The Rise of AI Litigation: How Emerging Tech Companies Can Safeguard Their IP and Liability Exposure

By James A. Wolff

March 25, 2025

As artificial intelligence continues to drive innovation at an unprecedented pace, it has also become a battleground for litigation, particularly concerning intellectual property misappropriation, data scraping, and model transparency.

The legal frameworks traditionally relied upon to govern software, data use, and trade secrets are proving increasingly inadequate in addressing the complexities posed by generative models and autonomous decision-making systems.

This lack of clarity over ownership rights, fair use doctrines, and the extent to which proprietary data can be leveraged for machine learning development has created an uncertain legal environment.

Startups and established enterprises alike are now forced to confront mounting litigation risks, regulatory scrutiny, and competitive threats in a legal landscape that is evolving faster than legislative and judicial bodies can respond.

At the center of this paradigm shift is the question of data provenance and usage rights. Many AI systems are trained on vast repositories of publicly available and proprietary datasets, often scraped from online sources without explicit authorization.

This practice has triggered lawsuits alleging infringement of copyright, database rights, and trade secrets, with plaintiffs arguing that AI developers are engaging in wholesale misappropriation of intellectual property without compensation.

Credit: NicoEINino/Adobe Stock

A significant area of contention in AI-related litigation concerns the use of copyrighted materials to train datasets. Recent cases have demonstrated how courts are beginning to address the unauthorized use of intellectual property in machine learning models.

In one example, the Southern District of New York held that the plaintiff plausibly alleged that defendants had removed copyright-management information (CMI) from its articles used to train large language models (LLMs), constituting a violation of the Digital Millennium Copyright Act (DMCA) under 17 U.S.C. §1202(b). *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-CV-1515 (JSR), 2025 WL 556019, at *3 (S.D.N.Y. Feb. 20, 2025).

However, the court dismissed claims under the DMCA's prohibition on distributing copies of copyrighted material without CMI, citing insufficient factual support. This decision reflects the judiciary's willingness to scrutinize how AI companies handle copyrighted materials, particularly in the context of data preprocessing and model training.

Similarly, in another decision, the court denied discovery requests sought by an AI company in a copyright infringement suit, emphasizing that the fair use defense in AI-related cases hinges on the nature of the defendant's use rather than the plaintiff's internal documents. *New York Times Co. v. Microsoft Corp.*, No. 23-CV-11195 (SHS) (OTW), 2024 WL 4874436, at *2 (S.D.N.Y. Nov. 22, 2024).

The rejection of fair use in AI training was further reinforced when a Delaware court ruled that repurposing Westlaw's headnotes for an AI-driven legal tool constituted copyright infringement. *Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc.*, No. 1:20-CV-613-SB, 2025 WL 1234567 (D. Del. Feb. 11, 2025).

These rulings reinforce the principle that AI companies seeking to invoke fair use cannot rely on discovery to justify broad-scale ingestion of copyrighted content, thereby narrowing the scope of potential defenses in cases involving unauthorized dataset usage.

Another significant recent decision underscores the standing and remedial complexities surrounding standing and the availability of remedies in AI copyright litigation, illustrating the procedural hurdles that plaintiffs face when seeking redress for the unauthorized use of their content in machine learning models.

In that case, plaintiffs alleged that OpenAI used their copyrighted works to train ChatGPT without proper attribution, violating the DMCA. *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24 CIV. 01514, 2024 WL 4711729, at *1 (S.D.N.Y. Nov. 7, 2024); *See Also Getty Images (US), Inc. v. Stability AI, Inc.*, No. 1:23-cv-00135-GBW (D. Del. filed Feb. 3, 2023).

However, the court held that the plaintiffs lacked Article II standing to seek retrospective and injunctive relief, finding no concrete injury and deeming the alleged harm too speculative, signaling potential procedural hurdles for content creators challenging AI companies over unauthorized dataset use.

Beyond data scraping, concerns over intellectual property misappropriation extend to the outputs generated by AI systems and their underlying decision-making processes. An increasing number of cases challenge the originality and authorship of AI-created content, raising novel questions about whether derivative works infringe upon underlying training datasets.

Companies using AI for content generation, software development, and product design now face heightened legal exposure if model outputs can be traced to copyrighted or proprietary material.

