2023 | VOL. 40 | NO. 2

Inside

A publication of the Corporate Counsel Section of the New York State Bar Association

Chartering Your Corporation's Course: Reducing Litigation Risk When Amending a Corporate Charter With Multiple 'Classes' of Stock

Privilege, Waiver, and Good Faith: False Claims Act Defense After *Schutte*

Scraping the Surface of Trade Secrets: Data Scraping and Artificial Intelligence

Scraping the Surface of Trade Secrets: Data Scraping and Artificial Intelligence

By Jonathan E. Barbee

Earlier this year, X (formerly known as Twitter) sent Meta a cease-and-desist letter accusing it of hiring former Twitter employees and using those employees to learn the trade secrets behind Twitter's platform.[1] Meta was developing its new Threads platform, which would compete directly with Twitter. The focus of the accusations was Meta's alleged misappropriation of Twitter's trade secrets through those former employees, but the letter also accused Meta of scaping data about Twitter followers. While the letter did not classify the Twitter follower data as trade secrets, or indicate that Meta's alleged data scraping would rise to the level of misappropriation of trade secrets, it showed that data scraping and the misappropriation of trade secrets can have a common nexus and arise from similar scenarios.

"Data scraping," or "web scraping," is a process where a computer program—such as an artificial intelligence (AI) program—collects data from the web or another source. AI tools, such as ChatGPT, are designed around the concept of data scraping in many ways because the artificial intelligence in a tool like ChatGPT derives from collecting information and data from outside sources, especially information available on the internet. The data being scraped can reside in the recesses of the internet or deep in a document repository, which is why data scraping can provide access to large volumes of information that humans would not be able to mine manually.

In some ways, it is not surprising that X did not allege that the scraping of Twitter follower data misappropriated its trade secrets. After all, for data to be considered a trade secret, it must be kept confidential, and Twitter follower data appears to be publicly-available. But, with slightly different facts, where permission was needed to access the Twitter follower data, or the data was not completely available to the public, then scraping Twitter follower data could have stronger implications for misappropriation.

Back in 2020, before ChatGPT caught the world's attention, the Eleventh Circuit grappled with the issue of data scraping and trade secrets. Specifically, the Eleventh Circuit faced the question of whether scraping publicly-available data on the internet could lead to the misappropriation of trade secrets. In that case, *Compulife Software Inc. v. Newman*, 959 F.3d 1288 (11th Cir. 2020), Compulife, a company that produced a database of life insurance quotes, sued a competitor for misappropriating trade secrets and violating a state anti-hacking statute, among other things. While Compulife's database was available to the public, individual users were only able to retrieve individual quotes from the database—with the database structured that way, it would simply take too long for a person to retrieve every quote in the database.[2] In that sense, the public did not have access to the *entire* database—the public only had access to the individual set of quotes requested. Even though the database was technically available to the public, it could not practically be recreated or copied in its entirety by any single person.

The twist in the *Compulife* case arose because the defendant used a data scraping computer program—known as a "bot" and similar to an AI tool—to request over 40 million quotes from Compulife's database.[3] The Eleventh Circuit explained, "Although Compulife has plainly given the world implicit permission to access as many quotes as is *humanly* possible, a robot can collect more quotes than any human practicably could."[4] "So," the Eleventh Circuit continued, "while *manually* accessing quotes from Compulife's database is unlikely ever to constitute improper means, *using a bot to collect an otherwise infeasible amount of data may well be*."[5] Based on that reasoning, the Eleventh Circuit concluded that a data scraping program could misappropriate trade secrets by copying an entire database of information, even when that database is technically available to the public.[6] It is far from clear whether other courts will follow the same rationale, but the Eleventh Circuit established one principle that other courts may consider: Even if an AI tool *can* access data, that does not mean that the AI tool *should* access that data.

As AI tools continue to evolve and become widely adopted, data scraping is bound to become more prevalent, which is likely to lead to an increase in data scraping litigation. In order to prepare for this heightened scrutiny of data scraping, in-house counsel should consider the following practical tips when using AI tools in their organizations.

