethics guidance on competence, confidentiality, supervision and disclosure obligations. *See*, *e.g.*, ABA Standing Comm. on Ethics and Pro. Resp., Formal Op. 512 (July 29, 2024); "Mandatory Certification Regarding Generative Artificial Intelligence," Judge Brantley Starr, N.D. Tex. (May 30, 2023); United States Court of Appeals for the Fifth Circuit,



Court Decision on Proposed Rule (June 12, 2024) (declining to adopt a special rule on use of artificial intelligence in drafting briefs).

The same questions keep surfacing across all these efforts. Who is responsible for supervising the technology? How should outputs be validated? What makes use of the system defensible if challenged later?

The problem is that these questions are being asked at the wrong level. "Legal AI" often gets treated as a single category when, in reality, the tools are performing fundamentally different functions. A billing system and a generative AI model drafting substantive legal analysis may rely

on similar underlying technology, yet the professional obligations they implicate are not the same.

Until the industry begins distinguishing between the functions the AI systems are actually performing, the conversation about governance, defensibility and validation will keep collapsing technologies that should be evaluated under entirely different frameworks.

Legal Operations and the Practice of Law

A useful distinction is between operational work that supports legal services and the practice of law itself.

This is not a new division. The profession separated operational support functions from substantive legal work long ago when assigning responsibility, supervision and liability. ABA Model Rule 5.3 addresses supervision of non-lawyer assistance. Rules 5.1 and 5.2 govern supervision responsibilities among lawyers. The Corporate Legal Operations Consortium (CLOC) draws the same line, defining legal operations as business and technical functions that enable legal service delivery rather than constitute it.

AI systems now operate on both sides of that line.

Operational legal support is the infrastructure around legal practice: matter management, spend tracking, intake routing, information governance, knowledge organization, and the production of structured data used downstream in legal workflows. The people performing this work are often not lawyers, and even when lawyers perform it, they are not necessarily exercising legal judgment. AI tools in this category automate or accelerate the administrative and technical processes that support legal service delivery.

The practice of law, or work requiring the exercise of legal judgment, is different. It involves advising clients, evaluating responsiveness and privilege, conducting legal research, drafting legal analysis, preparing witnesses and rendering opinions. AI systems used in this category of work are not acting independently. They are tools lawyers use while performing substantive legal work, and the lawyers remain responsible for the resulting judgment.

Why the Distinction Matters

The distinction matters because the same underlying model can operate for both functions. A language model extracting dates from contracts into metadata fields is actually doing operational support work. The same model summarizing contractual obligations to support client advice is participating in substantive legal analysis. The technology is identical. The activity is not.

Governance is where the distinction matters most visibly. In operational settings, the applicable framework looks like the long-standing model for supervising vendors and non-lawyer service providers: diligence, contractual controls, auditability and oversight of deliverables. In substantive legal work, the obligations shift. ABA Formal Opinion 512 makes the point directly: lawyers using generative AI remain bound by existing professional obligations relating to competence, confidentiality, communication, supervision and candor toward tribunals.

These are not interchangeable obligations. A generalized “AI governance” policy that does not distinguish between operational support and substantive legal work may require lawyer-level review where vendor management would suffice, while missing the situations that require direct attorney supervision.

Defensibility runs into the same problem. In operational settings, defensibility is procedural. Were workflows applied consistently? Can outputs be audited? Are results reproducible? In substantive legal work, defensibility is judgment oriented. The question is whether a lawyer can stand behind the resulting analysis as professionally reasonable. Conflating these two standards creates mismatches: procedural documentation treated as sufficient where substantive review is required, and exhaustive

substantive review demanded where process validation already addressed the risk.

Validation works the same way. Some AI outputs can be tested against known benchmarks. A clause classifier has a measurable error rate. A metadata extraction system can be evaluated against a validated sample set. These tools can be measured against identifiable ground truth. Other outputs are harder to measure. Classification decisions in document review can still be evaluated statistically through recall, precision and sampling methodologies developed in technology-assisted review workflows and recognized in federal case law and Sedona Conference guidance. *See*, *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012); *Rio Tinto PLC v. Vale S.A.*, 306 F.R.D. 125 (S.D.N.Y. 2015); The Sedona Conference, "Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery," 15 Sedona Conf. J. 217 (2014). Interpretive outputs, such as summaries, chronologies, factual narratives and thematic syntheses, are less amenable to evaluation against a single objective benchmark. In those settings, attorney review becomes the operative safeguard.

The line is not always clear. Some systems move quickly between operational and substantive functions, and broader AI governance frameworks may be appropriate in those settings. But treating all legal AI systems as analytically identical obscures the different professional obligations associated with different uses.

The legal profession already has many of the tools needed to govern these systems responsibly. The problem is not a shortage of rules or methodologies. It is the practice of treating all AI systems as though they raise the same supervisory and validation questions.

What the Distinction Enables

Once the distinction is made, the surrounding governance issues become much easier to handle.

Procurement teams can evaluate technologies according to the function they will perform and apply oversight proportional to the risk. Courts can write standing orders that target specific forms of AI-assisted legal work rather than treating all AI use as a single disclosure or governed category. Bar associations can issue ethics guidance tied to the obligations implicated by a particular use case. In-house legal departments can build acceptable-use policies that distinguish operational deployment from substantive legal work and apply heavier scrutiny where the lawyer's judgment is at stake.

None of this requires a new regulatory structure. The profession already has substantial guidance through the rules of professional conduct, case law on supervision and competence, and validation methodologies developed across years of technology-assisted review practice. The proper exercise is determining what the tool is doing before applying rules to it.

Before asking whether AI use is permissible, defensible or reliable, practitioners have to identify what function the system is supporting in the legal workflow. Treating all legal AI as a single category obscures the actual source of professional risk: not the existence of AI itself but the role the system plays in the provision of legal services.

*****

Esther Birnbaum is the executive vice president of data intelligence at HaystackID, where she leads strategic initiatives integrating advanced AI technologies and data solutions across legal, compliance, and governance workflows.