

Circumscribing the Spider: Trademark Law and the Edge of Data Scraping

Michael P. Goodyear*

INTRODUCTION

In the twenty-first century, data is an incredibly valuable commodity. Past generations were powered by gold, oil, and other natural resources, but in the Internet Age, data is one of the most valuable assets for companies, fueling corporate behemoths like Google.¹ Like with gold and oil, companies can mine—or “scrape”—data to accumulate wealth and power, as well as knowledge. Web scraping, also referred to as data scraping, spidering, or just plain scraping, is the extraction of online data for purposes such as staying informed on competitor prices, measuring customer sentiment, selling data, and producing competing products.² Most people have encountered services strongly influenced by scraped data, such as when reviewing a flight comparison website or purchasing e-commerce products whose prices are set by benchmarking them against scraped competitors’ prices.³ In the era of data, web scraping is a

* J.D., University of Michigan Law School (2020); A.B., University of Chicago (2016). Thank you to the *Kansas Law Review* editors for all of their efforts in bringing this work to print.

1. *Data Valuation: Understanding the Value of Your Data Assets*, DELOITTE (2020), <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Finance/Valuation-Data-Digital.pdf> [<https://perma.cc/F3Y5-WR82>]; Bennett Cyphers, *Google Says It Doesn’t ‘Sell’ Your Data. Here’s How the Company Shares, Monetizes, and Exploits It*, ELEC. FRONTIER FOUND. (Mar. 19, 2020), <https://www.eff.org/deeplinks/2020/03/google-says-it-doesnt-sell-your-data-heres-how-company-shares-monetizes-and> [<https://perma.cc/X9WX-985F>]; *How Do Tech Companies Make Money From Our Personal Data?*, FORBES (June 27, 2019, 11:20 AM), <https://www.forbes.com/sites/quora/2019/06/27/how-do-tech-companies-make-money-from-our-personal-data/?sh=20dded384788> [<https://perma.cc/VE5W-RQZZ>].

2. See Eran Halevy, *Once Only for Huge Companies, ‘Web Scraping’ Is Now an Online Arms Race No Internet Marketer Can Avoid*, ENTREPRENEUR (Apr. 20, 2018), <https://www.entrepreneur.com/article/311261> [<https://perma.cc/R43M-XKCJ>].

3. See Colm Kenny, *Web Scraping*, ZYTE, <https://www.zyte.com/learn/what-is-web-scraping> [<https://perma.cc/5XN9-2LW6>] (last visited Oct. 6, 2021).

prodigious business.⁴

Despite its ubiquity in the market, however, scraping is also a practice that many companies want to prevent.⁵ Commentators have referred to this phenomenon as an “arms race” where companies both try to scrape competitors’ data and prevent those same competitors from scraping their own data.⁶ In these arms races, companies use technological strategies to block scraping or trick competitors into scraping incorrect information.⁷ However, companies have also taken to the courts to prevent scraping. Industry leaders such as LinkedIn⁸ and Southwest Airlines⁹ have sued scrapers to protect their data. Such anti-scraping litigation has used a wide variety of legal claims, including trespass to chattels, copyright infringement, violation of the Computer Fraud and Abuse Act (“CFAA”), and breach of terms of use.¹⁰

Trademark law is one such anti-scraping claim that has been underexamined in scholarship and underutilized in the courts. Scholarly literature on scraping has so far focused on the nature of scraping¹¹ or how to restrict the act of scraping.¹² In particular, several of these articles have

4. Greg Elmer, *Scraping the First Person*, in COMPROMISED DATA: FROM SOCIAL MEDIA TO BIG DATA 112, 113 (Ganaele Langlois, Joanna Redden & Greg Elmer eds., 2015).

5. Klint Finley, ‘Scraper’ Bots and the Secret Internet Arms Race, WIRED (July 23, 2018, 7:00 AM), <https://www.wired.com/story/scraper-bots-and-the-secret-internet-arms-race> [<https://perma.cc/XM8S-ZM6D>]; Halevy, *supra* note 2.

6. Finley, *supra* note 5; Halevy, *supra* note 2.

7. Finley, *supra* note 5; Halevy, *supra* note 2.

8. See hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985 (9th Cir. 2019), *vacated*, No. 19-1116, 2021 WL 2405144 (U.S. June 14, 2021).

