



Andrew Haling

PATHWAY LAW FIRM

Litigating the Ai decision

HOW DO YOU PROVE HARM WHEN THE DECISION LIVES INSIDE AN AI AGENT ALLEGED TO BE A “PROPRIETARY BLACK BOX”?

By the end of 2026, worldwide spending on artificial intelligence is forecast to surpass \$2 trillion. (Gartner, Gartner Forecasts Worldwide AI Spending to Grow 47% in 2026 (May 19, 2026).) Adoption has kept pace with the money. In a global survey of nearly 2,000 organizations, 88% reported regularly using AI in at least one business function. (McKinsey & Co., *The State of AI in 2025: Agents, Innovation, and Transformation* (Nov. 5, 2025).) In human resources, 92% of chief human resource officers surveyed anticipated that AI would be further integrated into the workforce. (SHRM, *2026 CHRO Priorities and Perspectives*.) And in a survey of 193 auto insurers by insurance regulators, 88% reported that they use, plan to use, or plan to explore AI or machine-learning models in their operations. (Nat’l Ass’n of Ins. Comm’rs, *Artificial Intelligence* (2025).)

This unprecedented investment and adoption of a new technology will predictably create pressure from business leadership to generate a return on investment and accelerate the adoption of the technology. As plaintiffs’ lawyers, we’ve seen this story before. A massive investment too big to fail, pressure on leadership to implement and turn a profit, and a rush to adoption. But who pays the price? Our current and future clients. Corners will be cut and guidelines will be ignored. The result will be harm to the public and to the individuals we represent, and AI will rapidly become the fulcrum in plaintiff litigation against corporations (and probably even governments). Already some pioneering cases are being litigated.

Gene Lokken was 91 years old when he fell and fractured his leg and shoulder. After surgery, his doctors sent him to a skilled nursing facility for the rehabilitation they said he needed. A few weeks in, the health insurance payments stopped. According to a federal lawsuit brought by his estate, the decision to cut off his post-

acute care did not come from a treating physician who had examined him.

It came, the complaint alleges, from AI, a predictive model called nH Predict that forecast how long a patient “should” need care by comparing an individual to a database of millions of prior patients. Real decisions in life-and-death matters allegedly being determined by AI. This is no longer science fiction, but reality. The smart money is AI will spread throughout organizations in making decisions as pressure mounts to transition AI from proof of concept to wider scale implementation.

Whatever the facts ultimately prove to be in *Estate of Lokken v. UnitedHealth Group* (D.Minn., No. 23-cv-3514), the case is a preview of a question that will soon land on the desk of nearly every plaintiff’s lawyer: what happens when the defendant either denies a client a treatment, a job, an apartment, their rights, or their dignity, or makes a decision on safety that your client bears the consequence of? More urgently for the practitioner: how do you prove it when the decision lives inside an AI system alleged to be a “proprietary black box” the defendant will fight your access to it every step of the way?

This article is about that second question. Across many jurisdictions plaintiffs’ attorneys will increasingly win or lose in AI-related discovery disputes that ultimately will be decided at the motion stage. Framing of the AI issues is critical to win at the motion stage.

The growing trend

In health insurance, *Lokken* is joined by *Barrows v. Humana, Inc.* (W.D.Ky., No. 3:23-cv-654), a similar case challenging the same nH Predict tool on similar theories. The plaintiffs’ theory is essentially that the AI system violates the health-insurance contracts the insurer entered into with beneficiaries, with an end-goal of keeping the premiums and decreasing the expenses. In August 2025, the court

declined to require the plaintiffs to first exhaust Medicare’s administrative-appeals process and allowed their breach of contract, implied-covenant, unjust-enrichment, and fraud claims to proceed, reasoning that those claims were not “inextricably intertwined” with a benefits determination because they challenge the use of AI in place of the individualized human assessment the contract promised. (*Barrows v. Humana, Inc.* (W.D.Ky. Aug. 14, 2025, No. 3:23-cv-654) 2025 U.S. Dist. Lexis [slip op.].)

