

The hidden limits of AI indemnification — Managing IP risk across the generative AI stack

By Sasha S. Rao, Esq., and Todd M. Hopfinger, Esq., Sterne, Kessler, Goldstein & Fox

SEPTEMBER 9, 2026

Companies deploying generative AI frequently assume that a model provider's intellectual-property indemnification provision will protect them if an infringement claim arises. In practice, however, the scope of that protection often depends less on the indemnity itself than on the exclusions that follow.

Because infringement claims often arise at the customer or end-user level, while indemnification terms are negotiated between model providers and their direct customers, companies may discover significant coverage gaps only after a dispute emerges.

Many of those exclusions address activities that are common, and often essential, to AI deployment, including customer inputs, integrations, modifications, applications built on foundation models, and agentic functionality.

Because infringement claims often arise at the customer or end-user level, while indemnification terms are negotiated between model providers and their direct customers, companies may discover significant coverage gaps only after a dispute emerges.

These dynamics highlight the importance of evaluating indemnification provisions across the entire AI stack rather than reviewing any agreement in isolation.

AI carries unique IP liability risks

Who is liable?

The liability risks of generative AI are no longer theoretical. Recent copyright litigation has challenged both the use of copyrighted materials to train foundation models and the generation of allegedly infringing outputs. See, e.g., *The New York Times Co. v. Microsoft Corp.*, et al.

(No. 24-4872, S.D.N.Y.), *Kadrey, et al. v. Meta Platforms, Inc.* (No. 23-3417, N.D. Cal.), and *Disney Enterprises, Inc. v. Midjourney, Inc.* (No. 25-5275, C.D. Cal.).

One difficulty in allocating AI-related liability is that no single party controls the entire technology stack.

These disputes highlight why indemnification has become a critical component of AI contracting. Further, AI systems may generate technical solutions or methods that potentially implicate patent rights when implemented by users or AI agents.

One difficulty in allocating AI-related liability is that no single party controls the entire technology stack. A model provider develops the foundation model, an intermediary may fine-tune or integrate that model into a broader product, and an end user ultimately determines how the system is used through prompts, enterprise data sources, external tools, or agents. Unlike traditional software, responsibility is distributed across multiple participants.

As a result, liability often materializes downstream while indemnification rights are negotiated upstream. A customer may commit to indemnify its users without fully understanding the limitations of the indemnification available from the model provider.

Likewise, a model provider may have little visibility into how its technology is ultimately deployed. This creates the possibility of contractual gaps in which liability exists but corresponding indemnification protection does not.

What triggers liability?

AI indemnification presents challenges that do not exist in traditional software agreements. Conventional software generally performs defined functions in a predictable manner, allowing responsibility for alleged infringement to be tied to identifiable functionality.

Generative AI systems, by contrast, are probabilistic, non-deterministic technologies whose outputs may reflect contributions from model providers, application developers, end users, external data sources, and agents.

As a result, determining whether infringement arose from the model, implementation, prompt, data source, or autonomous system may itself become a disputed issue. This makes fault-based exclusions particularly difficult to apply.

The practical question, however, is not whether indemnification exists, but whether it remains available when the technology is deployed as intended.

Anthropic, for example, excludes indemnification for use of its services or outputs in a manner that a customer “knows or reasonably should know” infringes another’s rights. *Commercial Terms of Service*, Anthropic, § K.3(d) (June 17, 2025), <https://bit.ly/4xz1DID>.

Google similarly excludes certain claims where a customer knew or should have known that generated output was likely infringing. *Service Specific Terms*, Google Cloud, § 20.i(1) (July 29, 2026), <https://bit.ly/4qU3uoT>.

Such provisions may convert coverage disputes into litigation over causation and the customer’s state of mind.

Indemnification and exclusions

Most major model providers offer some form of intellectual-property indemnification because customers generally lack visibility into model training data and would otherwise assume significant infringement risk.

For example, Anthropic indemnifies customers against claims that authorized use of its services or outputs violates a third party’s intellectual-property rights. Anthropic, § K.1.

OpenAI likewise agrees to indemnify customers against certain third-party claims alleging that its services infringe intellectual-property rights. *OpenAI Services Agreement*, OpenAI, § 13.1 (December 1, 2025), <https://bit.ly/3V8falB>.

The practical question, however, is not whether indemnification exists, but whether it remains available when the technology is deployed as intended.

Many model providers exclude claims arising from modifications, combinations with third-party technology, customer-provided data, or use of patented inventions contained in outputs. Anthropic’s commercial terms include each of these exclusions. Anthropic, §§ K.3(a)–(e).

OpenAI contains similar carve-outs for third-party modifications, combinations with third-party products, customer content, and customer applications. See, OpenAI, § 13.1.

The challenge is that these exclusions may overlap with ordinary AI deployment rather than extraordinary misuse. Unlike traditional software agreements, where modification and combination exclusions often apply to atypical uses, many of the activities potentially implicated by AI exclusions constitute the ordinary and intended deployment of modern AI systems.