9. See Sw. Airlines Co. v. Roundpipe, LLC, 375 F. Supp. 3d 687 (N.D. Tex. 2019).

10. See Kathleen C. Riley, *Data Scraping as a Cause of Action: Limiting Use of the CFAA and Trespass in Online Copying Cases*, 29 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 245, 265 (2018) (describing some of the most common anti-scraping claims).

11. See Jeffrey Kenneth Hirschey, *Symbiotic Relationships: Pragmatic Acceptance of Data Scraping*, 29 BERKELEY TECH. L.J. 897 (2014) (discussing the legal landscape of scraping prior to the decision in *Craigslist v. 3Taps* and suggesting that data hosts and scrapers could work more cooperatively together); Vlad Krotov, Leigh Johnson & Leiser Silva, *Tutorial: Legality and Ethics of Web Scraping*, 47 COMM’NS ASS’N INFO. SYS. 555 (2020) (discussing the law and ethics around data scraping); Han-Wei Liu, *Two Decades of Laws and Practice Around Screen Scraping in the Common Law World and Its Open Banking Watershed Moment*, 30 WASH. INT’L L.J. 28 (2020) (comparing foreign laws on data scraping and suggesting that the trend of “Open Banking” may lead to a convergence in scraping law practices, as well as a reduction in demand for scraping); Tess Macapinlac, *The Legality of Web Scraping: A Proposal*, 71 FED. COMM’NS L.J. 399 (proposing an amendment to the CFAA to legalize the web scraping of publicly available websites).

12. See Jacquellena Carrero, *Access Granted: A First Amendment Theory of Reform of the CFAA Access Provision*, 120 COLUM. L. REV. 131 (2020) (suggesting a legislative modification of the CFAA to better balance the First Amendment and privacy interests in scraping); Ioannis Drivas, *Liability for Data Scraping Prohibitions Under the Refusal to Deal Doctrine: An Incremental Step Toward More Robust Sherman Act Enforcement*, 86 U. CHI. L. REV. 1901 (2019) (arguing for an

focused on the CFAA's potential for preventing the scraping of publicly available information.¹³ However, following the U.S. Court of Appeals for the Ninth Circuit's 2019 decision in *hiQ Labs, Inc. v. LinkedIn Corp.*,¹⁴ holding that LinkedIn could not block the scraping of public data,¹⁴ and the Supreme Court decision in *Van Buren v. United States*, holding that the CFAA only covers unauthorized access not unauthorized use,¹⁵ the viability of CFAA claims against scraping is in doubt.¹⁶ Therefore, websites may start to search for new, innovative legal approaches to scraping. So far, the legal strategies to restrict scraping have largely focused on blocking scraping wholesale rather than limiting how scraped data is used. Trademark claims under the federal Lanham Act offer an underexamined alternative to restrict particular uses of scraped data that constitute trademark infringement, false advertising, or dilution. Until now, only one article—written in 2007, when only one case addressing

application of antitrust law and the Sherman Act “refusal to deal” doctrine to limit competitors’ monopolies over their public data by allowing scraping); J. Alexander Lawrence & Kristina Ehle, *Combating Unauthorized Webscraping*, 20 COMPUT. L. REV. INT’L 171 (2019) (discussing options for website owners to prevent scraping post-*hiQ v. LinkedIn*); Riley, *supra* note 10 (discussing the main claims against scrapers and proposing using copyright as the main vehicle for limiting scraping); Andrew Sellars, *Twenty Years of Web Scraping and the Computer Fraud and Abuse Act*, 24 B.U. J. SCI. & TECH. L. 372 (2018) (discussing CFAA civil claims’ ability to restrict scraping practices over specific phases in CFAA litigation over the past twenty years); Benjamin L. W. Sobel, *A New Common Law of Web Scraping*, 25 LEWIS & CLARK L. REV. 147 (2021) (proposing a tort of bad faith breach of terms of service to compensate for the weakening of CFAA claims after the Ninth Circuit decision in *hiQ v. LinkedIn*); Nicholas A. Wolfe, *Hacking the Anti-Hacking Statute: Using the Computer Fraud and Abuse Act to Secure Public Data Exclusivity*, 13 NW. J. TECH. & INTELL. PROP. 301 (2015) (discussing potential loopholes in the CFAA, including in relation to scraping).

