

AI Isn't the Hard Part: Building Artisanal Intelligence Inside Legal Teams

By Sirisha Gummaregula

May 5, 2026

Across legal departments, the pattern is increasingly familiar: a tool is selected, a pilot runs successfully, and then progress stalls. The technology performs as advertised, but the transformation does not follow.

The gap is not within the tool's capability. It is the operating model around it.

There is a difference between deploying AI in an existing workflow and rethinking how legal work gets done. Deployment produces real gains within an existing model: tasks get faster, and some work shifts to a lower level of seniority. But the organizations seeing more fundamental change are the ones redesigning their operating model around what the technology makes possible. QuisLex calls this Artisanal Intelligence: the disciplined application of human judgment to ensure AI-enabled legal work is complete, consistent, and reliable. It requires designing workflows around how AI is actually used for a specific organization, building quality controls suited to its risk profile, maintaining human judgment, and continuously updating as technology and legal work evolve. Most legal departments are only beginning to do that.

Humans in the Lead

Building that operating model starts with defining human responsibility in an AI-assisted workflow.

The phrase "human in the loop" is often employed to signal responsible use without defining what is

required in practice. It implies AI is the principal actor and humans step in at the end. In high-stakes legal work, that inversion sets up a false sense of control, creating exposure the organization cannot defend.

Lawyers reviewing outputs at the end of a process they did not design from a system whose failure modes they do not understand do not constitute defensible governance. Humans must be in the lead: deciding where AI is used, designing how data flows through the system, defining validation checkpoints, and retaining accountability for outcomes.

What that looks like in practice is straightforward: humans design; AI executes; humans govern.

To do that, lawyers need to understand where probabilistic systems fail and how to design verification around those points. These capabilities sit at the intersection of legal judgment, process design, and systems thinking and cannot be improvised at the moment of failure.

Designing Before Selecting

Before a tool is selected, the more fundamental question is whether the workflow itself should be



Sirisha Gummaregula

redesigned around what AI makes possible. A contract review process should not simply have AI slotted in. It should be redesigned to reflect what AI can analyze at scale and where human judgment adds the most value. Only then can the foundation be properly defined with how matters flow, how decisions are made, what information is assembled, and where accountability sits. Those answers determine whether AI deployment produces trusted and reliable outcomes or merely faster outputs.

Much in-house legal work is business work with legal consequences. The legal dimension adds significance, but the business context determines whether the outcome is right. Contracts document commercial relationships; compliance programs govern business risk. Because of this relationship, platform selection cannot be treated as an isolated technology decision. It is a choice shaped by the specific workflow being redesigned, the matter types it serves, and the risk it must guard against. Some tasks may be better suited to enterprise AI trained in business context, while others require specialized legal tools.

The legal obligations, however, remain unchanged. ABA Formal Opinion 512 confirms that responsibility cannot be delegated to a machine, and courts have reinforced this with sanctions and verification requirements. AI output should be treated like a junior associate's draft with the supervising attorney accountable for the final product. That accountability is crucial because AI failure modes are less visible and occur more quickly. Designing the operating model from the outset to support this oversight is what makes it defensible, not assuming humans being nominally present in the process take care of everything.

Measuring What Actually Matters

AI changes legal work in ways that require a different approach to measuring value. Many organizations still focus on time saved rather than outcomes produced. A contract review taking 2 hours instead of 10 only creates value if it produces work that is as reliable, identifies the same issues, and can be defended.

Cost questions then differ from those most legal departments are asking: Can legal handle greater complexity without proportional cost increases? Are decisions more consistent across similar matters? Are issues identified earlier in the workflow? These require measurement focused on outcomes. Speed without assurance is not efficiency; it is compressed exposure.

In contract negotiations specifically, the question is not only whether the markup is correct but whether the system had access to commercial context, client priorities, and counterparty position before generating it. Work product that is legally accurate but commercially misaligned can increase risk.

Understanding where AI actually breaks down is what makes those outcome measurements meaningful, and it is where the operating model must be strongest.

Failures Nobody Sees Coming

Understanding why human accountability cannot be delegated requires understanding how AI fails in legal workflows.

Hallucinations often take center stage as a problem already sanctioned by courts. But this is not the most dangerous failure. The more serious shortfalls are less visible: they produce outputs that appear correct, pass standard review, and are acted upon until something goes wrong downstream.

Five distinct failure modes are emerging, each requiring its own detection methodology.

Silent omission, the most consequential, typically originates in a context assembly failure, where the system was never working from the right inputs in the first place, causing it to miss provisions that are structurally peripheral or linguistically atypical relative to its instruction framing without signaling that gap. The output appears complete even when it is not.

In a complex MSA, the revenue recognition position is rarely explicit, instead emerging across payment terms, acceptance criteria, and service credits distributed throughout the agreement. An AI system may extract these elements correctly in isolation but still not identify the risk they collectively create. This is not

a hallucination; it is an omission. The system signals neither uncertainty nor the gap.

Boundary failures occur when the system answers the question asked but misses a qualifying provision altering the analysis. *Confident inconsistency* arises when similar queries produce different results across comparable documents. *Context drift* takes place in multistep workflows when the system's understanding shifts over time. Only then do we get to *hallucination* itself: fabricated factual content presented with the same confidence as accurate content.

In practice, these are detected not by reviewing outputs alone, but by testing whether the system has engaged with the full document universe, equivalent provisions are treated consistently across a dataset, and workflow parameters remain stable across execution steps.

Governance programs addressing only hallucination take care of just the last and most easily isolatable of these five risks. The first four have an insidious impact: upstream errors in context assembly propagate down the workflow and appear at review as complete work product. By then, the exposure may already exist.

Effective quality control covers everything from system design, input assembly, permissions, and workflow monitoring before execution to structured validation and review after outputs are produced. The legal departments best positioned to manage this are those treating AI governance not as a compliance exercise but as a sophisticated operational discipline embedded at every stage.

When Judgment Atrophies

As AI handles more execution, attorneys responsible for oversight gradually exercise less of the judgment that makes that oversight meaningful. Supervision becomes a formal step rather than a substantive one. The judgment meant to serve as the backstop does not disappear, but sets in quietly as cognitive rust, until the moment it is needed but no longer reliable.

Building an operating model genuinely accounting for this means treating human capability as something to be actively maintained alongside AI capability, not assumed to persist on its own. No audit trail can capture it, which is precisely what makes it the hardest to govern and the most consequential to design around.

The Hard Part

Legal departments treating AI as a technology decision will see incremental gains. Those treating it as an operating model issue and investing in maintaining the human judgment the model depends on can instead build functions that are more consistent, more scalable, and more defensible.

Artisanal Intelligence is the name for this operating model discipline. The tools are already powerful, but whether their output can be trusted and defended depends entirely on the craft, judgment, and governance built around them. Most organizations have not yet started that work. The ones that do will be harder to displace.

Sirisha Gummaregula is CEO and founder of QuisLex, a leading alternative legal services provider with 22 years of experience in complex legal and compliance work for Global 500 corporations and Am Law 100 and Magic Circle firms. QuisLex holds a U.S. patent for measuring quality in legal workflows, was the first legal service provider to obtain ISO 9001 quality management certification, and has applied Six Sigma methodology to legal services since its founding. QuisLex has filed a provisional patent application covering the methodology underlying its five failure mode framework for legal AI governance.