In *Kisting-Leung v. Cigna Corp.* (E.D. Cal., No. 2:23-cv-1477), plaintiffs challenge a different tool – Px Dx – used to allegedly deny claims in bulk. There, the court allowed a claim under California’s Unfair Competition Law, predicated on Health and Safety Code section 1367.01, subdivision (e), to proceed, holding that the statute fell within ERISA’s savings clause and so was not preempted. (*Kisting-Leung v. Cigna Corp.* (E.D. Cal. Mar. 30, 2025) 780 F.Supp.3d __ [2025 WL 958389].) These three cases are being pursued by the Clarkson Law Firm, whose attorneys include CAALA members.

The wave extends well past insurance. In *Mobley v. Workday, Inc.* (N.D. Cal., No. 3:23-cv-770), the court conditionally certified a nationwide class of job applicants over age 40, who allege that Workday’s AI-driven applicant-screening tools discriminated against them by age. (*Mobley v. Workday, Inc.* (N.D. Cal. May 16, 2025) [order granting preliminary certification].)

At an earlier stage, the court held that when a vendor’s algorithm performs a traditional hiring function – screening, ranking, rejecting – the vendor can be liable as the employer’s agent. This begs an important question in the AI legal landscape: who is responsible for AI’s misconduct?

In *Louis v. SafeRent Solutions, LLC* (D. Mass., No. 1:22-cv-10800), a

tenant-screening algorithm that allegedly disadvantaged Black and Hispanic housing-voucher holders produced a \$2.275 million settlement and an injunction against the challenged scoring feature – one of the first AI-discrimination cases to end in real money and real structural relief.

These are some of the bellwether pioneering cases. But the disputes in them will expand deeper into other aspects of litigation. Insurance, employment, transportation, premises safety, housing, lending, government benefits: the tools differ, the causes of action differ, but the underlying fact pattern is the same. A decision that once required human judgment will be delegated, allegedly, to an AI system. Companies will be attracted to create an AI system that essentially replaces the trained expert who traditionally exercised judgment based on training, experience, and human realities.

Why companies automate

To litigate these cases, you must understand why the defendant built the AI system/agent, because the motive shapes the proof. Every plaintiff's attorney knows that linking wrongdoing to profit reveals the root of the truth for juries to deliver justice. Tell the truth that connects the act to motivation. Automating a decision saves money on three fronts at once, and each has an evidentiary consequence.

First, it saves labor. Human expertise – physicians, underwriters, trained professionals – is expensive. Instead of hiring human experts to conduct hands-on individual assessments, companies will increasingly seek to apply generalized expertise onto individual circumstances.

Second, it suppresses outliers. An AI system trained on averages tends to decide the average way, which means the unusual but truthful circumstance gets swept into the same denial bucket. Again, you've heard this framing before, likely from the habitual defense-oriented biomechanic expert.

Third, this automation system will likely be used knowing there is a percentage of wrongful decisions or denials. Then weigh the cost of paying for the errors in litigation against the overall

savings.

Again, this is the same story plaintiffs' lawyers have fought for years. AI technology is new, yet the motivations and calculations behind it are the same. When denials are cheap to produce and rarely contested, the error volume will climb unless a plaintiff lawyer holds the system accountable.

Another major challenge to AI systems is that defendants will fight to keep the reasoning and architecture of the system hidden. When a human being makes a decision, that human leaves a trail, or at least can be a witness who is deposed and asked questions about the methodology or lack thereof. That human can be asked what they considered, what they assumed, what facts were relied upon, and why. There are credentials to probe and a history to examine. There may be notes, a rationale, a file. Replace that person with an AI system and it will spit out a polished output that can easily be designed to include only information supporting the output rather than challenge or contradict it.

Every day there are articles about lawyers using AI that generates hallucinations, but this phenomenon will infect every industry rushing AI systems into use. When the AI generates the output, the design of the system insulates the defendant because there is no deponent, no curriculum vitae, no evidence trail, no contemporaneous reasoning. The defendant is able to hide what facts and reasoning were included and not included. When a person most knowledgeable responds at deposition with "I don't know" or "I don't recall" to a critical piece of information, a jury will be invited to judge that person's credibility. The AI output covers up this critical lapse because there is no witness on the stand.