13. See Carrero, *supra* note 12 (suggesting a modification of the CFAA to balance competing interests); Macapinlac, *supra* note 11 (suggesting an amendment to the CFAA to legalize scraping); Sellars, *supra* note 12 (discussing the history of CFAA claims against scraping); Wolfe, *supra* note 12 (discussing CFAA loopholes for scraping).

14. *hiQ Labs, Inc. v. LinkedIn Corp.*, 938 F.3d 985, 995 (9th Cir. 2019), *vacated*, No. 19-1116, 2021 WL 2405144 (U.S. June 14, 2021). Although *hiQ* was vacated and remanded by the Supreme Court in light of *Van Buren*, it is unlikely that the outcome will change. See Randi Singer & Michael Goodyear, *Scraping Suit Stymied after Van Buren*, WEIL INTELL. PROP./MEDIA ALERT (June 16, 2021) (“Although it did not explicitly say so, the Court’s interpretation largely mirrored that of the Ninth Circuit in *hiQ* and earlier cases.”).

15. *Van Buren v. United States*, 141 S. Ct. 1648, 1662 (2021).

16. See Camille Fischer & Andrew Crocker, *Victory! Ruling in hiQ v. LinkedIn Protects Scraping of Public Data*, ELEC. FRONTIER FOUND. (Sept. 10, 2019), <https://www EFF.org/deeplinks/2019/09/victory-ruling-hiq-v-linkedin-protects-scraping-public-data> [<https://perma.cc/4UGG-Y99L>]; Aaron R. Cooper, David Bitkower, Clifford W. Berlow & Grace C. Signorelli-Cassady, *Why the Supreme Court’s Decision in Van Buren May Be Felt Beyond Criminal Law*, JENNER & BLOCK (Dec. 11, 2020), https://jenner.com/system/assets/publications/20547/original/Supreme_Court_Decision_Van%20Buren.pdf?1607720058 [<https://perma.cc/EM4K-RJ6J>] (“A decision favoring *Van Buren* would extend that rule to the rest of the country, thereby requiring such websites to take further action to more closely monitor visitors and take other steps to revoke authorization of an offending party [for the purposes of scraping].”).

trademarks and scraping had been decided—has examined the relationship of scraping to trademark law.¹⁷ This Article hopes to significantly expand the consideration and understanding of Lanham Act claims in regard to scraping.

In Part I, this Article will provide an overview of the practice of scraping, as well as the legal claims that plaintiffs typically use in response. Part II will examine possible claims under the Lanham Act to restrict the use of scraped data, namely trademark infringement, false advertising, and dilution. Part III will then discuss the primary defense to these claims in the scraping context: nominative fair use. After setting up the legal framework and identifying the possible actions under the Lanham Act, this Article will then review the seven cases that have discussed trademark law and scraping in Part IV. Finally, this Article will look to the future in analyzing six recently filed cases involving Lanham Act claims against data scraping practices in Part V before concluding.

I. THE PRACTICE OF SCRAPING AND ITS LAWFULNESS

“Scraping” consists of using a computer program to inspect, collect, and aggregate data from different webpages.¹⁸ This method allows entities to gather publicly available online data on an enormous scale.¹⁹ Scraping is distinct from the practice of accessing websites’ underlying data structures through formalized data requests, usually via application programming interfaces (“APIs”).²⁰ Instead of using the underlying code, scraping gathers data directly from the content on the website’s page, either via the screen outputs or the HyperText Markup Language (“HTML”) code displayed by the website.²¹ In short, scraping functions as automatic web browsing. Scraping technologies also search, copy, and retrieve data from websites regardless of whether they have permission to do so.²²

17. Sean O’Reilly, *Nominative Fair Use and Internet Aggregators: Copyright and Trademark Challenges Posed by Bots, Web Crawlers and Screen-Scraping Technologies*, 19 LOY. CONSUMER L. REV. 273 (2007).

