INTELLIGENCE • EVIDENCE • IMPACT

AI Does Not Remove Liability. It Moves It.

What the 2026 insurance exclusions reveal about the real cost of replacing human judgment

Sean Mahon, OSC
Founder and Principal, The Collective Risk Intelligence LLC
August 2026

Disclosure

I run a firm that sells analyst-led, human-validated intelligence work. The argument below concludes that human validation is becoming harder to remove from professional delivery than most business cases assume. A reader is entitled to weigh that interest against the argument. I have therefore built the case on documents anyone can check: insurance forms, an enforcement action, a federal docket, and published research. Where I state an opinion, I label it.

The claim being sold

The business case for replacing professional labor with AI systems is usually presented as a change in cost architecture. Human teams carry salaries, benefits, payroll taxes, recruiting, training, management overhead, paid leave, and severance exposure. Those costs are largely fixed. They persist through slow quarters and they do not scale down on demand.

AI systems are presented as the opposite: licensing, model usage, and infrastructure that expand and contract with the work. Fixed payroll becomes variable consumption. When demand falls, spend falls with it.

That framing is not wrong about the mechanics. It is incomplete about what else moves.

What actually transfers

Replacing a person with a system does not delete the liability attached to the work. It relocates that liability from one category into another, and the two categories are not equivalent.

Employment liability is mature. Wrongful termination, discrimination, wage and hour, and severance exposure are governed by decades of statute and case law, are routinely underwritten, and are priced with reasonable confidence. An organization carrying a human workforce knows roughly what its employment risk costs, because a market has been pricing that risk for a long time.

Liability for autonomous output is none of those things. It is novel, unsettled, and being actively repriced. And unlike employment risk, it is currently being written out of the policies that professional firms already hold.

> *This is not risk elimination. It is risk transfer into a market that is presently declining to accept the transfer.*

The market has already responded, in writing

The clearest evidence for this is not commentary. It is policy language filed with regulators.

The Insurance Services Office (ISO), the Verisk unit that drafts the standardized policy forms most US commercial insurers use as base language, introduced three AI-related endorsements for optional use in commercial general liability policies. Form CG 40 47 excludes coverage under Coverage A and Coverage B for bodily injury, property damage, or personal and advertising injury arising out of generative artificial intelligence. Form CG 40 48 excludes Coverage B for personal and advertising injury on the same basis. Form CG 35 08 applies the exclusion to the Products and Completed Operations Liability Coverage Part (Konkel & Ayaz, 2026).

Individual carriers have gone further. Berkley Insurance Company issues an endorsement titled Artificial Intelligence Exclusion (Absolute), form PC 51380 00, dated June 2024. It amends three coverage parts: directors and officers, employment practices liability, and fiduciary liability. The endorsement bars payment for loss on account of any claim based upon, arising out of, or attributable to any actual or alleged use, deployment, or development of artificial intelligence by any person or entity (Berkley Insurance Company, 2024).

The enumerated subparts are broader than the headline. The exclusion reaches an insured's failure to identify or detect content created through a third party's use of AI; inadequate or deficient policies, practices, procedures, or training relating to AI, or the failure to develop or implement any; breach of any duty regarding the creation, use, detection, or containment of AI; any product or service incorporating AI; representations allegedly made by a chatbot or virtual customer service agent; statements concerning the company's AI capabilities or business plans; violation of any law regulating AI; and any demand or regulatory requirement that the company investigate, assess, monitor, or respond to the risks of AI (Berkley Insurance Company, 2024).

Read together, those subparts mean the governance program a firm builds to manage the exposure sits inside the exclusion alongside the exposure itself. Deploying AI without controls is excluded. Deploying it with controls that prove deficient is also excluded. So is the cost of answering a regulator who asks what the controls are.

The definition carries as much weight as the operative language. The form defines artificial intelligence as any machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments (Berkley Insurance Company, 2024). Read literally, that reaches machine assistance firms have relied on for years, well before generative tools arrived. Hamilton Insurance Group has adopted narrower but still sweeping language directed at generative systems specifically (Konkel & Ayaz, 2026).

One detail deserves emphasis, because it is frequently lost in summary. Coverage counsel commentary has described this endorsement as applying to directors and officers, errors and omissions, and fiduciary policies. The form itself amends employment practices liability, not errors and omissions. That distinction

is not cosmetic. Employment practices liability is the coverage line that responds to hiring discrimination claims, which is the precise category of harm that automated screening produces at scale.

Underwriters are not moralists. They are the parties whose entire business is estimating what a category of error will cost. When they decline to carry a risk, or carve it out of forms they were previously willing to write, that is a priced judgment about expected loss.

Why the error profile is different

The distinction underwriters appear to be reacting to is not that machines make mistakes and people do not. It is how the mistakes propagate.

