Copyright Issues with Generative AI



By: (L-R): Matthew Braunel - Partner (St. Louis), Michael Parks, Partner (Chicago), Alex Weidner - Associate (St. Louis)

Copyright Issues with Generative AI

Several cases are working their way through the Court system that might impact how Generative AI tools are trained, the ways in which users interact with Generative AI tools and the work product that Generative AI tools generate. Most of the legal issues surrounding these cases sound in copyright and that is where this article will focus.

Copyright Basics

U.S. Copyright law finds its basis in Art. I, Section 8, Clause 8 of the Constitution, referred to as the patent and copyright clause. The purpose of this clause, according to the Constitution, is "[t]o promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." The Copyright Act, Title 17 of the United States Code, provides additional context – permitting copyright protection for original works of authorship fixed in any tangible medium of expression. 17 U.S.C. § 102. The bar for originality is relatively low. *Feist Publications, Inc. v.*

Rural Telephone Service Co., Inc. 499 U.S. 340 (1991) (facts in telephone book are not copyrightable, but compilations of those facts can be).

The Exclusive Rights

Copyright protection affords the author six exclusive rights in the protected work – reproduction, preparation of derivative works, distribution and three types of public performance. 17 U.S.C. § 102. The exclusive rights relating to reproduction and preparation of derivative works are the most frequently invoked in cases involving Generative AI tools. The right of reproduction generally prohibits copying, in whole or in part, of protectable portions of a Work without the permission of the copyright owner. 17 U.S.C. § 102(1). The right of preparation of derivative works prohibits creating a new work based on one or more existing works without the permission of the copyright owner. 17 U.S.C. § 102(2). Common examples of derivative works include translations, movie versions of literary materials and condensations of preexisting works. *See* United States Copyright Office Circular 14. The exclusive rights do expire as set forth in 17 U.S.C. §§ 302, 303 and 304, and figuring out the correct expiration date can sometimes feel like solving an algebra problem. Although the authors of this paper are algebra-loving intellectual property lawyers, for simplicity, the exclusive rights currently last about 100 years, although they have expired in less time in earlier enacted versions of the statute. For reference, the standard for determining copyright infringement involves comparing the Work with the alleged infringing content for substantial similarity. *See e.g.*, *Boisson v. Banian, Ltd*, 273 F.3d 262, 268 (2d Cir. 2001). We will withhold further analysis of substantial similarity, as the current cases either require a simple analysis based on verbatim copying or a much more complicated analysis with respect to the exclusive right to prepare derivative works.

Fair Use

Congress has codified certain limitations on the exclusive rights of the copyright owner in 17 U.S.C. § 107 – Fair Use. The statute provides that a violation of an exclusive right under § 106 is not an infringement if done for a "fair use" purpose such as criticism, comment, news reporting, teaching, scholarship or research. 17 U.S.C. § 107. To determine if fair use applies, Congress has set forth several factors:

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(2) the nature of the copyrighted work;
(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
(4) the effect of the use upon the potential market for or value of the copyrighted work.

Generative AI Training

The general public was exposed to Generative AI tools in November 2022, but their development began in the 1940s. For modern purposes, and dropping some of the legally charged vocabulary, Generative AI tools have been trained on written Works by reading and retaining information and then using that information, on some level, to generate what appears to be new content.

There is not a lot of transparency into the Works upon which the Generative AI tools were trained. Some of the filed Complaints cite to digital compilations begun in the 1970s, online databases of books containing hundreds of thousands of titles, and crawls of the internet, including Google patents, Wikipedia and the websites of many media companies.

The creators of this original content, including well-known and popular authors, celebrities and groups that support authors, claim that their Works were included in these training materials without their permission. The technology companies deny the allegations and claim that any use of the Works is fair use – not a violation of copyright rights.

The Exclusive Rights vs. Fair Use

The interplay of § 106 and § 107 sets the stage for the legal battle between content creators and technology companies. The content creators claim that technology companies copied their Works in training the Generative AI tool. Some of these allegations include that the Generative AI tools create derivative works in their processes, including that the output of the Generative AI tool is (or perhaps could be) a derivative work of the content creator's Work.

Seemingly Settled Questions

The cases filed after November 2022 are not the first time some of these issues have been addressed. There were allegations of copyright infringement of vast amounts of content creators' content by Google in the early 2000s as part of

Google's Library Project, the stated purpose of which is "for users to search on Google through millions of books." Google indicated it would make available books that were in the public domain (presumably those in which copyright rights had expired). Two organizations that represent the interests of authors brought claims against Google for digitizing works without permission. Class settlements over $100 million were reached, but not confirmed. In 2015, the Second Circuit Court of Appeals affirmed a finding of no infringement based on fair use – Google's Library Project enhanced sales for the content creators and provided a public service without violating intellectual property law. The Supreme Court denied *cert.*

The Current Claims

Several current cases are provided below. These allegations generally build on the technology provider's training of the Generative AI based on protected Works and extend that protection to the Generative AI's seemingly newly created content. In some cases, there are allegations that a large percentage of all content creators will be put out of work by the technology. *See e.g.*, Complaint, ¶ 112, *Authors Guild et al v. OpenAI Inc. et al*, No. 1:23-cv-08292 (S.D.N.Y. Sep 19, 2023) ("Goldman Sachs estimates that generative AI could replace 300 million full-time jobs in the near future, or one-fourth of the labor currently performed in the United States and Europe."); *id.* at ¶ 114("An Authors Guild member who writes marketing and web content reported losing 75 percent of their work as a result of clients switching to AI").

