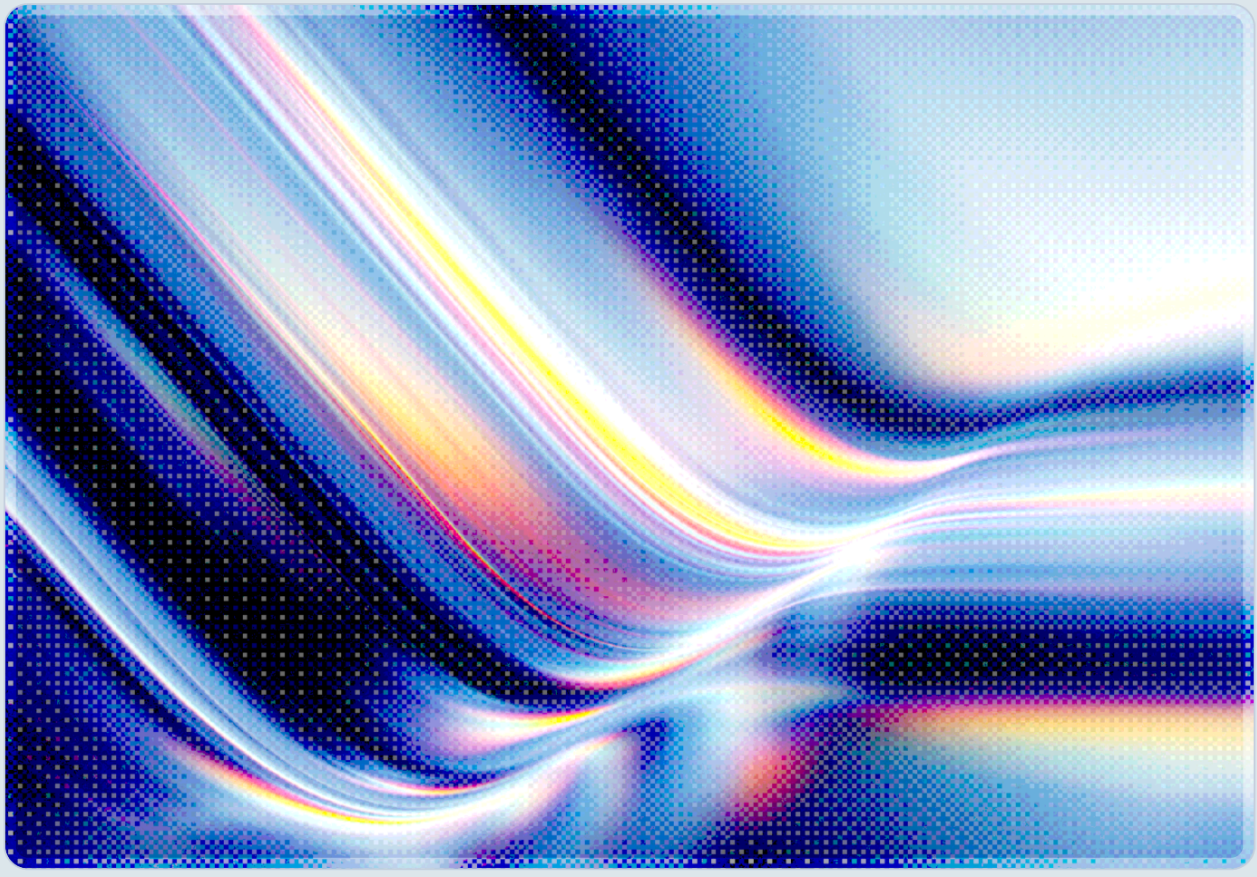
CSAI Foundation | Cloud Security Alliance