Human error is typically individual and bounded. One analyst misreads one document on one engagement. The error is discoverable, attributable, and contained by the fact that the next file is handled by a different person on a different day.

A configured system does not vary. It applies the same rule to every record it touches, at volume, without fatigue, until someone notices the pattern. The error is not repeated. It is instantiated.

There is a documented case of exactly this, and it involves no vendor and no model. In 2023, iTutorGroup settled an enforcement action brought by the Equal Employment Opportunity Commission for 365,000 dollars. The agency alleged the employer had programmed its online application software to automatically reject female applicants aged 55 or older and male applicants aged 60 or older, rejecting more than 200 qualified applicants (U.S. Equal Employment Opportunity Commission, 2023). No default setting produced that result. A person defined a criterion, the criterion entered a system, and the system applied it without exception.

The contemporary version is on a federal docket, and a court has now drawn the liability line at almost exactly the point this argument turns on. In Mobley v. Workday, Inc., the Northern District of California granted preliminary certification of a nationwide collective under the Age Discrimination in Employment Act, covering applicants aged 40 and over whose applications were scored, sorted, ranked, or screened by Workday's AI recommendation system and who were denied an employment recommendation (Mobley v. Workday, Inc., 2025).

One passage in that order is worth reading closely. The court distinguished between a system that participates in the hiring decision and one that merely executes criteria the employer selected. Had the screening applied only criteria specifically chosen by the employer, the court reasoned, Workday would be no different from an Excel spreadsheet operated by the employer that sorted candidates by age, and would not be liable on an agency theory. The claim survived because Workday's software was alleged not to be implementing employer criteria in a rote way, but participating in the decision-making process by recommending some candidates and rejecting others (Mobley v. Workday, Inc., 2025).

The scale figures in the same order illustrate propagation better than any hypothetical could. Workday represented that 1.1 billion applications were rejected through its platform during the relevant period, and argued that notice could reach hundreds of millions of potential plaintiffs. The court held that alleged breadth of conduct is not a basis for withholding notice (Mobley v. Workday, Inc., 2025). The case is ongoing and the certification is preliminary; Workday may move to decertify after discovery, and the court noted that the second stage will take a more exacting look at the record.

I examined the same question from the operational side in an assessment of automated hiring systems, which found that across every platform reviewed, the criteria, thresholds, and workflow rules driving automated disposition are configured by the employer or its authorized users rather than imposed by the vendor as a fixed default (The Collective Risk Intelligence, 2026).

That convergence is what connects the two halves of this argument. The platform is not the author of the decision. It is the enforcement mechanism for a decision someone else made. Liability follows authorship, and authorship sits with whoever wrote the criteria.

The condition attached to the coverage that survives

This is the part of the picture that changes the economics, and it has been almost entirely absent from the public conversation.

Coverage for AI-related professional error has not disappeared uniformly. What has emerged instead is coverage that is conditioned. Where professional indemnity and errors and omissions policies still respond to a claim involving AI output, the surviving cover generally depends on a human having reviewed and approved that output before it reached the client.

> *The human in the delivery chain is no longer only a quality measure. It is becoming a term of insurability.*

That reframes the entire cost comparison. Most AI business cases treat human review as discretionary overhead, a friction cost to be minimized as confidence in the system grows. If review is instead a condition of coverage, it is not discretionary at all. It is a fixed input to remaining insurable, and the savings model that assumed it away was never accurate.

This is my reading rather than a settled legal conclusion, and it should be tested against the specific policy on the wall rather than against a general description of the market. But the direction of travel is documented, and any firm deploying autonomous systems into client deliverables should be reading its own renewal language rather than inferring the position from an article.

The counterweight

An honest treatment has to state where this argument is weaker than it first appears.

- The market is bifurcating rather than uniformly retreating. Several carriers now write affirmative AI coverage, and at least one has publicly stated it plans no AI exclusion at all.
- Exclusions are construed narrowly and against the insurer. Courts have rejected attempts to treat phrases such as arising out of as automatic bars to any claim with a remote connection to excluded conduct (Konkel & Ayaz, 2026).
- Mixed claims may preserve the duty to defend. Where an action alleges both AI-related conduct and ordinary wrongful acts, an exclusion may not eliminate the insurer's defense obligation (Konkel & Ayaz, 2026).

- An exclusion may not be read so broadly that it swallows the coverage. For businesses deeply integrated with AI, an exclusion eliminating cover for the primary line of business may conflict with the reasonable expectations of the insured (Konkel & Ayaz, 2026).
- Earlier policies may respond. Occurrence-based programs predating these endorsements generally contain no AI limitation, and the later addition of an exclusion may itself support the argument that earlier forms were broader (Konkel & Ayaz, 2026).

