

Recent District Court decision casts doubt on terms of use barring data scraping

June 24, 2024 | Client Update | 2-minute read

The Northern District of California recently held that state law contract and tort claims seeking to prevent data scraping are preempted by the Copyright Act. The decision could have significant implications for companies seeking to enforce terms of use prohibiting the scraping of data to train generative AI tools.

Training generative AI tools requires large amounts of data, and data scraping technology is often used to meet this demand. While courts across the United States grapple with the question of how such technology intersects with the existing copyright regime, many companies have updated their terms of use to prohibit such activity.

In 2023, X Corp. sued Bright Data to enforce such terms of use and assert state contract and tort claims. Bright Data sells datasets comprised of data scraped off social media sites as well as tools that enable customers to scrape data themselves. X sought to prevent Bright Data from copying the publicly available user data from X's social media platform and from selling tools that would allow third parties to do the same.

In May, the court dismissed X's claims, holding that they were preempted by the Copyright Act. The Court explained that according to X's User Terms, X users "own [their] Content" and grant X a nonexclusive license that allows X to make their content publicly available. This nonexclusive license does not allow X to exclude others from reproducing X users' content and the court found that the Copyright Act preempts state law claims granting rights that are "equivalent to any of the exclusive rights within the general scope of copyright." The court reasoned that allowing X's claims to proceed would "entrench its own private copyright system that rivals, even conflicts with, the actual copyright system enacted by Congress." Accordingly, the court found that the extent to which public data may be freely copied from social media platforms, including via data scraping, should be governed by the Copyright Act and not by conflicting contract provisions.

The decision has far-reaching implications for any company trying to protect third-party data through contractual provisions. Nothing in the court's logic limits the application of this case to just data scraping prohibitions. Indeed, any contractual provision that limits the use of third-party data – which the company does not own or to which it does not have an exclusive license – appears to be at risk of preemption under the Copyright Act. Given the existing uncertainty over the applicability of the copyright regime to AI training, this decision further increases the risk and uncertainty companies face when seeking to ensure data on their platforms is not used to train AI tools.

Summer Associate Naomi Zhao contributed to this update.

If you have any questions regarding the matters covered in this publication, please reach out to any of the lawyers listed below or your usual Davis Polk contact.

David Lisson

+1 650 752 2013
david.lisson@davispolk.com

Matthew J. Bacal

+1 212 450 4790
matthew.bacal@davispolk.com

This communication, which we believe may be of interest to our clients and friends of the firm, is for general information only. It is not a full analysis of the matters presented and should not be relied upon as legal advice. This may be considered attorney advertising in some jurisdictions. Please refer to the firm's privacy notice for further details.