The Legal Landscape of Web Scraping

I. A Scraping-Driven Economy

Data, having been deemed "the oil of the digital era," is now considered the most valuable asset on earth.[1] But, unlike oil, data is becoming *more* accessible to the general public. Although the availability of data can be attributed to numerous factors, including "the emergence of global-scale technology platforms"[2] that lower the cost of hosting data, perhaps the biggest contributor of data availability is the proliferation of web scraping technologies—that is, automated technologies that swiftly collect massive amounts of data on the web.

Countless emerging and established companies rely on web scraping to power their offerings. For example, scraped data is used to train AI technologies, to offer price comparisons between similar products, to power web-based search functions, and to help law enforcement identify wanted persons. But at the same time disruptive technologies are exploring the vast utility of scraped data, websites that host scraped data and other stakeholders are continuing to challenge the legality of scraping. While scraping is not *per se* illegal, it has risks.

In the United States, there is no single legal or regulatory framework that governs scraping. The legal regime governing scraping has been largely reactive—developing in real time as stakeholders (including websites and regulators) make claims relating to the collection and use of their data. Further complicating the legal analysis is that this analysis is often fact-intensive and turns on considerations such as the nature of the data being scraped, the origins of the data, the technology (if any) used to prevent data scraping and the technology used to scrape, and the existence and content of a website's terms of service. In other words, there is no single answer to whether a given scraping practice could be actionable.

That said, this Client Alerts offers an overview of the law affecting stakeholders in the world of data scraping—be it data scrapers or websites and individuals whose data is being collected. This Client Alert addresses some of the most common theories of liability asserted against data scrapers and purchasers of scraped data and includes key takeaways from litigation to date.

II. Overview Of Frequently-Asserted Claims

1. The CFAA

The Computer Fraud and Abuse Act ("CFAA") imposes civil and criminal liability for improperly accessing a "protected computer"—i.e., any computer connected to the internet.[3] It provides that "[w]hoever

---

[1] THE ECONOMIST, *The world's most valuable resource is no longer oil, but data*, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (May 6, 2017); *see also generally* Peter K. Yu, *Data Producer's Right and the Protection of Machine-Generated Data*, 93 TUL. L. REV. 859, 860-63 (2019) (discussing the value of big data and the factors contributing to that increased value).

[2] MCKINSEY DIGITAL, *Strategy for a digital world*, *available at* https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/strategy-for-a-digital-world (Oct. 8, 2021).

[3] 18 U.S.C. § 1030 *et. seq.*

. . . intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer . . . shall be punished."[4]

Recently, tech companies such as LinkedIn and Meta have invoked the CFAA to enjoin others from accessing data hosted on their platforms. The plaintiffs have claimed that an entity that uses web scraping to "access a computer without authorization" violates the CFAA. Although the caselaw interpreting the CFAA is rapidly evolving, a few principles can be gleaned from these cases.

First, to the extent a company is accessing "public" information, the CFAA is not likely to apply. That was the case in *hiQ Labs, Inc. v. LinkedIn Corp.*, where LinkedIn alleged that hiQ, a data-analytics start-up, violated the CFAA when it scraped publicly-available member profile pages to fuel its analytics offerings. After analyzing the CFAA's text and legislative history, the Ninth Circuit explained that "the CFAA's prohibition on accessing a computer 'without authorization' is violated when a person circumvents a computer's generally applicable rules regarding access permission, such as username and password requirements, to gain access to a computer."[5] As such, the Court held: "It is likely that when a computer network generally permits public access to its data, a user's accessing that publicly available data will not constitute access without authorization under the CFAA."[6]

Second, application of the CFAA is highly technical, often turning on specific aspects of a company's "access[]" to a "computer."[7] "Use of the computer is integral to the perpetuation of a fraud under the CFAA[.]"[8] As a result, a company may be accessing third-party *data* without accessing the *computer* on which that data originates. This distinction was critical in *Meta v. BrandTotal*, where Meta sued BrandTotal on the basis that BrandTotal's use of data originating on Meta's platforms violates the CFAA (among other claims).[9] There, one of BrandTotal's offerings was said to be using "'reactive' data collection" technology.[10] Relying on BrandTotal's expert report, the Court found that the technology was "accessing and processing the data that Meta has sent to the individuals users . . . [but was] not proactively 'accessing' or 'communicating with' Meta's servers."[11] Accordingly, the Court found that Meta could not state a CFAA claim as to that aspect of BrandTotal's technology.[12] Companies collecting data from third-parties should consider whose computer—if anyone's—they might be said to be "accessing" in analyzing whether their conduct would be actionable under the CFAA, and whether there may be other viable technologies for data collection. And companies embroiled in litigation on either side of a CFAA claim should retain sophisticated technical experts who can explain whether a company's offering amounts to "access" of a computer system.

