LAW.COM

Expert Opinion **Artificial Intelligence**

AI and the Law, Part 2: The Capability Trap

In Part 2 of his "AI and the Law" series, David J. Partida discusses how courts are increasingly applying traditional negligence doctrines to AI systems, focusing on duty, breach, causation, and damages in the context of autonomous decision-making.

September 01, 2026 at 10:00 AM By **David J. Partida**

Photo: Production Perig/stock.adobe.com

The negligence framework is not new. What is new is its application to systems that make consequential decisions without human intervention at every stage. Courts are beginning to work through that application by drawing on products liability, professional malpractice, and technology negligence doctrines that provide considerably more guidance than AI's novelty might suggest.

For deployers, the direction is not encouraging.

How the Negligence Doctrine Is Applying

Negligence requires duty, breach, causation, and damages. None of those elements disappear because AI was involved. The analysis changes, however, because of the way AI systems operate.

Duty is unlikely to present much difficulty. Courts have long recognized that organizations deploying technology for commercial purposes owe a duty of care to those foreseeably affected by its outputs.

A company using AI to make decisions affecting customers, applicants, patients, or counterparties occupies much the same position as any other operator of a system capable of causing foreseeable harm.

Breach is where the AI-specific analysis becomes more consequential. Courts evaluating technology-related negligence ask whether the defendant acted as a reasonably prudent organization would have under similar circumstances.

Decisions allowing bias and negligence claims to proceed based on allegations that vendors failed to audit training data for discriminatory patterns suggest that meaningful validation and bias testing are becoming baseline expectations for reasonable deployment. *Mobley v. Workday, Inc.*, No. 23-cv-00770-RFL, 2025 WL 1424347 (N.D. Cal. May 16, 2025) (granting preliminary

collective certification of ADEA age discrimination claim on behalf of putative class of millions of applicants over 40 whose applications were processed by Workday's AI screening system).

Regulators are moving in the same direction, increasingly treating inadequate testing, training, and bias assessment as evidence of unfair or unreasonable practices. *(In re FTC v. Rite Aid Corp., FTC Docket No. 2023190* (Dec. 19, 2023) (finding failure to test accuracy, enforce image quality standards, train employees, or assess demographic disparate impact constituted unreasonable deployment practice under Section 5 of the FTC Act).)

And an aggressive plaintiff's bar is actively targeting employers for their use of AI in connection with termination and layoff decisions. (Dara Kerr, Meta used AI to tag workers who took leave to be laid off, lawsuit claims. (July 14, 2026); *Does 1-26 v. Meta Platforms, Inc.*, No. 4:26-cv-07122-WHO (N.D. Cal. July 13, 2026) (July 14, 2026 Northern District of California lawsuit alleging Meta used AI-driven performance scoring and employee monitoring to select workers on protected leave or with disabilities for layoffs).)

Recent empirical evidence reinforces the point that legal exposure often arises not from intentional discrimination, but from failing to test systems at the level where harms actually occur.

A large-scale academic study of AI hiring systems found that aggregate testing masked substantial position-level racial disparities that became visible only when individual hiring workflows were analyzed separately. Nearly 26% of applications submitted by Black applicants were routed through positions producing outcomes that would trigger EEOC disparate-impact scrutiny.

The study also documented how widespread deployment of a single vendor's model across multiple employers created "systemic rejection" effects that no individual employer may have recognized on its own. (Nick Lichtenberg, *Largest study of AI hiring algorithms to date finds 'clear racial disparities'—over 25% of Black applicants tainted by bias*, Fortune (May 26, 2026), citing "Algorithmic Monocultures in Hiring," forthcoming ACM FAccT Conference paper.)

For negligence purposes, the implications are significant. Deployers may have difficulty arguing that risks were unknowable if reasonable testing, monitoring, or segmentation analysis would have identified them before harm occurred.

It will be critical for all enterprises to see how these issues unfold in cases like the *Meta* case (Complaint*, Does 1-26 v. Meta Platforms, Inc.*, No. 4:26-cv-07122-WHO (N.D. Cal. July 13, 2026) and other similar suits where these nascent issues are just beginning to unfold.

The analysis becomes more complicated where a system's failures stem not from a single defect, but from cascading effects the deployer never evaluated.

An AI system that takes a technically correct action based on incomplete context can trigger infrastructure failures across dependencies it was never designed to model.

Restarting a service to reduce latency or rerouting traffic to manage load may appear entirely reasonable in isolation yet produce catastrophic outcomes when the system lacks visibility into other stressed components.

Courts assessing whether a deployer acted reasonably will ask not only whether the AI was adequately validated before deployment, but whether the deployer understood and accounted for the ways an autonomous system could amplify risk within a complex operating environment.