None of that reverses the underlying point. It means the exposure is contested rather than settled, which is precisely the condition under which an organization should be least confident that it has transferred the risk away.

The savings assumption is fragile on its own terms

Set liability aside for a moment. The premise that AI converts labor into a lower and more controllable cost is itself under pressure.

A systematic study of token consumption across eight frontier models on SWE-bench Verified found that agentic tasks consume roughly 1,000 times more tokens than code reasoning or chat interactions, driven primarily by input rather than output as context accumulates across steps. Consumption was highly variable, with runs on the same task differing by as much as 30 times. Higher token usage did not translate into higher accuracy; accuracy tended to peak at intermediate cost and saturate beyond it. And the models systematically underestimated their own token costs, with weak to moderate predictive correlation (Bai et al., 2026).

The last finding is the operationally significant one. A cost that varies thirtyfold on identical work, does not correlate with quality, and cannot be forecast reliably by the system incurring it is variable, but it is not controllable. Those are different properties, and business cases routinely conflate them.

Practitioner data points the same direction. The FinOps Foundation's State of FinOps 2026, drawn from 1,192 practitioners responsible for more than 83 billion dollars in annual cloud spend, found that the share of teams managing AI spend rose from 31 percent two years earlier to 98 percent, and that AI cost management is now the single most sought-after skill practitioners name (FinOps Foundation, 2026). A discipline built to govern reserved instances is being asked to forecast consumption for workloads younger than a fiscal year.

One caution for readers pursuing this further. The secondary literature on AI unit economics is inconsistent to the point of being unusable. Reported per-token price declines range from 67 percent year over year to 600-fold since 2020, and reported agentic consumption multipliers range from 5 to 1,000 times, depending on which aggregator is citing which analyst. Those ranges are not reconcilable, and the figures above are limited to sources I could trace to a primary publication.

The metric that follows

If the argument holds, the question organizations are asking is the wrong one.

Cost per task is measurable, favorable to automation, and nearly meaningless on its own, because it prices the production of an answer rather than the production of an answer somebody can stand behind. In

regulated, investigative, financial, legal, and clinical work, the deliverable is not the output. It is the output plus the ability to explain why it should be trusted.

The comparison that matters is risk-adjusted cost per defensible outcome. On the AI side that includes model consumption, agent activity, infrastructure, integration, governance, human validation, error remediation, retained liability, and vendor dependency. On the human side it includes compensation, benefits, management, training, technology, quality control, error remediation, and retained liability. Both are real. Neither is captured by comparing a salary against a subscription.

Framed that way, the operative question is not whether an AI team is cheaper than a human team. It is at what level of consumption, complexity, oversight, and retained liability the AI operating model stops being the lower-cost option.

> *The cheapest way to produce an answer and the cheapest way to be right are not the same calculation. Only one of them shows up on the invoice.*

References

Bai, L., Huang, Z., Wang, X., Sun, J., Mihalcea, R., Brynjolfsson, E., Pentland, A., & Pei, J. (2026). How do AI agents spend your money? Analyzing and predicting token consumption in agentic coding tasks. arXiv. https://arxiv.org/abs/2604.22750

Berkley Insurance Company. (2024). Artificial intelligence exclusion (absolute) [Endorsement form PC 51380 00 (06-24)]. https://www.hunton.com/assets/htmldocuments/noindex/PC-51380-00-06-24-Artificial-Intelligence-Exclusion-Absolute.pdf

FinOps Foundation. (2026). State of FinOps 2026. https://data.finops.org/

Konkel, J., & Ayaz, H. (2026, April 13). AI exclusions in insurance policies: Broad language, uncertain impact. Policyholder Pulse, Pillsbury Winthrop Shaw Pittman LLP. https://www.policyholderpulse.com/ai-exclusions-insurance-policies/

Mobley v. Workday, Inc., No. 3:23-cv-00770-RFL (N.D. Cal. May 16, 2025) (order granting preliminary collective certification, Dkt. No. 128). https://www.govinfo.gov/content/pkg/USCOURTS-cand-3_23-cv-00770/pdf/USCOURTS-cand-3_23-cv-00770-1.pdf

The Collective Risk Intelligence. (2026). Automated hiring systems: Applicant tracking systems, AI screening tools, and applicant-flow outcomes (TCRI-RA-002). https://www.tcriskintel.com

U.S. Equal Employment Opportunity Commission. (2023, September 11). iTutorGroup to pay $365,000 to settle EEOC discriminatory hiring suit. https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit

The Collective Risk Intelligence provides analyst-led investigative intelligence, due diligence, and risk analysis for business, legal, investment, and high-stakes decision environments.

Sean Mahon, Founder and Principal
smahon@tcriskintel.com · www.tcriskintel.com