Judicial scrutiny of model transparency is also particularly relevant given the opacity of AI decision-making processes. Regulators and courts increasingly demand explainability in AI systems, particularly in cases where companies argue that model outputs are sufficiently transformative or fall under fair use defenses.

New York Times reflects this evolving tension, demonstrating how courts are scrutinizing fair use claims in AI-related copyright disputes and emphasizing that justification must come from the copier's actions rather than external factors. In rejecting Microsoft's attempt to obtain discovery from the plaintiff in a copyright suit, the court reinforced that fair use must be evaluated based on the copier's actions rather than on internal documents or rights holder practices.

This ruling signals that AI developers cannot rely on external factors to justify AI model decisions that incorporate copyrighted works. Instead, they must be transparent in how their models generate outputs and must be prepared to defend those decisions based on their own actions.

The court's decision in *Intercept Media,* underscores a growing judicial focus on AI transparency, particularly regarding the handling of copyright management information (CMI). By allowing claims under the DMCA to proceed, the court signaled a willingness to scrutinize how AI companies manage CMI during data ingestion and model training.

This development suggests that courts may increasingly hold AI developers accountable for obfuscating the provenance of their training data, thereby imposing stricter transparency obligations on companies that fail to disclose their use and alteration of copyrighted material in AI training processes.

Raw Story also highlights standing and redressabil-

ity in AI litigation. Courts are beginning to scrutinize whether plaintiffs can seek retrospective and injunctive relief when AI companies use their copyrighted materials. To establish standing, plaintiffs must demonstrate a concrete and particularized injury directly caused by the defendant's use of their works and show that a favorable ruling would provide meaningful redress.

Here, the court found that speculative harms and the inability to show specific economic or reputational damage were insufficient. While courts are not yet imposing sweeping transparency requirements, the growing body of case law suggests that companies relying on opaque AI systems will face greater litigation risks as regulatory enforcement mechanisms continue to evolve.

Given the rapidly shifting legal landscape, the traditional approaches to IP protection and liability mitigation that once sufficed for software and data-driven enterprises are no longer adequate. AI companies must adopt a forward-looking legal strategy that addresses the unique risks posed by evolving case law and regulatory developments.

Central to this strategy is the need to implement robust data governance policies that delineate clear sourcing and licensing practices. Companies relying on publicly available datasets should proactively assess whether scraping or data ingestion methods are defensible under fair use and copyright doctrines. Where uncertainties exist, obtaining licenses or utilizing synthetic data alternatives may provide a more secure path to avoiding costly infringement claims.

Beyond data governance, AI companies must strengthen their intellectual property protections by securing patents for proprietary model architectures, training methodologies, and novel AI applications. While copyright protections for AI-generated content remain in flux, patents provide a more stable mechanism for safeguarding competitive advantages in algorithm development and deployment.

In cases like *Raw Story* and *Intercept Media*, stronger contractual agreements and licensing structures could have limited claims over unauthorized dataset use, while clearer attribution policies might have mitigated liability under the DMCA.

Trade secret protections should also be reinforced through strict internal controls, employee agreements, and cybersecurity measures to prevent inadvertent exposure or theft of proprietary model designs and training datasets, particularly as courts scrutinize AI transparency and fair use defenses in litigation like *New York Times*.

Contractual safeguards are critical in mitigating liability exposure, particularly as seen in cases like *Intercept Media* and *Raw Story*, where disputes over dataset use and attribution might have been avoided with clearer licensing terms. AI companies should structure agreements with vendors, partners, and customers to include indemnification provisions that allocate risk appropriately.

Given the uncertainties surrounding AI-generated outputs, well-defined licensing agreements could have preempted claims in *Getty Images* and *New York Times* by explicitly addressing rights, limitations, and usage permissions. Clear contractual language on liability disclaimers, compliance obligations, and dispute resolution mechanisms provides a necessary layer of protection in an increasingly litigious environment.

As AI litigation escalates, companies that fail to take proactive measures risk financial liability and reputational damage impacting their market position. The high-profile lawsuits shaping the legal contours of AI regulation today will set the precedents that will define the industry for years to come.

AI innovators must recognize that legal risks are not theoretical but tangible threats to long-term viability. By integrating risk management early, companies can navigate AI's next decade of AI-driven transformation while safeguarding their intellectual property and minimizing litigation exposure.

James A. Wolff *is Counsel and Chair of the Emerging Technologies Law Group at Warshaw Burstein, LLP*