Cases to Watch

Silverman et al v. OpenAI, Inc. et al, No. 3:23-cv-03416 (N.D. Cal. Jul 07, 2023): Comedian Sarah Silverman, along with others, filed a copyright infringement case against OpenAI based on the alleged use of copyrighted material in the training of OpenAI's large language models. OpenAI filed a motion to dismiss that was granted in part, with the Court dismissing claims for vicarious copyright infringement, removal or alteration of copyright management information, negligence, and unjust enrichment, with leave to amend. The Court declined to dismiss claims for unfair trade practices. An Amended Complaint was recently filed. The case has been consolidated with *Tremblay et al v. OpenAI, Inc. et al*, Docket No. 3:23-cv-03223 (N.D. Cal. Jun 28, 2023) and is ongoing.

L. et al v. Alphabet Inc. et al, No. 3:23-cv-03440 (N.D. Cal. Jul 11, 2023): This case involves class-action claims against Google based on alleged web scraping

conducted, in part, to obtain materials used for training AI. The plaintiffs assert claims for unfair competition, negligence, violations of the Comprehensive Computer Data Access and Fraud Act, invasion of privacy, intrusion upon seclusion, larceny, conversion, trespass to chattels, intentional interference with an existing contract, breach of a third-party beneficiary contract, unjust enrichment, and copyright infringement. An Amended Complaint was filed in January, and motion to dismiss briefing is ongoing.

Authors Guild et al v. OpenAI Inc. et al, No. 1:23-cv-08292 (S.D.N.Y. Sept. 19, 2023): Authors Guild, a group of fiction and nonfiction authors, filed a copyright infringement action against OpenAI and Microsoft based on the alleged use of the authors' copyrighted books in the training of OpenAI's large language models. The complaint alleges current and ongoing harm to authors resulting from the use of their copyrighted works to train AI systems. While certain entities have answered the complaint, motion to dismiss briefing is ongoing for other entities.

The New York Times Company v. Microsoft Corporation et al, No. 1:23-cv-11195 (S.D.N.Y. Dec 27, 2023): The New York Times alleges claims for, among other things, copyright infringement, the removal or alteration of copyright management information, and misappropriation based on the alleged use of New York Times content in the training of AI systems. In particular, the New York Times points to alleged instances in which portions of its articles were reproduced in response to prompts to the AI system. A partial motion to dismiss has been filed, but oral argument has not yet been scheduled. The case is ongoing.

Thompson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc., 2023 WL 6210901 (D. Del. Sept. 25, 2023) – parties cross motions for summary judgment of infringement and fair use both denied due to factual issues over the underlying copyright infringement and fair use issues. Set for trial August 2024.

Non-AI Copyright Issues

In *Petrella v. Metro-Goldwyn-Mayer, Inc.*, the Supreme Court held that the equitable doctrine of laches does not bar copyright claims that are timely within the three-year limitations period. 572 U.S. 663 (2014). The Court's decision was based, in part, on the fact that the 3-year statute of limitations inherently limits relief so the application of laches was unnecessary. *See id.* at 672 (indicating that plaintiffs may "gain retrospective relief running only three years back from the date the complaint was filed."). In the wake of that decision, a circuit split has arisen between the Second and Ninth Circuits – the two preeminent circuits for

copyright law – as to whether the Copyright Act's 3-year statute of limitations is a complete bar to recovery for infringement that occurred more than three years prior to when suit was filed or whether a doctrine called the "discovery rule" may continue to permit recovery for periods beyond the limitations period. *Compare Sohm v. Scholastic, Inc.*, 959 F.3d 39, 49-50 (2d Cir. 2020) (holding that, even under the discovery rule, a copyright plaintiff may not recover for infringement occurring more than three years before the plaintiff filed suit) *with Starz Ent., LLC v. MGM Domestic Television Distrib., LLC*, 39 F.4th 1236, 1242-44 (9th Cir. 2022) (holding that Petrella does not bar recovery for infringement that occurred more than three years before the filing of an otherwise timely suit under the discovery rule). In a 2023 decision, the Eleventh Circuit joined the Ninth Circuit's view and, when the decision was appealed, the Supreme Court granted certiorari to resolve the growing circuit split. *Nealy v. Warner Chappell Music, Inc.*, 60 F.4th 1325, 1331 (11th Cir.), *cert. granted in part*, 144 S. Ct. 478, 216 L. Ed. 2d 1313 (2023). The Supreme Court's decision is expected later this year.

What's Next?

It's difficult to predict what the Court system will ultimately conclude about the use of copyrighted works to train AI models. Because the current cases are still percolating in the district courts, it remains to be seen whether a Circuit dispute will arise over the issues and, if so, whether the Supreme Court will step in to resolve the matter. Perhaps there's an AI system out there that can accurately predict what will happen, but we're not holding our breath for that.