Agentic AI Liability: A Live Legal Battleground

Courts, Legislatures, and Insurers Are Still Drawing the Lines

2026-07-03

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

The legal question of who bears responsibility when an autonomous AI agent causes harm remains unresolved in each of the venues examined here: courts are extending product liability and agency doctrines built for human intermediaries onto software that increasingly acts without one, state legislatures are moving faster and less consistently than federal law or contract practice can absorb, and insurers are pulling back coverage precisely as the litigation surface expands. In *Garcia v. Character Technologies*, a federal judge in the Middle District of Florida ruled that a chatbot application can be treated as a "product" subject to strict liability, negligence, and wrongful-death claims, rejecting a First Amendment defense that would have shielded the platform as protected speech [6][1]. In *Mobley v. Workday*, a federal judge in the Northern District of California allowed discrimination claims to proceed against an AI vendor on the theory that its screening software functioned as an "agent" of the employers that deployed it, exposing vendors, not just deployers, to direct liability under federal civil rights law [2]. Neither ruling is final, and both are being watched closely because they represent among the first judicial attempts to fit software that acts autonomously into legal categories, product, agent, employee, that were built around human or human-directed conduct.

Legislatures are responding unevenly and, in places, contradictorily. California's AB 316, effective January 1, 2026, forecloses the defense that an AI system's own autonomy caused the harm, meaning a company cannot escape liability by arguing the software acted on its own [3]. Colorado took the opposite trajectory: after twice delaying its landmark AI Act, the state's legislature passed SB 189 in May 2026, pushing the effective date to January 1, 2027 and stripping out the duty-of-care and impact-assessment obligations that had been central to the original law's ambition, replacing them with a narrower disclosure-and-transparency regime [4]. At the same time, the European Commission's 2025 withdrawal of its proposed AI Liability Directive leaves the European Union without a harmonized fault-based framework, shifting the practical weight onto the revised Product Liability Directive (EU 2024/2853), which member states must transpose by December 9, 2026 and which explicitly brings standalone software and AI systems within strict product liability [16][5]. For security and risk leaders, the immediate implication is that liability exposure for agentic AI deployments now varies by jurisdiction in ways that are actively moving, not settling, and that standard vendor contracts and insurance policies were largely written before this exposure existed.

Background

Litigation and legislation targeting generative AI chatbots emerged first, and agentic AI liability is developing directly on that foundation rather than starting fresh. The wrongful-death suit against Character Technologies, filed in October 2024 after a teenager's suicide following what the complaint described as an intense emotional relationship with a companion chatbot, produced the first major ruling that a mass-market AI application is a "product" for liability purposes rather than a form of protected expression [1][6]. That characterization matters significantly: it opens the door to strict liability and design-defect claims that do not require proving negligence, only that a safer design was feasible. Additional suits against OpenAI have followed a similar pattern: a March 2026 case in the Northern District of Illinois alleges that ChatGPT generated fabricated legal citations that caused economic harm to a business, and a wrongful-death suit filed in California Superior Court in August 2025 alleges the product fostered emotional dependency in a minor and provided harmful instructions [7]. Plaintiffs' firm Edelson PC separately filed seven suits against OpenAI and its CEO in April 2026 spanning negligence, wrongful death, aiding and abetting, product liability, and unfair-competition theories, indicating that the litigation is broadening in both volume and legal theory rather than converging on a single cause of action [8]. These cases target outputs from conversational systems, but in CSA's assessment the same design-defect and failure-to-warn theories are likely to extend to agents that take autonomous action, and arguably with a stronger claim, since an agent that books a purchase, sends a message, or executes code has moved from speech to conduct.

Agency law is the second front, and it opened with Mobley v. Workday. Derek Mobley alleged that Workday's AI-powered applicant-screening tools discriminated against him and similarly situated applicants on the basis of race, age, and disability, and that some of his applications were rejected within an hour of submission, a turnaround the complaint argued was too fast for any human reviewer to have been involved [2]. The court dismissed the theory that Workday operated as an "employment agency" but allowed the case to proceed on the theory that Workday acted as an agent of the employers using its software, reasoning that the tool performed a traditional hiring function, screening and recommending candidates, that federal anti-discrimination law does not permit an employer to delegate away [2][9]. Workday's own SEC filings and court submissions disclosed that roughly 1.1 billion applications had been rejected using its software during the relevant period, giving the case large scale for what is still, procedurally, an early-stage ruling [17]. The significance for agentic AI generally is that the court's reasoning does not turn on chatbot-style conversational output at all; it turns on whether software is functionally standing in for a human decision-maker, which is the defining feature of an agentic system by design.