---

[4] 18 U.S.C. § 1030(a)(2)(C); *see also Musacchio v. United States*, 136 S. Ct. 709, 713 (2016). California's Comprehensive Computer Data Access and Fraud Act ("CDAFA"), California Penal Code § 502, is a State analogue to the CFAA. While not identical statutes, the CFAA and CDAFA are often analyzed together.

[5] *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F.4th 1180 (9th Cir. 2022).

[6] *Id.* at 1201.

[7] 18 U.S.C. § 1030(a)(2)(C).

[8] *Dresser-Rand Co. v. Jones*, 957 F. Supp. 2d 610, 614-15 (E.D. Pa. 2013) ("Dresser-Rand's CFAA claim … fails to meet the basic requirement of accessing a computer. … Wadsworth may have accessed Dresser-Rand documents, but he never accessed Dresser-Rand computers, as required under the CFAA.").

[9] *See Meta v. BrandTotal Ltd.*, 605 F. Supp. 3d 1218 (N.D.Cal. June 6, 2022).

[10] *Id.* at 1260.

[11] *Id.*

[12] *Id.* at 1261.

Third, a company considering bringing a CFAA claim should consider whether the potentially adverse company is acting "without authorization." Many scrapers access data through the permission of third-parties (i.e., their own clients) who provide login credentials. Platform owners often attempt to prevent this practice by seeking to revoke a company's authorization, including through a cease-and-desist letter. In these cases, the question arises whether a computer "owner"—i.e., the website or app developer—may revoke access granted by an authorized user (i.e., an account holder) such that a company's further access to that data may be a CFAA violation. The Ninth Circuit, at least, has held that it can.[13]

2. Copyright Infringement and the Digital Millennium Copyright Act

To varying degrees of success, website owners have asserted claims against scrapers for copyright infringement, including for violations of the Digital Millennium Copyright Act ("DMCA").[14]

A plaintiff alleging copyright infringement "need only allege (1) ownership of a valid copyright and (2) copying of original elements of the work."[15] Claims asserting copyright infringement based on scraping often turn on whether a plaintiff has any rights vis-à-vis the copyrighted (*i.e.,* scraped) work. Although online platforms often advise their members that the data members post belongs to those members, a platform may nevertheless be able to assert a copyright violation if the information scraped extends *beyond* member data. For example, in *Facebook v. Power Ventures*, a judge in the Northern District of California explained that, although the information Power Ventures intended to extract from Facebook was user data, if they "first have to make a copy of a user's entire Facebook profile page in order to collect that user content, such action may violate Facebook's proprietary rights."[16] Even if a scraper targets only data posted by a user, a website host may be able to assert a claim of copyright infringement if that website can claim—through its terms of use or otherwise—that it has an exclusive ownership interest over that work.[17]

Some data hosts have capitalized on the DMCA to stop scrapers. The DMCA, like the CFAA, provides for both criminal and civil liability[18] and includes provisions that scraping may trigger. For example, the DMCA prohibits any person from "circumvent[ing] a technological measure that effectively controls access to" a copyrighted work.[19] The DMCA defines circumvention to include "avoid[ing], bypass[ing], remov[ing], deactivat[ing], or impair[ing] a technological measure, without the authority of the copyright owner."[20] Thus, scrapers who employ algorithms to circumvent a website's technological barriers (such as CAPTCHA and limitations on access rates) that are intended to exclude bots may find themselves facing a DMCA claim.

