Developers Are Testing Defenses In Generative AI Litigation

By **Christopher Valente, Michael Meredith and Amy Wong** (September 15, 2023)

Although still in their infancy, a growing number of recently filed lawsuits associated with generative artificial intelligence training practices, products and services have provided a meaningful first look into how U.S. courts may address the privacy, consumer safety and intellectual property protection concerns that have been raised by this new and inherently evolving technology.

The legal theories that have served as the basis of recent claims have varied widely, but are often overlapping, and have included:

- Invasion of privacy and property rights;
- Patent, trademark and copyright infringement;
- Theft, conversion or misappropriation; and
- Violations of state consumer protection laws.

The factual foundation for each of these claims is that publicly available personal, private or protected information has been improperly collected or scraped from the internet to train or develop generative AI products.

While the outcomes of these early generative AI cases are far from certain, preliminary indications suggest that courts are not succumbing to the hype and rhetoric and are approaching generative AI claims with a healthy level of skepticism.

Yet, many of the potential defenses have still not been tested in the context of generative AI. The coming months will be pivotal in setting the tone for generative AI litigation moving forward.

This article aims to provide a snapshot of possible defenses in the rapidly growing field of generative AI law in the U.S.

Christopher Valente

Michael Meredith

Amy Wong

Possible Defenses

There are a variety of defenses that have already been effectively asserted by defendants in generative AI litigation.

Common themes include lack of standing, reliance on the fair use doctrine, and the legality of so-called data scraping.

The following is a brief summary of the key principles underlying each of these possible defenses that AI developers may rely on in future litigation.

Lack of Standing

In July, the U.S. Court of Appeals for the Seventh Circuit in Dinerstein v. Google

LLC affirmed the dismissal of breach of privacy claims brought on behalf of a putative class of patients of the University of Chicago Medical Center for lack of standing.[1]

The Dinerstein decision could provide AI developers with precedent for an important, and possibly complete, defense to claims that rely on the assumption that mere use of copyrighted, consumer or personal data to train AI models constitutes a legally cognizable harm.

The holding in Dinerstein suggests, to the contrary, and regardless of the legal theory selected, that the individual owners of copyrighted, personal or private data used in AI training must demonstrate a plausible, concrete injury to establish standing to pursue those theories.

In Dinerstein, the plaintiffs alleged that UCMC breached its contractual privacy arrangements with its patients, invaded their privacy, and violated Illinois' consumer protection statute by using several years of anonymized patient medical records to train an AI model that could be used in software to anticipate patients' future healthcare needs.

The U.S. District Court for the Northern District of Illinois ultimately dismissed the plaintiffs' claims due to lack of standing and for failure to state a claim, noting that plaintiffs failed to establish damages associated with the disclosure of their anonymized patient data or defendants' tortious intent.

Affirming the dismissal, the Seventh Circuit "agreed with [the] decision to dismiss the case" but indicated that the analysis should "begin[] and end[] with standing."[2]

Specifically, it explained that, because the plaintiffs failed to allege that any patient data was used to identify any specific member of the class and the defendants contractually and "explicitly agreed not to identify any individual,"[3] the plaintiffs could not establish the existence of any "concrete and particularized, actual or imminent" harm necessary to "supply the basis for standing."[4]

Fair Use

Another broad defense that might be successfully pursued by AI developers against any copyright claim is the well-recognized doctrine of fair use.

Fair use is a defense to claims of infringement when copyrighted material is used in a transformative way. Transformative use can occur when copyrighted material is used to serve different market functions or expand the utility of the copyrighted work.

The doctrine appears particularly appropriate for the AI training process, which does not involve the traditionally impermissible copying and commercial reproduction of copyrighted work and, instead, only analyzes copyrighted material to detect patterns in an effort to develop a new function or application, namely, a large language model or other generative AI product.

To date, no U.S. court has explained the appropriate application of the fair use doctrine in the context of generative AI models or AI-generated materials. However, the doctrine has provided a complete defense in similar situations.

For example in the 2015 case of Authors Guild v. Google Inc.,[5] the U.S. Court of Appeals for the Second Circuit concluded that a search engine's publication of small portions of copyrighted books was transformative because it improved access to that information.

In 2002 the U.S. Court of Appeals for the Ninth Circuit, in Kelly v. Arriba Soft Corp.,[6] held the same with respect to searchable images of copyrighted visual artwork.

In response to lawsuits alleging copyright infringement, some AI developers have already suggested the fair use doctrine's applicability. The U.S. Supreme Court's 2021 decision in Google v. Oracle,[7] which determined that Google's use of portions of Oracle's code to create its Android operating system was fair use, may also support the use of this defense in the context of generative AI.

GitHub and Microsoft have also argued that the plaintiffs in Doe v. GitHub Inc., filed in the U.S. District Court for the Northern District of California in November 2022, affirmatively chose not to assert claims of copyright infringement because they "would run headlong into the doctrine of fair use."[8]

Stability AI, similarly, has also defended its model's training processes by stating, "anyone that believes that this isn't fair use does not understand the technology and misunderstands the law."[9]

Legality of So-Called Data Scraping

Finally, while generative AI developers may have relied on scraping of the internet to develop training datasets for their products, they are far from the first group of companies to "scrape" the internet for commercially useful information.

In fact, it is a common practice among data science and technology companies.

One such company, hiQ Labs Inc., for example, famously scraped information from the publicly available profiles of online users of the business networking site LinkedIn in order to provide employers with data and analysis regarding potential recruits and their job-seeking behaviors.

In the 2022 case of hiQ Labs Inc. v. LinkedIn Corp.,[10] the Ninth Circuit rejected claims that the practice of scraping publicly available data constitutes an invasion of privacy or violation of the Computer Fraud and Abuse Act.

In its decision, the court focused on the distinction of publicly available data and data marked "private," and held that accessing publicly available data does not constitute access without authorization under CFAA unless the data has been marked private.

AI developers will likely be able to take advantage of the precedent established in hiQ Labs to defend their data collection practices and can further expect that the hiQ Labs decision will likely feature prominently in the numerous cases pending in the Northern District of California.

Future Trajectory

While the current wave of generative AI litigation continues to work its way through the courts, recent trends suggest that plaintiffs attorneys may be eager to expand beyond the generative AI developers to target companies that adopt or use generative AI products or solutions.

As such, both developers and users of generative AI products and solutions would do well to monitor the viability of the defenses outlined above as they prepare and implement risk management strategies for their generative AI products or solutions.

Christopher J. Valente is a partner, Michael W. Meredith is an associate and Amy Wong is a partner at K&L Gates LLP.

K&L Gates partner Michael J. Stortz and associate Peter E. Soskin contributed to this article.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] Dinerstein v. Google, LLC, 73 F.4th 502, 508 (7th Cir. 2023).

[2] Id.

[3] Id. at 516.

[4] Id. at 508.

[5] 804 F.3d 202 (2d Cir. 2015).

[6] 336 F.3d 811 (9th Cir. 2002).

[7] 141 S. Ct. 1163 (2021).

[8] Motion to Dismiss Operative Complaint in Consolidated Actions, Doe v. GitHub, Inc., Case No. 4:22-cv-06823 (N.D. Cal. 2022), ECF No. 108.

[9] Riddhi Setty, First AI Art Generator Lawsuits Threaten Future of Emerging Tech, BLOOMBERG LAW (January 20, 2023), available athttps://news.bloomberglaw.com/ip-law/first-ai-art-generator-lawsuits-threaten-future-of-emerging-tech.

[10] Case No. 17-3301 (9th Cir. 2022).