A third and less visible front is running through contract and insurance markets rather than courtrooms. Legal advisories aimed at manufacturers and enterprises deploying agentic systems in operational roles, such as autonomous supply-chain and inventory decisions, warn that standard AI vendor agreements were negotiated for advisory tools that surface recommendations to a human, not for systems authorized to act, and that liability caps tied to fees paid can be dwarfed by the cost of a single autonomous decision gone wrong [10]. Insurers are responding by narrowing rather than expanding coverage: the Insurance Services Office introduced a generative AI exclusion for commercial general liability policies in January 2026, and carriers including Chubb, Travelers, and Berkshire Hathaway have received state regulatory approval, in more than 80 percent of filings reviewed, to add explicit AI exclusions to general liability, directors and officers, and errors-and-omissions policies [18][11]. The practical effect is that the entities most exposed to the emerging case law, the AI vendors and enterprise deployers named in it, are simultaneously losing access to the "silent AI" coverage that many had relied on when no policy language addressed AI risk directly.

Security Analysis

Courts Are Applying Old Categories to New Behavior, With Inconsistent Results

The Character.AI and Workday rulings share a structural feature: both courts declined to treat the AI system's autonomy as a reason to exempt it from liability frameworks designed for human or human-directed actors. That instinct is consistent with California's AB 316, which now makes that judicial reasoning statutory by explicitly barring an "the AI did it autonomously" defense across the entire supply chain, developer, fine-tuner, integrator, and deploying enterprise alike [3]. But, in CSA's assessment, the underlying doctrinal fit remains imperfect. Product liability law asks whether a design was unreasonably dangerous or whether a warning was adequate, questions built around a static, mass-produced object; an agentic system that adapts its behavior based on context, tool access, and prior interactions does not sit comfortably inside that frame, and scholars tracking the litigation have argued that multi-agent systems in particular are outpacing liability frameworks that were stretched to fit even single-agent tools [12]. Agency law, the theory that succeeded in Mobley, fits somewhat better conceptually because it already contemplates a principal being bound by an agent's authorized actions, but it was built around agents with legal personhood who can be deposed, cross-examined, and held individually accountable, none of which applies to a software agent. Security teams evaluating legal exposure should not assume that a favorable characterization in one jurisdiction, product versus agent versus tool, will transfer to another; the cases above reached their results using different doctrines because the facts, and the judges, differed, not because the law has converged on a single answer.

State and Regional Legislation Is Diverging, Not Converging

The gap between California's AB 316 and Colorado's SB 189 illustrates that state legislatures are not moving toward a common standard. California has hardened the rule against deployers in the strongest terms yet enacted, closing the autonomy defense entirely while preserving ordinary defenses based on causation, foreseeability, and comparative fault [3]. Colorado, by contrast, retreated from its own more ambitious framework: the original Colorado AI Act would have required developers of high-risk systems to exercise reasonable care against algorithmic discrimination and would have imposed active risk-management and impact-assessment duties on deployers, but SB 189 stripped those provisions out in favor of a narrower transparency-and-disclosure regime and pushed the effective date to January 1, 2027, following what commentators described as pressure that included a White House callout of the state's original approach [4]. The European Union's posture compounds the uncertainty rather than resolving it: having withdrawn the AI Liability Directive in February 2025 for lack of member-state agreement, the EU now relies on the revised Product Liability Directive, which brings AI and standalone software within strict liability for defective products but does not address the fault-based, discovery-asymmetry problems, such as a plaintiff's inability to see inside an opaque model, that the withdrawn directive had been designed to solve [5][13][16]. Organizations operating agentic AI across US states and into the EU face materially different liability exposure and evidentiary burdens depending on where an incident occurs and where the affected party can bring suit, with no indication that this variance will narrow before more state legislative sessions and the EU's December 2026 transposition deadline pass.

