Farella Braun + Martel

ACC Association of Corporate Counsel
SAN FRANCISCO BAY AREA

AI & Data Scraping: Copyrights, Contracts & Other Legal Risks

Alex Reese, Partner, Farella Braun + Martel

Janel Thamkul, Deputy General Counsel, Anthropic

September 19, 2024

This presentation is provided for informational purposes and does not constitute legal advice.

Agenda

- What is scraping?
  - Historical uses for scraping
  - Who scrapes? How AI companies have affected scraping activity
- The legal claims typically used to combat scraping
- Recent cases that affect the legal landscape
  - Copyright: Ongoing lawsuits; fair use theories
  - Contract: *Meta v. Bright Data* and *X v. Bright Data*
  - Computer Fraud & Abuse Act: *hiQ v. LinkedIn* and *Van Buren*
- Tips for responsible scraping, and tips for responding to scraping
- Q&A

Farella
Braun + Martel

What is scraping?

- Scraping is the automated collection of data from online sources.
- Who scrapes? Almost everyone.
  - Market research.
  - Pricing analysis.
  - Social media advertising / monitoring
- Many companies also rely on scraped data
  - AI models and AI tools are trained on scraped data

Farella
Braun + Martel

What is scraping?

- Historically: Scraping focused on raw data, such as prices, number of social media followers, CV data, etc.
  - Whether the data is publicly available or behind a password wall affects risk
- Today:
  - Growth of AI companies has changed both the quantity and types of data needed
  - Discussion

Farella Braun + Martel

Overview: Legal Claims that Often Apply to Scraping

- Copyright claims
  - Infringement
  - Anti-circumvention
- Breach of contract: Most websites have terms of service prohibiting scraping.
  - Consider: Enforceability of terms unilaterally posted (browse-wrap agreements).
- Computer Fraud and Abuse Act (18 U.S.C. sec. 1030): Originally an anti-hacking claim; carries civil and criminal penalties.

Farella
Braun + Martel

Digging Deeper: Recent Cases on Copyright Infringement and Scraping

- Multiple ongoing lawsuits alleging copyright infringement
  - Several ongoing class actions filed by visual artists, journalists, authors, and music publishing groups.

- Primary defense in these cases is fair use, requiring analysis of:
  - (1) purpose and character of use,
  - (2) nature of the work,
  - (3) amount of work used, and
  - (4) effect on the market.

Farella
Braun + Martel

Digging Deeper: Recent Cases on Copyright Infringement and Scraping

- Fair use issues:
  - Is the use transformative?
  - Do LLMs contain copies, or use data to learn patterns?
- Fair use cases on transformative technologies
  - Consider *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021)
  - Consider *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015)

Farella
Braun + Martel

Digging Deeper: Recent Cases on Copyright Infringement and Scraping

- Another breed of copyright claim: Anti-circumvention.
  - 17 U.S.C. sec. 1201 prohibits circumventing “a technological measure that effectively controls access to a” copyrighted work.
  - Social media platforms have threatened anti-circumvention claims for scraping behind the password wall.
  - This claim has never been tested.

Farella
Braun + Martel

Digging Deeper: Recent Cases on Contract Claims and Scraping

Meta Platforms v. Bright Data, 23-cv-00077-EMC, (N.D. Cal. Jan. 23, 2024)

- Court granted Bright Data's motion for summary judgment.
- Meta's Terms of Service do not govern activity unless you are "using" (i.e. logged into) the platform.

X Corp. v. Bright Data Ltd., C 23-03698 WHA, (N.D. Cal. May. 9, 2024)

- Court granted Bright Data's motion for summary judgment.
- Meta's Terms of Service do not govern activity unless you are "using" (i.e. logged into) the platform.

Farella Braun + Martel

Digging Deeper: Recent Cases on CFAA and Scraping

- Social media platforms have alleged that scrapers violate the CFAA by:
  - Creating fake accounts
  - Using legitimate user credentials with user permission
  - Keeping publicly available data after account is made private
  - Accessing public data following cease and desist

- *hiQ Labs, Inc. v. Linkedin Corp.*, 273 F. Supp. 3d 1099, 1103 (N.D. Cal. 2017)
- *Van Buren v. U.S.,* 141 S. Ct. 1648 (2021)

Farella Braun + Martel

Tips for Responsible Scraping

- There is a spectrum of risk associated with scraping.
- Things to consider for responsible scraping:
  - Type of data being collected—likely to include personal information?
  - How the data is being used
  - Is the data publicly available?
  - Is sign-in required? Is bypassing Captcha required?
  - How to interact with Robots.txt files
  - Has your company explicitly agreed to terms of use / service?

Farella
Braun + Martel

Tips for Responsible Scraping

- If you don’t scrape directly but use AI tools:
  - Consider asking for a model card
  - Seek documentation from the developer on which data was used
  - Ask for the company’s policy on data sourcing
  - Consider contractual risk mitigation (reps/warranties; indemnification)
- Receiving a cease and desist alters the risk calculus

Farella
Braun + Martel

Tips for Combatting Scraping

- Prominently post terms of service prohibiting scraping
  - Example: “You may not collect any data or content using automated means (such as robots, spiders, or scrapers) without prior written permission.”
  - Example: “You may not conduct automated queries (including screen and database scraping, spiders, robots, crawlers, bypassing “captcha” or similar precautions, or any other automated activity with the purpose of obtaining information from” this website.
- Consider where terms of service are posted

Farella Braun + Martel

Tips for Combatting Scraping

- Use Captchas and similar technology where appropriate
  - This can give rise to a copyright claim
- Send a powerful cease and desist
  - Quote your terms of service
  - Consider asserting anti-circumvention theory
  - CFAA theory not likely as useful as it used to be, except for private data

Farella
Braun + Martel

Questions + Contact Information

Alex Reese
Partner
Farella Braun + Martel
415.954.4914
areese@fbm.com

Janel Thamkul
Deputy General Counsel
Anthropic

Farella
Braun + Martel