

# Data Scraping, Privacy Law, and the Latest Challenge to the Generative AI Business Model

JULY 17, 2023

D. REED FREEMAN JR., DAN JASNOW, EVA J. PULLIAM, CONSTANCE ZHOU

**Share This Page** [EMAIL](#) [LINKEDIN](#) [TWITTER](#) [FACEBOOK](#)

In a putative class action filed on June 28, 2023, in the Northern District of California, and in other similar cases, plaintiffs allege that OpenAI, Microsoft, and their respective affiliates violated the privacy rights of millions of internet users through the large-scale scraping of their personal data from social media, blog posts, and other websites, and using those data to train machine learning models.

Furthermore, the plaintiffs assert that the defendants violated the Computer Fraud and Abuse Act (CFAA) by intentionally accessing protected computers without authorization and obtaining information through ChatGPT plug-ins integrated across various platforms and websites.

The plaintiffs also allege that the defendants failed to adequately disclose that users' data may be used to train machine learning models and generative AI tools.

Likely the most significant allegation revolves around data scraping. The implications could be profound if plaintiffs prevail on their theory that consent is necessary for large-scale scraping of personal data from the internet for use in training AI tools. It's difficult to imagine how AI companies would obtain retroactive consent from millions of users who have already provided their data to websites.

## Other Pending Lawsuits

These new data scraping lawsuits emerge amidst a surge of other pending litigation challenging the data scraping activities of generative AI tools.

The other cases often center on copyright infringement or license-based claims. For example, anonymous coders filed claims against the AI-assisted coding tool Copilot in the Northern District of California for alleged breaches of open-source software licenses and the Digital Millennium Copyright Act.

In early 2023, stock photo provider **Getty Images** sued Stability AI, a smaller AI start-up in Delaware federal court, alleging the illegal use of its photos to train an image-generating bot.

In Europe, under the General Data Protection Regulation (GDPR), organizations must obtain explicit consent from EU citizens before collecting or processing their personal data. This stipulation may create difficulties for AI models that rely on scraped data, as they must ensure proper consent is acquired, particularly when managing personal data. Furthermore, the ban on ChatGPT in Italy, although later lifted, highlights the growing concerns about AI's compliance with data protection regulations. These legal disputes underscore the unresolved issues for AI developers and their users regarding the legality of scraping and utilizing data from the public internet to train AI tools like ChatGPT.

## Implications

As the legal landscape surrounding artificial intelligence rapidly evolves, the question of whether generative AI tools can lawfully use data from the public internet for training purposes remains unanswered. The increasing legal challenges faced by generative AI tools emphasize the complexities and unresolved issues that artificial intelligence brings to the forefront of legal and ethical debates. Some key concerns include:

- **Data privacy:** Generative AI tools often use data from the public internet, including social media posts, blogs, and other user-generated content, to train their algorithms. This raises concerns about the privacy rights of individuals if their data has been used without their consent, potentially leading to violations of data protection laws and regulations.
- **Licensing and open-source issues:** AI tools may use open-source software or data with specific licensing requirements, which can result in legal disputes if these requirements are not met or are violated during the AI development process.
- **Intellectual property:** AI-generated content can closely resemble or even reproduce copyrighted material, raising questions about copyright infringement. Additionally, the use of copyrighted data in training AI models could lead to legal claims regarding the unauthorized use of such material.

Legislators and regulators around the world are rushing to impose new legal regimes around artificial intelligence. In the meantime, as these new cases show, a large body of existing law may impose guardrails around it. How the most transformative technology in a generation will be shaped is at stake.

Contacts

---



---

**D. Reed Freeman Jr.**

PARTNER



---

**Dan Jasnow**

PARTNER AND AI, METaverse & BLOCKCHAIN  
INDUSTRY GROUP CO-LEADER



---

**Eva J. Pulliam**

PARTNER



---

**Constance Zhou**

ASSOCIATE

---

### **Related Industries**

AI, Metaverse & Blockchain

### **Related Practices**

Privacy, Data Protection & Data Security

## **Continue Reading**

---

ALERTS

### **New York's Fashion Sustainability and Social Accountability Act – Round 2 Underway**

JULY 17, 2023 | ANTHONY V. LUPO, ANGELA M. SANTOS, R. ERICA ROQUE, FELICIA A. XU,  
EMILY B. LEWIS

ALERTS