The 1.2-second problem: "Human in the loop" as a defense

You should anticipate the defendant's answer that in fact there was a human making the decision. The AI system is merely a tool. There was a professional who made the decision: the human in the loop. The defendant will downplay the power of AI in litigation (while simultaneously promoting it to its investors).

Do not accept the framing at face value; interrogate what the human actually did.

In a 2023 ProPublica and Capitol Forum investigation into Cigna's PxDx system, reporters found that over a two-month period, company doctors denied more than 300,000 claims while spending an average of roughly 1.2 seconds on each. One former Cigna physician described the process bluntly: "We literally click and submit." (Ross & Herman, How Cigna Saves Millions by Having Its Doctors Reject Claims Without Reading Them, ProPublica (Mar. 25, 2023).) That's the sell to corporate leadership and investors – less human labor, less human expertise, one human doing the job of many. The company saves money. Look at the advertisements: the AI company promises to replace the labor force while claiming a human was, technically, in the loop.

There is a body of science that explains why this happens, and it needs robust and vehement argument by plaintiffs' lawyers to the court. Human-factors researchers have documented for decades a pair of related phenomena – *automation bias* (the tendency to over-trust and defer to an automated recommendation) and *automation complacency* (reduced vigilance in monitoring a system perceived as reliable).

An influential 2010 synthesis found these effects appear in novices and experts alike and cannot simply be trained away. (Parasuraman & Manzey, Complacency and Bias in Human Use of Automation: An Attentional Integration (2010) 52 Human Factors 381.) A 2012 systematic review documented the same tendency specifically among clinicians using decision-support tools. (Goddard, Roudsari & Wyatt, Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators (2012) 19 J. Am. Med. Informatics Ass'n 121.)

The lesson for the courtroom is precise: designing a workflow around a one-click approval predictably hollows out intellectual decision-making and judgment. When a defendant asserts a human made the decision, the design of the interface and the metrics imposed on the reviewer become central discovery targets, not background detail. What is the design of

the system meant to do – encourage humans to engage the facts, spend time on the matter, and exercise judgment, or the opposite? Without knowing the case, one can often predict the reality: the return on investment is based on spending less on human capital and encouraging speed and efficiency.

None of this is an argument against artificial intelligence. Used responsibly, these tools can bring consistency, quality, and workflow. They can raise awareness of outliers and highlight areas that need attention.

But when AI is paired with maximizing profit and spending less on human labor, it creates a recipe for danger. A properly built system should deliver the evidence supporting its decisions, flag areas that need more human investigation, and provide an audit trail of the evidence considered and the reasoning applied.

The argument should be framed as this: access to the AI system and its pattern of usage is paramount to determining whether there was a meaningful human in the loop. Proprietary technology may deserve protective orders, but access to the AI system can reveal what evidence was intentionally ignored, downplayed, or suppressed.

The defense case

Defendants will contend that AI systems can process volumes of data more efficiently and support human decision-making. They argue their tools are decision aids that flag cases for professional review rather than decision-makers that replace it, and that a qualified human retains authority to override any output.

On discovery specifically, defendants press concerns that source code and training data are trade secrets of real value, that broad demands impose disproportionate burden, and that a plaintiff's theory should not become a license to reverse-engineer a competitor's technology. (Many AI studies on "hallucination" indicate that the root cause of the error can be the training data itself.) The defendant will emphasize the importance and role of the human clicking the button.

Framing the discovery fight before it starts

You will probably know, or suspect, that AI played a key role in the underlying case. To prove it, you need to understand the AI system. The defendant will resist on three predictable grounds – relevance, burden, and trade secret – and it will try to win that resistance early in discovery to cut off the plaintiff before any meaningful discovery can start. How you frame the issues in your requests, before that motion practice begins, largely determines what you will be allowed to see.

In *Lokken*, plaintiffs moved to compel material about nH Predict. The court granted discovery into how the tool was developed, how it was used, who built it, whether it was designed to supplant physicians, the oversight it received from an internal AI review board, and documents from before the tool's rollout that would show how practices changed once it came online. But the court *denied* production of the model's source code, training data, and underlying rules, reasoning that those internals were not relevant to what remained a breach-of-contract claim.