18. See Anthony J. Dreyer & Jamie Stockton, *Internet ‘Data Scraping’: A Primer for Counseling Clients*, N.Y. L.J. (July 15, 2013, 12:00 AM), <https://www.law.com/newyorklawjournal/almID/1202610687621/Internet-‘Data-Scraping’%3A-A-Primer-for-Counseling-Clients> [<https://perma.cc/83ZQ-SJCX>].

19. Drivas, *supra* note 12, at 1903; Kenny, *supra* note 3.

20. Hirschev, *supra* note 11, at 897.

21. *Id.* at 897–98.

22. O’Reilly, *supra* note 17, at 274.

Scraping strategies vary. Some companies have employees manually gather data, others create and use custom scraping scripts for each website they want to target, and yet others employ generic web scraping tools to trawl many websites for data.²³ More sophisticated companies will typically either create custom scraping algorithms for the websites they want to scrape or use premade scraping tools to target a large number of websites.²⁴

The practice of scraping is comprised of three distinct steps: (1) building the crawling path and collecting target URLs; (2) developing and employing the scraper; and (3) retaining and parsing the data.²⁵ First, one compiles a library of URLs from which data is to be collected (the crawling path).²⁶ Next, the company must build the actual web scraper, or select the premade scraping tool it wishes to use, and employ it to scrape the crawling path.²⁷ The crawler acts as the beacon for the scraper “bots,” directing them to the web pages to which data should be extracted.²⁸ The scraper then requests data from the URLs and receives an HTML file in response.²⁹ Companies often use several proxies linked to different IP addresses to access a wider variety of geographic-specific data and to distribute traffic load on the URL that is being scraped.³⁰ Finally, the company must save and parse the collected data to make it understandable and usable.³¹ Artificial intelligence (“AI”) and machine learning (“ML”)

23. Hoda Raissi, *What Is Web Scraping?*, INSIDEBIGDATA (Jan. 26, 2019), <https://insidebigdata.com/2019/01/26/what-is-web-scraping/> [https://perma.cc/V48R-R3A6].

24. *See id.*

25. *See* Julius Cerniauskas, *The New Beginnings of AI-Powered Web Data Gathering Solutions*, TOWARDS DATA SCI. (Aug. 4, 2020), <https://towardsdatascience.com/the-new-beginnings-of-ai-powered-web-data-gathering-solutions-a8e95f5e1d3f> [https://perma.cc/WX8Y-BNE5] (arguing that there are four discreet steps, but the proxy step is not always followed); Tony Paul, *Price Comparison on Amazon: How Web Scraping Helps Companies Win the E-Commerce Game*, DATAHUT (June 23, 2021), <https://www.blog.datahut.co/post/price-comparison-on-amazon-how-web-scraping-helps-companies-win-the-e-commerce-game> [https://perma.cc/83N6-TQEH] (describing scraping as a three step process wherein one (1) identifies the URLs; (2) uses scraping technology to request the HTML file; and (3) parses the data); Kenny, *supra* note 3 (identifying five key steps: (1) identify the target websites; (2) collect URLs of the relevant pages; (3) request the HTML from these pages; (4) use locators to find the data in the HTML; and (5) save the data in a structured format).

26. Cerniauskas, *supra* note 25.

27. *See id.*

28. Kenny, *supra* note 3.

29. *Id.*; Paul, *supra* note 25.

30. Cerniauskas, *supra* note 25; Janet Williams, *Comprehensive Guide to Choosing a Proxy Service for Your Web Scraper*, PROMPTCLOUD, <https://www.promptcloud.com/blog/comprehensive-guide-to-choosing-a-proxy-service-for-your-web-scraper> [https://perma.cc/3Z6C-XY45] (last visited Oct. 6, 2021).