---

[13] *See id.* at 1267 ("Once Meta revoked BrandTotal's continued use of its various programs to actively collect data while panelists were logged into Facebook—which it had the power to stop, but did not before February of 2021—it violated the CFAA."); *see also LVRC Holdings LLC v. Brekka*, 581 F.3d 1127, 1135 (9th Cir. 2009) ("[A] person uses a computer 'without authorization' under [the CFAA] … when the employer has rescinded permission to access the computer and the defendant uses the computer anyway."); *see also United States v. Nosal*, 844 F.3d 1024, 1029 (9th Cir. 2016) (same); *see also Rimini St., Inc. v. Oracle Int'l Corp.*, 473 F. Supp. 3d 1158, 1183 (D. Nev. 2020) ("While Rimini is correct in stating that an Oracle licensee may designate a third party to act as an agent and download the files on its behalf, pursuant to the support website's terms of service, Oracle still retains the right to terminate access.").

[14] 17 U.S.C. § 1201 *et seq*.

[15] *Facebook, Inc. v. Power Ventures, Inc.*, 2009 WL 1299698, at *3 (N.D. Cal. May 11, 2009).

[16] *Id.* at *4.

[17] *Craiglist v. 3Taps Inc.*, 942 F. Supp. 2d 962, 973 (N.D. Cal. 2013).

[18] *Couponcabin LLC v. Savings, Inc.*, 2016 WL 3181826 (N.D. Ind. June 8, 2015).

[19] 17 U.S.C. § 1201(a)(1)(A).

[20] *Id.* § 1201(a)(3)(A).

Relatedly, the DMCA prohibits companies from offering (even if not using directly) a technology that may be used to circumvent technological measures that are intended to protect copyrightable data.[21] A company that offers scraping software and does not itself scrape may nevertheless face exposure under the DMCA if the data scraped by its clients includes copyrightable works.

Like the CFAA, one's exposure under the DMCA often turns on a fact-intensive, technical analysis. The question often becomes whether a given technological measure "effectively controls access to a work,"[22] and courts considering this question often impose a high bar on plaintiffs to show that they had effective technological barriers protecting copyrightable works. For example, in *Couponcabin LLC v. Savings, Inc.*, Couponcabin alleged that Savings violated the DMCA by scraping its website. A Northern District of Indiana court held that Couponcabin failed to plead that its work was "effectively controlled" by a technological measure because, "even after the Plaintiff's implementation of 'technological safeguards and barriers,' its website remains accessible to users of servers and/or internet service providers that have not been blocked by Plaintiff's technology."[23] Quoting a Sixth Circuit decision, the court noted that "[j]ust as one would not say that a lock on the back door of a house 'controls access' to a house whose front door does not contain a lock . . . it does not make sense to say that [§ 1201(a)] of the DMCA applies to otherwise-readily-accessible copyrighted works."[24]

The DMCA also prohibits "intentionally remov[ing] or alter[ing] any copyright management information."[25] Copyright management information, or "CMI", is information "conveyed in connection with a work" that "inform[s] the public that something is copyrighted in order to prevent infringement."[26] CMI includes, *inter alia*, "[t]he title and other information identifying the work, including the information set forth on a notice of copyright," "the name of, and other identifying information about, the author of a work," and "[t]erms and conditions for use of the work."[27] In the scraping context, this provision may be implicated where copyrighted works are scraped and copied in a manner that does not include the associated CMI. Cases alleging a violation of this provision often turn on whether the CMI is in fact "conveyed in connection with a work," with courts disagreeing as to how close in proximity the CMI must be to the copyrightable work.[28]

In summary, websites that seek to maintain copyright protection over data they host should take steps to ensure that they have copyright interests in the data they seek to protect and that their technological barriers sufficiently protect those works.[29] For their part, scrapers should consider whether the data they are targeting is subject to the protections of copyright law.

3. Breach of Contract

Many websites that host data have terms of service governing data access and usage. Website operators have sued and threatened to sue scrapers and those who purchase scraped data for breach of contract, pointing

---

21 *Id.* § 1201(a)(2), (b)(1).

22 *Id.* § 1201(a)(3)(A).

23 *Couponcabin*, 2016 WL 3181826, at *6.

24 *Id.* (quoting *Lexmark Int'l, Inc. v. Static Control Components, Inc.*, 387 F.3d 522, 547 (6th Cir. 2004) (alteration in *Couponcabin*)).

25 17 U.S.C. § 1202.

26 *FurnitureDealer.Net, Inc. v. Amazon, Inc.*, 2022 WL 891473, at *19 (D. Min. March 25, 2022).

