IP Alert | Reddit Sues Perplexity Alleging Illegal Scraping of Google Excerpts of Copyrighted Reddit Posts, Setting Up Yet Another AI Copyright Battle

By Sydney Huppert, Kirk Sigmon, and Victoria Webb

Last week, Reddit, Inc. (“Reddit”) sued Perplexity AI, Inc. (“Perplexity”) and other co-defendants in the Southern District of New York in Reddit, Inc. v. SerpApi LLC et al., No. 1:25-cv-08736 (S.D.N.Y., Oct. 22, 2025).Among other allegations, Reddit claims that Perplexity violated the Digital Millennium Copyright Act (“DMCA”) by using scraped Google-generated excerpts of Reddit content to help power its AI answer machine instead of obtaining copyright-protected content directly from Reddit. This case is yet another development in ongoing battles against AI companies allegedly using scraped content. But, somewhat unlike the recent lawsuit Encyclopedia Britannica brought against Perplexity (which primarily focuses on copyright and trademark infringement), Reddit’s case against Perplexity focuses squarely on the DMCA and the legality of circumventing online measures that Reddit—and Reddit’s authorized partners—use to protect its online content.

Perplexity’s Model Uses Online Content for “Grounding” its Products

Perplexity markets itself as an “AI-powered Swiss Army Knife for information discovery and curiosity.” Put simply, it claims to be an “answer machine” that leverages both large language models (“LLMs”) and real-time web searches to compile a response to user queries. The use of such real-time web searches is part of a technique known as “grounding,” and, according to Perplexity, allows it to “battle-test” and significantly improve the accuracy (and recency) of Perplexity’s results. To incorporate data from real-time web searches, Perplexity uses content collected from websites (including, allegedly, scraped content collected by third party companies,

such as Perplexity's co-defendants). This has not been a universally popular strategy: multiple organizations, such as Encyclopedia Britannica, recently sued Perplexity for allegedly infringing their copyrights, and those cases also remain pending.

Reddit Alleges Perplexity's Purported Use of Scrapped Data Violates the DMCA

Unlike these prior cases against Perplexity (which mostly focused on alleged copyright/trademark infringement), here, and among other allegations, Reddit alleges that Perplexity and its co-defendants violated the anti-circumvention aspects of the DMCA. According to Reddit's complaint, Perplexity used data, scraped by its co-defendants, that included excerpts and snippets of Reddit posts in Google Search Engine Results Pages ("SERPs"). In other words, Reddit argues that Perplexity and its co-defendants improperly gained access to and used copyrighted posts, images, and other content by retrieving that content via Google SERPs, bypassing the need to retrieve that content from the Reddit website directly. Interestingly, Reddit seemingly does not expressly contend that it has any legal right to those copyrighted works: rather, its complaint focuses on the idea that Perplexity and its co-defendants circumvented Reddit's mechanisms to protect those copyrighted works.

Reddit claims in its complaint that defendants circumvented different types of technological measures. First, Reddit alleges that the scraping of Google's SERPs (by Perplexity and all three co-defendants) circumvented Reddit's own measures to prevent unauthorized scraping, such as IP address-based rate limits, user identification tools, and the like. Reddit claims that it adopted these measures to limit others' access to Reddit data unless they agree to particular terms (presumably involving some form of payment). For example, according to Reddit's complaint, Google has entered into an agreement with Reddit, permitting Google to scrape Reddit's content and generate excerpts and summaries of that content on Google's SERPs. In contrast, according to Reddit, Perplexity has not entered into such an agreement with Reddit. Instead, Reddit alleges that Perplexity's co-defendants (which allegedly use, among other things, proxies to circumvent user ID and IP address rate limits on Reddit and Google) are unlawfully circumventing technological measures to control access to copyrighted content on Reddit's social platform under the DMCA, and that Perplexity allegedly joined in on that unlawful circumvention via its use of the scraped data. Specifically, Reddit alleges that the scraping from Google SERPs circumvents both Google's and Reddit's technological measures to control access to copyrighted content from Reddit.

Reddit Argues it Loses Income from Perplexity's Scraping—while Perplexity Responds on Reddit, Arguing It Only Uses Reddit like Humans Do

Reddit's complaint seeks damages, injunctive, and other equitable relief based on its allegations of DMCA violations, unfair competition, unjust enrichment, and civil conspiracy. According to Reddit, Perplexity and its co-defendants' unauthorized access and use of its material harms Reddit's ability

to monetize its data (and Reddit alleges in the Complaint that Perplexity's actions threaten the licensing arrangements that Reddit has made with willing partners).

Perplexity, for its part, has already responded in part to this lawsuit (by, amusingly, posting on Reddit), arguing that they are merely doing what human users do: sharing content and links from Reddit. And Perplexity claims the case is nothing more than a "sad example of what happens when public data becomes a big part of a public company's business model." Perplexity also ventures to "guess" that the case is "about a show of force in Reddit's training data negotiations with Google and OpenAI," adamantly claiming that it "doesn't train foundation models!"

Perplexity's Reddit post immediately amassed several comments. For example, commenters explored who actually owns rights in Reddit posts. Another Redditor encouraged Perplexity to "fire your social media team" or "whatever lawyer signed off on this statement," noting that aspects of Perplexity's post appeared to admit to several key points in the Complaint and that "[t]here is a reason companies don't do this!" And at least one Redditor fact-checked Perplexity's statement using Perplexity—with Perplexity's answer machine purportedly finding Perplexity's Reddit post ranged from "True" to "False," "Unverified" or "Partly True," for example.

In any event, Perplexity does not appear willing to go down without a fight, noting that they "won't cave." This language suggests that the case has some potential to move forward and answer novel questions raised in the AI context.

Lawsuit has Significant Possibility for Impact on Industries Using Scraped Content

This case, in conjunction with the Encyclopedia Britannica case filed earlier this fall, could have a massive impact on the ability of AI companies to use data scraped from online resources. Indeed, many different developers of AI products leverage real-time searches to improve accuracy on recent events and similar topics. Similarly, if use of scraped content could expose developers to legal risk under the DMCA, they might stop using this data, which may add yet another hurdle for those looking for new sources of training data. In fact, some evidence suggests that this case might also impact search engine optimization ("SEO") firms, which allegedly also scrape Google SERPs to help their clients understand their position relative to competitors. As the first case on this issue, we look forward to further developments from the Court.

Posted: October 28, 2025