Causation presents perhaps the most technically novel challenge. AI systems can generate outputs through processes that are difficult to reconstruct or explain.

Courts have encountered analogous problems in products liability cases involving complex manufacturing processes and have developed doctrines—including *res ipsa loquitur* and burden-shifting in appropriate circumstances—that allow plaintiffs to proceed even where the precise mechanism of harm remains uncertain.

It is reasonable to expect courts to draw on those same tools in AI cases, particularly where the deployer controls the relevant information about how the system operates.

The Capability Trap

One of the least appreciated consequences of rapid AI development is that increasingly capable systems may also raise the applicable standard of care. If a demonstrably superior model was commercially available, and a deployer instead chose an inferior one that produced a harmful outcome the superior system likely would have avoided, a plaintiff can credibly argue that the better-performing system establishes the benchmark for reasonable conduct.

Courts applying a "reasonably prudent deployer" standard in a market where validated, higher-performing models exist are likely to consider what those models were capable of doing when evaluating what the defendant should have done.

The issue becomes even more significant when superior systems are not simply more accurate, but also more observable. Once vendors begin offering tools capable of detecting model drift, disparate impact, systemic rejection patterns, or cross-system concentration risks, failing to use those capabilities may itself become evidence of unreasonable deployment. What begins as a competitive advantage has a way of becoming tomorrow's expected control.

The principle is hardly unique to AI. Products liability has long scrutinized manufacturers that fall below the state of the art, and professional malpractice measures conduct against what

competent practitioners in the field actually do. As the AI market matures and best practices become more established, the gap between what leading systems can accomplish and what a particular deployer chose to use is likely to become an increasingly important feature of AI negligence litigation.

Existing Regulatory Duties Already Provide the Framework

For many regulated industries, courts need not wait for new AI-specific doctrine. An AI-driven lending decision that produces disparate impact may violate the Equal Credit and Opportunity Act regardless of whether the AI itself was technically “at fault.” An AI-assisted trading system that fails to achieve best execution may implicate existing fiduciary duties.

A healthcare platform whose AI triage system consistently produces incorrect recommendations may face professional liability under legal standards that treat AI-generated advice the same as professional advice when presented to patients in that manner. In many cases, courts applying these existing frameworks to AI-related disputes will not require new legal doctrine. The principal questions will be factual rather than doctrinal.

The New York framework for negligent misrepresentation, established in *Ultramares Corp. v. Touche* and developed extensively by subsequent decisions, reaches false statements made during commercial relationships where the speaker had reason to know the recipient would rely on them. (*Id.*, 255 N.Y. 170, 174 N.E. 441 (1931) (Cardozo, C.J.).)

A company that uses AI to generate professional advice, product information, or client-facing recommendations, then distributes those outputs without adequate review, occupies much the same position as any other commercial actor that publishes inaccurate information to a relying audience.

The Document That Loses the Case

The most important evidence may not be the model's performance metrics at deployment. It is everything that happened afterward: the validation that was never performed, the performance drift that went undetected, the internal report identifying concerns that received no follow-up, or the governance meeting that was postponed until after the incident.

In the data breach context, the central question was never simply whether a breach occurred. The question was whether the defendant had taken the steps a reasonably prudent organization would have taken based on what it knew at the time.

Plaintiffs’ counsel in AI negligence cases will ask many of the same questions and pursue many of the same categories of evidence through discovery: emails identifying known problems, risk

assessments that were never completed, monitoring reports that were ignored, and governance processes that existed only on paper.

The emerging hiring-algorithm evidence underscores how damaging retrospective discovery can become. Many of the disparities identified in the Stanford-led study became apparent only after researchers disaggregated hiring data at the position level—the very type of analysis plaintiffs will argue deployers should have conducted themselves. Once the capability to identify a particular risk exists, explaining why no one looked for it becomes considerably more difficult.

Many AI cases will ultimately turn on a familiar question. By the time litigation begins, there may already be a technical explanation for what happened. The more difficult question is whether the deployer acted reasonably.

If there is no documented record of validation standards, monitoring protocols, governance decisions, or follow-up when concerns were identified, that absence is likely to become one of the most important exhibits in the case.

Building that record is not simply a technology project. It is also a legal one.

David J. Partida *is a litigator at Wilk Auslander, where his practice includes complex commercial litigation and emerging technology disputes.*

Page printed from: https://www.law.com/newyorklawjournal/2026/09/01/ai-and-the-law-part-ii-the-capability-trap/print/

NOT FOR REPRINT

© 2026 ALM Global, LLC, All Rights Reserved. Request academic re-use from www.copyright.com. All other uses, submit a request to asset-and-logo-licensing@alm.com. For more information visit Asset & Logo Licensing.