27 17 U.S.C. §1202(c).

28 *See FurnitureDealer.Net*, 2022 WL 891473, at *19 (discussing diverging approaches).

29 This includes obtaining copyright registration of such works before suing for infringement.

to the website's terms of service that often purport to limit the manner in which data on the website can be accessed and used. For example, in the *BrandTotal* case discussed above, Meta prevailed on a claim that BrandTotal breached its contract with Meta by collecting data from Facebook and Instagram via automated technology in violation of Meta's terms of use.[30]

A threshold question is whether the scraper is bound by a platform's terms of service. Although this question often calls for a fact-specific inquiry, generally, "browsewrap agreements" that purport to bind any visitor to a website simply by visiting that website are deemed unenforceable.[31] Under Ninth Circuit precedent, "[u]nless the website operator can show that a consumer has actual knowledge of the agreement, an enforceable contract will be found based on an inquiry notice theory only if: (1) the website provides reasonably conspicuous notice of the terms to which the consumer will be bound; and (2) the consumer takes some action, such as clicking a button or checking a box, that unambiguously manifests his or her assent to those terms."[32] Thus, courts have found "clickwrap agreements" that "requir[e] a computer user to 'consent to any terms or conditions by clicking on a dialogue box on the screen in order to proceed with [a] . . . transaction,'" to be enforceable.[33]

4. Additional Common Law Claims

In addition to breach of contract claims, website hosts often sue those engaged in scraping for common law claims of trespass to chattels and unjust enrichment .

Although the contours of a trespass to chattels claim vary by state, a plaintiff alleging trespass to chattels in the web scraping context generally must allege that a defendant accessed its computer system without authorization and caused damage. These claims often turn on whether a plaintiff can allege that a scraper damaged its computer systems. Courts have differed on what constitutes such "damage." In some cases, a plaintiff's allegation of an increased burden on server capacity caused by a scraper's activity may constitute damage for purposes of the claim.[34]

Platforms also often allege that scrapers are unjustly enriched by scraping. The simplicity of this theory makes it appealing to a plaintiff. Generally a plaintiff alleging unjust enrichment need allege only "a receipt of a benefit and unjust retention of the benefit at the expense of another."[35] Unjust enrichment generally does not require proving information beyond that which a plaintiff attempts to prove with statutory-based claims, and thus often rises and falls with statutory claims based on the same conduct.[36]

---

[30] *BrandTotal*, 605 F.Supp. at 1258.

[31] *See, e.g.*, *Berman v. Freedom Fin. Network, LLC*, 30 F.4th 849, 856 (9th Cir. 2022).

[32] *Id.*

[33] *See, e.g.*, *Hancock v. Am. Tel. & Tel. Co.*, 701 F.3d 1248, 1255 (10th Cir. 2012) (quoting *Feldman v. Google, Inc.*, 513 F. Supp. 2d 299, 236 (E.D. Pa. 2007)); *see also Valelly v. Merrill Lynch, Pierce, Fenner & Smith Inc.,* 464 F. Supp. 3d 634, 640 (S.D.N.Y. 2020) ("Although a clickwrap agreement's terms and conditions must be clear and conspicuous, they need not all be simultaneously and immediately visible; the terms may be binding and enforceable even if they are only accessible through a hyperlink.").

[34] *See Snap-on Bus. Sols. Inc. v. O'Neil & Assoc., Inc.*, 708 F. Supp. 2d 669, 679 (N.D. Ohio 2020).

[35] *See, e.g.*, *Prakashpalan v. Engstrom, Lipscomb & Lack*, 223 Cal. App. 4th 1105, 1132 (2014).

[36] *See, e.g.*, *In re Clearview AI, Inc., Consumer Priv. Litig.*, 585 F. Supp. 3d 1111, 1131 (N.D. Ill. 2022), *clarified on denial of reconsideration*, No. 21-CV-0135, 2022 WL 2915627 (N.D. Ill. July 25, 2022) ("Because plaintiffs have adequately [alleged] that the Clearview defendants used the benefit of their biometric data without paying them for its value, the Court denies the Clearview defendants' motion to dismiss the Virginia unjust enrichment claim.").