31. Cerniauskas, *supra* note 25; *see also* Kenny, *supra* note 3.

are also being used to recognize patterns between websites and help automate data gathering to reduce the need to manually develop and manage scrapers for different URLs.³²

Scraping is incredibly important to today's companies given that much of the world's knowledge resides on the Internet.³³ Collecting these myriad sources of data is critical to transforming them into useful and useable information.³⁴ With so many websites holding publicly accessible data, scraping can glean significant insights by providing structured data from this public information.³⁵ Scraped data is used for helpful and increasingly common tasks such as, *inter alia*, monitoring competitor prices, aggregating price comparisons, displaying travel options, and offering real estate choices.³⁶

Many of the platforms we use on a regular basis employ scraping to improve or power their services. Large online retailers such as Walmart and Amazon use scraping to gather competitors' price data to inform their own product pricing.³⁷ This market-informed approach to pricing partially explains the frequent changes in prices for online retailers.³⁸ For example, Amazon will change the average product's price every ten minutes based on both scraped competition pricing information and factors such as customer shopping patterns, inventory, and profit margins.³⁹ This elastic market pricing can lead to a more competitive marketplace, benefitting consumers, but large companies can also lower prices on these loss-leaders and raise prices on less common products.⁴⁰

32. Cerniauskas, *supra* note 25.

33. Raissi, *supra* note 23.

34. *Id.*

35. Kenny, *supra* note 3.

36. *Id.*

37. See Finley, *supra* note 5 (describing how companies like Walmart and Amazon have teams dedicated to scraping); Jeffrey Dastin, *Amazon Trounces Rivals in Battle of the Shopping 'Bots'*, REUTERS (May 10, 2017, 4:48 AM), <https://www.reuters.com/article/us-amazon-com-bots-insight/amazon-trounces-rivals-in-battle-of-the-shopping-bots-idUSKBN1860FK> [<https://perma.cc/G263-DHR3>] (describing how Amazon blocked Walmart's scraping attempts in 2017); see also Alberto Cavallo, *More Amazon Effects: Online Competition and Pricing Behaviors* 7 (Nat'l Bureau of Econ. Rsch., Working Paper No. 25138, 2018), https://www.nber.org/system/files/working_papers/w25138/w25138.pdf [<https://perma.cc/7HX2-V8V9>] (describing how algorithms devised from scraped data influence Amazon's pricing strategies).

38. See Paul, *supra* note 25 (stating that "[i]n 2013, Amazon made price changes to 40 million products in a [single] day").

39. Neel Mehta, Parth Detroja & Aditya Agashe, *Amazon Changes Prices on Its Products About Every 10 Minutes – Here's How and Why They Do It*, INSIDER (Aug. 10, 2018, 10:13 AM), <https://www.businessinsider.com/amazon-price-changes-2018-8> [<https://perma.cc/92Q5-Z2KP>].

40. See *id.*

In addition to traditional online marketplaces, online shopping aggregation platforms also scrape data from different online stores to display price comparisons on their own sites.⁴¹ These price comparison websites can be a boon to customers and businesses through directing consumers to the best prices.⁴² Indeed, they can sometimes save consumers significant amounts of time and money.⁴³ There can be drawbacks, however, such as aggregators charging businesses fees to be listed, leading to not all options being shown to consumers and thus not guaranteeing that consumers will actually see the best market price.⁴⁴

Online travel companies similarly scrape data about flights, train rides, and other modes of transportation, as well as trip experiences, and present price comparisons to users on their own platform.⁴⁵ Given the complex nature of online travel bookings, the travel industry uses scraping to obtain data on hotel listings, travel options, and hotel and travel product reviews.⁴⁶ This aggregated information can provide consumers with the best (and most affordable) options for their travel needs.⁴⁷ While travel websites can offer considerable benefits to consumers and businesses, they can also pose potential problems such as hidden fees or added fees for consumers to book through them.⁴⁸

41. O'Reilly, *supra* note 17, at 274; Bhagyeshwari Chauhan, *3 Ways Popular Price Comparison Websites Gather Data*, DATAHUT (Dec. 2, 2020), <https://www.blog.datahut.co/post/how-popular-price-comparison-websites-grab-data> [https://perma.cc/64V2-GYJR].

42. Chauhan, *supra* note 41.

43. *Top 3 Benefits of Using Price Comparison Websites*, AUGUSTA FREE PRESS (Jan. 10, 2019, 9:11 AM), <https://augustafreepress.com/top-3-benefits-of-using-price-comparison-websites> [https://perma.cc/UFZ2-SWDL].

44. *Are Comparison Sites Good for Your Business?*, BUSINESS.COM (June 2, 2020), <https://www.business.com/articles/comparison-websites-business-impact> [https://perma.cc/W7RM-MWXX].