Contractual Risk Allocation Has Not Caught Up to Autonomous Action

Enterprise legal teams reviewing existing AI vendor agreements are finding that the contracts were written for a different risk profile than the one agentic deployments now create. Liability caps tied to subscription fees, exclusions for consequential damages, and indemnification language calibrated to advisory tools that a human reviews before acting all assume a decision point that agentic systems, by design, remove or compress [10]. Advisories aimed at manufacturers deploying agentic systems in supply-chain and inventory roles recommend renegotiating around explicit authority limits with mandatory human-escalation thresholds, override and kill-switch protocols with defined vendor accountability, contractual allocation of responsibility for data quality feeding the agent's decisions, liability structures that reflect the scale of autonomous risk rather than fees paid, and audit-trail requirements sufficient to reconstruct why a given autonomous decision was made [10]. That last point, auditability, is doing double duty: it is both a governance control and, in CSA's assessment, likely to become the evidentiary foundation a defendant would want in hand to demonstrate reasonable oversight if a regulator or plaintiff later challenges a specific agent action, though no case cited in this

note has yet turned specifically on audit-trail evidence. Organizations that cannot reconstruct an agent's decision chain after an incident are, in CSA's assessment, likely to be litigating from a weaker position regardless of which liability theory ultimately applies.

Insurance Markets Are Pricing In Uncertainty by Excluding It

The insurance industry's response has been to narrow coverage rather than price the new risk, which leaves a gap precisely where the litigation described above is expanding. Commercial general liability, directors and officers, and errors-and-omissions policies are being amended with explicit generative and agentic AI exclusions covering bodily injury, property damage, defamatory AI-generated content, intellectual property infringement, and physical damages traced to AI-driven decisions, with state regulators approving the large majority of insurer filings seeking these exclusions [18][11]. This shift from ambiguous "silent AI" coverage, where AI-related losses were paid or denied under policy language that never mentioned AI, to explicit exclusion means that enterprises deploying agentic systems in operational roles, exactly the deployments most likely to generate the losses described in the Foley & Lardner and similar advisories, may find themselves self-insuring for agentic AI risk by default, whether or not that was a deliberate risk-management decision.

Recommendations

Immediate Actions

Organizations deploying or procuring agentic AI systems should inventory every agent with authority to take autonomous action, whether financial, operational, or customer-facing, and map each one against the jurisdictions in which affected parties could plausibly bring a claim, since AB 316, the Colorado framework, and EU product-liability rules impose materially different standards. Legal and procurement teams should review existing AI vendor contracts specifically for liability caps and consequential-damage exclusions written before the vendor's tools gained autonomous capability, and should request current certificates of insurance to confirm whether general liability or errors-and-omissions coverage now carries an AI exclusion. Security teams should confirm that every agentic system in production logs a reconstructable audit trail of inputs, tool calls, and outputs sufficient to support a reasonable-oversight defense after an incident.

Short-Term Mitigations

Enterprises should renegotiate vendor agreements for agentic tools to include explicit authority limits and mandatory human-escalation thresholds for high-consequence actions, override and kill-switch mechanisms with clearly assigned vendor and deployer responsibilities, and liability terms scaled to the potential impact of autonomous decisions rather than fees paid. Risk and legal teams should engage insurance brokers to assess whether AI-specific liability or technology errors-and-omissions products, now emerging as carriers formalize AI exclusions in standard policies, are needed to close coverage gaps, rather than assuming existing cyber or general liability coverage responds. Compliance functions should track the Colorado AI Act's transition to its January 2027 effective date and the EU Product Liability Directive's December 2026 transposition deadline as concrete dates against which current agentic deployments should be reassessed.

Strategic Considerations

Boards and executive leadership should treat agentic AI liability as an unsettled and actively litigated area rather than a solved compliance checkbox, since the doctrines being applied, product liability, agency law, and statutory autonomy bars, are being tested in real time and could shift meaningfully with the next appellate ruling in Garcia, Mobley, or the OpenAI-related suits. In CSA's assessment, organizations should build governance processes, including MAESTRO-aligned threat and risk modeling for agentic architectures, that produce a documented, auditable decision trail; while no cited ruling to date has directly addressed a framework of this kind, such documentation would likely support a reasonable-oversight defense as courts and regulators continue to define what that standard requires. Finally, given the divergence between California's hard line against the autonomy defense and Colorado's retreat toward disclosure-only obligations, organizations operating nationally should design internal governance to the strictest applicable standard rather than maintaining jurisdiction-specific compliance programs that are likely to require frequent revision as more states legislate.