An indemnification provision may appear broad on its face, yet provide significantly less protection if the exclusions reach the very activities required to deploy the AI system as intended.

Vendors routinely integrate foundation models into broader applications, connect them to enterprise software, and supplement them with retrieval systems, tools, and proprietary functionality. End users routinely interact with these systems through prompts, external data sources, custom instructions, and agentic workflows.

As a result, the critical question is not whether an exclusion exists, but whether it encompasses activities necessary to achieve the intended business use case. Put differently, an indemnification provision may appear broad on its face, yet provide significantly less protection if the exclusions reach the very activities required to deploy the AI system as intended.

Indemnification of inputs, outputs, and agents

Inputs

Indemnification for inputs generally falls into two categories: provider training data and customer-provided inputs.

Providers commonly indemnify claims tied to their own training data. For example, Google expressly extends its indemnity to claims arising from Google’s use of training data to develop certain pre-trained models. See Google Cloud, § 20.i(1). Anthropic similarly extends protection to claims involving training data used to develop models that form part of its services. See Anthropic, § K.1.

By contrast, providers frequently exclude customer prompts, fine-tuning materials, and other customer-supplied content. Anthropic excludes “inputs or other data provided by Customer.” *Id.*, § K.2(a). Other providers similarly exclude customer-provided training materials and, in some cases, require customers to indemnify the provider against claims arising from customer inputs. See, e.g., OpenAI, § 13.1 (“Customer Content”).

The distinction reflects control. Providers generally control model training. They do not control customer prompts, fine-tuning data, or downstream implementation choices.

Outputs

Output indemnification is often narrower because providers have limited control over downstream deployment.

Anthropic excludes modifications made by customers to services or outputs. Anthropic, § K.3(a). Google limits coverage to certain categories of unmodified generated output. Google Cloud, § 20(i)(i). OpenAI excludes output that has been modified, transformed, or combined with non-OpenAI products or services. OpenAI, § 13.1.

The practical issue is that activities central to AI deployment, including wrappers, retrieval systems, custom instructions, and product integrations, may fall within modification or combination exclusions depending on the contract language. Accordingly, output-related exclusions should be evaluated in light of the intended deployment architecture rather than in isolation.

Agents

Agentic AI creates additional indemnification challenges because the AI may itself perform the challenged conduct rather than merely generate content.

Providers have begun addressing this risk expressly. Anthropic excludes claims involving “the practice of a patented invention contained in an Output.” Anthropic, K.3(e). Google separately addresses actions performed by AI agents and allocates responsibility for those actions to the customer. Google Cloud, § 20.n.

As AI systems increasingly move from content generation to autonomous action, these exclusions may significantly narrow available protection. An indemnity negotiated around generated outputs may cover a shrinking portion of the actual product as deployment shifts toward agentic systems.

Practical contracting considerations

Counsel reviewing AI procurement, licensing, and deployment agreements should focus on four issues:

- Whether fine-tuning, retrieval-augmented generation, adapters, prompts, guardrails, and wrappers constitute prohibited “modifications.”
- Whether deployment with enterprise systems, APIs, databases, repositories, and third-party tools triggers “combination” exclusions.
- Whether customer-facing indemnification obligations exceed the protections available from upstream model providers.
- How responsibility is allocated for actions performed by AI agents.

The answers to these questions will often determine the practical scope of protection far more than the indemnification grant itself.

Conclusion

Generative AI indemnification provisions cannot be evaluated based solely on the existence of coverage. The more important question is whether that coverage survives the way the technology is actually deployed.

Unlike traditional software, generative AI operates through a chain of participants that may include model providers, application developers, enterprise customers, end users, and autonomous agents.

Liability may emerge at any point along that chain, while contractual risk allocation is often negotiated elsewhere. The result is that indemnification rights and indemnification exposure do not always align.

As organizations move from AI experimentation to enterprise deployment, reviewing indemnification language across the entire AI stack may be just as important as evaluating the underlying technology itself.

Sasha S. Rao and Todd M. Hopfinger are regular contributing columnists on AI and intellectual property for Reuters Legal News and Westlaw Today.

About the authors



Sasha S. Rao (L) is counsel in **Sterne, Kessler, Goldstein & Fox’s** trial and appellate practice group, where he focuses on matters including artificial intelligence (AI), pharmaceuticals, materials science, chemical arts, and automotive technology. He can be reached at SRao@sternekessler.com. **Todd M. Hopfinger** (R) is a director in the firm’s electronics practice group. He specializes in strategic intellectual property (IP) counseling and patent prosecution for technology companies. His areas of expertise include fields such as artificial intelligence (AI), bioinformatics, data science, blockchain, 5G, consumer technology, transportation and logistics (including autonomous vehicles),

digital advertising, and cybersecurity. He can be reached at thopfinger@sternekessler.com. The authors are based in Washington, D.C.

This article was first published on Reuters Legal News and Westlaw Today on September 9, 2026.