1. Use Caution With Internal AI Tools

By virtue of having their own intelligence, AI tools have a "mind of their own." When using internal AI tools, such as AI programs that are directly connected through a company's network to internal databases and work product created by employees, it is possible that internal AI tools will have access to confidential and proprietary information. Some of

that internal information may be considered a company's trade secrets. As part of their data scraping routines, internal AI tools could potentially scrape a company's own trade secrets and then disseminate those trade secrets in a way that compromises the protection of those trade secrets. For that reason, in-house counsel should take precautions to ensure that AI tools are not inadvertently scraping trade secrets from internal company sources. This could arise, for example, if an internal AI tool was asked to collect information from a company's databases for a client alert and, in doing so, included some of the company's trade secrets in its collection.

2. Know How AI Tools Are Trained

AI tools are trained by humans and can also be programmed to train themselves. When using AI tools, in-house counsel should be aware of whether those AI tools are being trained to scrape data and, if so, which sources are being scraped for information—including sources that may contain trade secrets. For example, in-house counsel should understand whether AI tools will scrape employees' personal computers and devices for data when connected to a company's network and whether any scraped data will be shared by AI tools with third parties outside of a company. Controlling how AI tools are trained can help in-housel counsel prevent AI tools from scraping data that could lead to issues, such as the inadvertent disclosure of trade secrets.

3. Review Licenses Carefully

In-house counsel should review the licenses for any databases and repositories that may be accessed by AI tools to ensure that licenses will not be violated by data scraping. In the same vein, in-house counsel should confirm that AI tools have the proper permissions and consent to collect data from sources being targeted by the AI tools.

4. Track the Activities of AI Tools

In-house counsel should monitor the activities of AI tools so that they can intervene if an AI tool starts collecting data that could be problematic. Even conducting audits of AI tools could be helpful in preventing AI tools from gathering data from sources that may contain trade secrets or other confidential and proprietary information. In-house counsel should work with their IT departments to install programs and systems that can monitor the activities of AI tools, such as forensics software that can track AI tools. Having a record of what AI tools have done, and where AI tools have scraped data, can be useful in the event of an incident, especially if the incident leads to an investigation.

5. Be Mindful of Anti-Hacking Statutes

AI tools, if left unchecked, can potentially engage in activities that could violate hacking statutes. For example, the Computer Fraud and Abuse Act (CFAA) is a federal statute that imposes civil and criminal liability for improperly accessing another person's computer without authorization.[7] AI tools that scrape data could potentially run afoul of the CFAA by collecting data from third-party computers and networks without the proper permissions. When data scraping flirts with hacking, in the sense that the scraping involves some sort of improper access to a third party's computer or network, then anti-hacking statutes like the CFAA may be implicated.

Jonathan E. Barbee is of counsel at MoloLamken LLP with a focus on intellectual property and technology-related litigation. He represents inventors, innovators, startups, and research institutions, both as plaintiffs and defendants. With a degree in electrical engineering, Mr. Barbee litigates complex patent, trade secrets, trademark, and copyright matters across an array of technologies and industries, including the high tech, medical devices, and pharmaceutical industries. In addition to being a member of the Executive Committee of NYSBA's Corporate Counsel Section and editor of *Inside*, Mr. Barbee is also an active member of the Trade Secret Committee of the New York Intellectual Property Law Association, the Equity, Diversity, and Inclusion Committee of the Association of University Technology Managers (AUTM), and the Board of Directors of Community Impact at Columbia University.

Endnotes

1. Hannah Murphy, *Twitter Threatens Trade Secrets Lawsuit Over Meta's Threads App, Financial Times* (July 6, 2023), *available at* https://www.ft.com/content/73c32acc-597e-47ee-9745-e7cfa80fbcdf.
2. *Id.* at 1314.
3. *Id* at 1300.
4. *Id* at 1314 (emphasis in original)
5. *Id.* (emphasis added).
6. *Id.*
7. 18 U.S.C. § 1030.