CSA Resource Alignment

CSA's MAESTRO framework provides the threat-modeling foundation organizations need to document the layered risks, from foundation model through orchestration and deployment infrastructure, that agentic systems introduce [14]. In CSA's assessment, the documentation MAESTRO produces would likely be useful as evidence of reasonable-oversight practices, though no cited court opinion, regulatory guidance, or plaintiff filing in this note has yet directly engaged with MAESTRO, AICM, or STAR for AI as a benchmark for that standard. CSA's AI Controls Matrix extends the Cloud Controls Matrix with AI-

specific control domains and a shared-responsibility model that maps onto the contractual risk-allocation questions raised above, particularly for organizations negotiating which party, model provider, integrator, or deploying enterprise, bears responsibility for a given control failure. CSA's STAR for AI program offers a mechanism for organizations to demonstrate, through independent assessment, an active risk-management posture, one that CSA expects will become increasingly relevant as courts and regulators continue to define the line between reasonable and negligent deployment, even though that line has not yet been drawn by reference to any specific assessment framework. CSA's prior research note on the AI agent governance framework gap describes the broader maturity shortfall, organizations governing agentic deployments with controls designed for simpler copilot tools, that underlies much of the liability exposure discussed in this note [15].

References

[1] Tech Justice Law Project. "Big Win in Our Character.AI Lawsuit: TJLP Statement on the Motion to Dismiss Decision." Tech Justice Law Project, May 21, 2025.

[2] Seyfarth Shaw LLP. "Mobley v. Workday: Court Holds AI Service Providers Could Be Directly Liable for Employment Discrimination Under "Agent" Theory." Seyfarth Shaw, 2024.

[3] Baker Botts LLP. "California Eliminates the "Autonomous AI" Defense: What AB 316 Means for AI Deployers." Baker Botts, January 2026.

[4] Seyfarth Shaw LLP. "Artificial Intelligence Legal Roundup: Colorado Postpones Implementation of AI Law as California Finalizes New Employment Discrimination Regulations." Seyfarth Shaw, 2026.

[5] IAPP. "European Commission withdraws AI Liability Directive from consideration." International Association of Privacy Professionals, 2025.

[6] Law360. "Character.AI Case Highlights Agentic AI Liability Questions." Law360, 2026.

[7] K&L Gates LLP. "AI Product Liability: The Next Wave of Litigation." K&L Gates, March 27, 2026.

[8] Edelson PC. "AI Lawsuits: The Cases Edelson Has Filed and Why They Matter." Edelson PC, 2026.

[9] CIO. "Workday's AI recruitment tool could be liable for discrimination." CIO, 2025.

[10] Foley & Lardner LLP. "Agentic AI Liability in Autonomous Supply Chain Decisions: Identifying and Preventing Legal Risks." Foley & Lardner, May 2026.

[11] Fenwick & West LLP. "The End of 'Silent AI'? Emerging AI Exclusions, Coverage Fragmentation, and Practical Implications for Policyholders." Fenwick, 2026.

[12] Berkeley Technology Law Journal. "Multi-Agent AI is Outpacing the Liability Frameworks Built for Single-Agent Systems." BTLJ, June 2026.

[13] Baker McKenzie. "Proposal for AI Liability Directive Withdrawn." Baker McKenzie, 2025.

[14] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." CSA, February 6, 2025.

[15] Cloud Security Alliance. "The AI Agent Governance Gap: What CISOs Need Now." CSA Labs, April 3, 2026.

[16] Gibson Dunn. "EU Product Liability Directive: Responding to Software, AI and Complex Supply Chains." Gibson Dunn, March 23, 2026.

[17] AIHR Institute. "Mobley v. Workday: 1.1 Billion Rejected Applications Put AI Hiring Under the Microscope." AIHR Institute, May 22, 2026.

[18] Insurance Intel. "Berkshire, Chubb, and Travelers Are Removing AI Coverage." Insurance Intel, April 27, 2026.