III. Privacy Considerations In Data Scraping

Web scrapers and those who host or rely on scraped personal data also should be aware of all applicable privacy regulations governing their activity and seek legal advice to ensure that they are complying with these regulations. There has been a surge in government investigations, enforcement actions, and class action lawsuits challenging data collection practices under existing laws, which is expected to continue with the passage of new laws and the proliferation of data collection.

The European Union's General Data Protection Regulation ("GDPR") imposes a far-reaching regime to protect personal information of individuals within E.U. member states and purports to apply to all companies processing data from European data subjects, regardless of where the companies are based.[37] In particular, the GDPR covers overseas organizations that satisfy one or both of two tests: (1) "the offering of goods or services" in Europe, or (2) "the monitoring of" behavior within Europe, even if the organizations are not established within the E.U. and do not process data there. Common examples of such covered organizations may include online retailers that target European consumers by using a local language and entities that price goods and services in a local European currency.[38]

One of the most widely-publicized cases of data scraping involves Clearview AI, a company offering facial recognition software that relied on billions of facial images scraped from the internet. Last year the Italian regulator Italian SA filed Clearview AI €20 million after it was found to be selling its database of billions of scraped facial images to other businesses.[39] The Italian SA explained that "the company infringed several fundamental principles of the GDPR including transparency—as it failed to adequately inform users—purpose limitation—as it processed users' data for purposes other than those for which they had been made available online— and storage limitation—as it did not set out any data storage period."[40]

U.S.-based companies or those conducting business in the U.S. may find themselves subject to other regulatory regimes. For example, the California Consumer Privacy Act ("CCPA"), signed into law in 2018, is the most comprehensive data privacy regulation in the United States. In November 2020, Californians approved the California Privacy Rights Act ("CPRA") that expands the CCPA. Most of the CPRA's provisions went into effect on January 1, 2023, with a look-back to January 2022. In pertinent part, the CCPA gives consumers (defined as California residents) the right to know what personal information is collected and stored and to demand that businesses delete that personal information.

Similar legislation and regulations seek to protect the privacy of biometric data. For example, the Illinois Biometric Information Privacy Act, or BIPA, is an Illinois statute that regulates the collection, use, retention, and destruction of individuals' biometric identifying information, such as fingerprints, retina scans, and facial geometry scans.[41] BIPA applies broadly to any private entity that operates or does business in Illinois (regardless of whether the entity is headquartered in Illinois). In addition to facing exposure under the GDPR,

---

37 Art. 32 GDPR.

38 *See* GDPR Recitals 23-24.

39 GPDP, Press Release, *Facial recognition: Italian SA fines Clearview AI euro 20 million Bans use of biometric data and monitoring of Italian data subjects*, https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751323#english (last visited Apr. 24, 2023).

40 *Id.*

41 740 ILCS 14/10.

the ACLU sued Clearview AI under BIPA, and the parties settled with a consent order that banned Clearview AI from offering its faceprint database to most private businesses.[42]

The requirements that these regulatory regimes impose on companies depend on the companies' classification—*e.g.*, as a data "processor", "controller", or "joint controller" (in the case of the GDPA) or as a "business" or "service provider" (in the case of the CCPA)—a question which itself is fact-specific and not always straight-forward, particularly for companies that rely on numerous data sources and have multiple offerings. In those cases, a company may be situated differently vis-à-vis privacy protection regulations for different aspects of its business.

***

If you have any questions about the issues addressed in this memorandum, or if you would like a copy of any of the materials mentioned in it, please do not hesitate to reach out to:

Corey Worcester
Email: coreyworcester@quinnemanuel.com
Phone: + 212 849 7471

Renita Sharma
Email: renitasharma@quinnemanuel.com
Phone: + 212 849 7413

Hope Skibitsky
Email: hopeskibitsky@quinnemanuel.com
Phone: + 212 849 7535

Zane Muller
Email: zanemuller@quinnemanuel.com
Phone: + 212 849 7302

To view more memoranda, please visit www.quinnemanuel.com/the-firm/publications/
To update information or unsubscribe, please email updates@quinnemanuel.com

[42] *See ACLU v. Clearview AI, Inc.*, Case No. 2020 CH 04353, Consent Order of Permanent and Time-Limited Injunctions Against Defendant Clearview AI, Inc., *available at* https://www.aclu.org/sites/default/files/field_document/signed_consent_order_5.11.22.pdf.