So, if plaintiffs did not get the source code, then the focus must shift to how the tool was used, to demonstrate that the human role was reduced to a formality. The categories below are a starting checklist that translates across practice areas – personal injury, employment, premises, and beyond. The two questions running through all of them are the same: was the human meaningfully in the loop, and what evidence did the system include, ignore, or suppress? Frame requests for production and deposition topics around these categories rather than around the algorithm itself, and you are asking for material a court is far more likely to order produced.

Discovery categories: Proving the human was – or was not – in the loop Existence and identification of the system.

Establish that an automated tool touched the decision at all, since your client likely was never told.

- The identity, vendor, and version of any software, model, or scoring tool used in the decision, and whether it was built in-house

or licensed.

- Policies, manuals, and workflow diagrams showing where in the process the tool is applied and what it outputs (a score, a recommendation, a denial).
- Whether and how the use of the tool was disclosed to the affected person.

Deployment and the business case.

Motive shapes proof – tie the tool to the promised savings.

- Materials given to leadership, boards, or investors describing the tool's purpose, projected cost savings, labor reduction, or return on investment.
- Contracts and statements of work with the vendor, including performance targets and pricing tied to volume or speed.
- Before-and-after data showing how decision outcomes, approval rates, or staffing changed once the tool came online.

The human role, measured not asserted. Test the "human in the loop" claim against what the human could actually do.

- Time-stamps and metadata for individual decisions, including the average time each reviewer spent per file.
- Productivity metrics, quotas, and scorecards imposed on the humans who reviewed or signed the outputs.
- Compensation, discipline, or termination records tied to reviewer speed, output volume, or deviation from the tool's recommendation.
- Training and instructions given to reviewers about whether, when, and how to override the tool.
- Override and reversal rates – how often humans disagreed with the tool, and what happened on appeal.

What the system included, ignored, or suppressed. Get at the evidence the output was built to leave out.

- The categories of input data the tool did and did not consider, and any inputs it was configured to exclude.

- Audit trails, logs, and edit histories showing what the tool generated, what a human changed, and when.
- Whether the system preserves the underlying reasoning and contrary evidence, or only the final output.
- Validation studies, accuracy testing, error rates, and known-limitation disclosures from the vendor or the company.

Oversight and knowledge of error. Show the company knew the cost of being wrong and used the tool anyway.

- Internal audits, AI-governance or review-board minutes, and risk assessments concerning the tool.
- Complaints, appeals data, regulator inquiries, and internal analyses of wrongful outcomes.
- Cost-benefit or reserve analyses weighing the expense of litigating errors against the savings from automation.

Depose the people, not just the machine. The system has builders and supervisors who can be examined.

- The engineers or vendor personnel who designed and configured the tool.
- The managers who set reviewer quotas and the reviewers themselves.
- A person most knowledgeable on the tool's design, deployment, validation, and audit capabilities.

A practical note on trade-secret objections: treat them as a problem to manage, not a wall. A robust protective order, an attorneys'-eyes-only tier, and, where needed, a special master or neutral expert are the ordinary tools for getting sensitive material produced under controlled conditions rather than not at all. Frame the fight as access to how the system was used and what it suppressed – not as a demand to publish a competitor's code – and the objection loses much of its force.

Conclusion

Although AI may be a new frontier in discovery, the story is old. A company invests heavily, leadership demands a return, and the pressure to cut cost and

move fast lands on our clients. What is different this time is where the evidence goes. When a machine makes the decision, the deponent, the credentials, the notes, and the contemporaneous reasoning can all disappear at once, and the client is often never told a machine decided at all. That is not a reason to fear the technology. Used responsibly, AI can make decisions more consistent and flag the very outliers that deserve a closer human look. The fight is about responsible use, and about accountability when use is not responsible.

For the plaintiff's bar, that fight will be won or lost in discovery, and discovery will be won or lost at the motion stage until precedents are won.



Andrew Haling is a partner at Pathway Law Firm, where he represents plaintiffs in serious personal injury and wrongful-death cases throughout California. He is the lead trial attorney as well as the manager of the litigation department. Additionally, he is focused on AI and its rapid impact on both legal work and our communities.