45. O'Reilly, *supra* note 17, at 273–74, 277.

46. Amrin Kareem, *Data Scraping in Travel Industry: Here's What You Need to Know*, DATAHUT (July 27, 2019), <https://www.blog.datahut.co/post/data-scraping-in-travel-industry> [https://perma.cc/R2WT-DHBP].

47. Doma Stankevičiūtė, *Scraping Flight Prices for Equal Competition*, APOPKA VOICE (Oct. 28, 2020), <https://theapopkavoice.com/scraping-flight-prices-for-equal-competition> [https://perma.cc/DWF6-3THR].

48. See Hugo Martín, *Those Hidden Resort Fees You Hate Paying? Travel Websites Fight Back*, L.A. TIMES (Aug. 27, 2019, 3:00 AM), <https://www.latimes.com/business/story/2019-08-26/booking-com-takes-on-hotel-resort-fees> [https://perma.cc/MMM2-2QSC] (discussing Booking.com's charging of fees to force hotels to be more transparent about their prices); Jacob Passy, *When You Should Use Travel Booking Sites Like Expedia and Priceline — and When You Shouldn't*, MARKETWATCH (June 1, 2017, 3:19 PM), <https://www.marketwatch.com/story/when-you-should-use-travel-booking-sites-like-expedia-and-priceline-and-when-you-shouldnt-2017-06-01> [https://perma.cc/F6JG-LMP9] (discussing how online travel agencies charge hotels large commissions when they book hotel rooms).

Scraping has also proved pivotal in the real estate industry. Scraping from public listings can allow realtors and prospective renters or buyers to evaluate different options based on size, location, price, amenities, and more.⁴⁹ Data scraping can offer significant benefits to consumers by adding more detail than is available in the raw data that websites contain, including comparing price changes over time.⁵⁰

As demonstrated by these few examples, scraping has undoubtedly revolutionized many industries. But not everyone wants to allow the practice. Many websites employ anti-scraping techniques to signal to bots that they do not want their data scraped.⁵¹ Some common techniques include IP tracking to identify and block bots, Captcha images, requiring logins, blocking crawlers with no User-Agent headers, frequently modifying the website's HTML to obfuscate changes, and even providing false information.⁵² Websites will also explicitly state in their terms of use that scraping is prohibited.⁵³ Even companies that engage in scraping will often try to limit the scraping of their own data.⁵⁴ Notably, Amazon blocked Walmart's bots in 2017 for engaging in the same tasks Amazon bots were undertaking on Walmart's website.⁵⁵ In another example,

49. Bhagyeshwari Chauhan, *Web Scraping in Real Estate: The Ultimate Tool for a Realtor*, DATAHUT (Nov. 28, 2017), <https://www.blog.datahut.co/post/web-scraping-in-real-estate-the-ultimate-tool> [https://perma.cc/9B9B-2YCP].

50. See generally Geoff Boeing & Paul Waddell, *New Insights into Rental Housing Markets Across the United States: Web Scraping and Analyzing Craigslist Rental Listings*, 37 J. PLAN. EDUC. & RSCH. 1 (2017) (describing information gaps in the rental market and how to better reflect the full U.S. rental market); João Manuel Azevedo Santos, *Real Estate Market Data Scraping and Analysis for Financial Investments* (Sept. 10, 2018) (Master's Thesis, Universidade do Porto) (SSRN) (advocating for scraping that will allow the aggregation of real estate data to present and predict market fluctuations).

51. See Josh Dzieza, *Why Captchas Have Gotten So Difficult*, VERGE (Feb. 1, 2019, 11:00 AM), <https://www.theverge.com/2019/2/1/18205610/google-captcha-ai-robot-human-difficult-artificial-intelligence> [https://perma.cc/8K9V-72GV] (explaining the changing use of Captchas by websites); Jack Nicas, *Why Can't the Social Networks Stop Fake Accounts?*, N.Y. TIMES (Dec. 8, 2020, 1:10 PM), <https://www.nytimes.com/2020/12/08/technology/why-cant-the-social-networks-stop-fake-accounts.html> [https://perma.cc/9CJ6-CLHQ] (describing social media platforms' efforts to block bots); Jiahao Wu, *5 Anti-Scraping Techniques You May Encounter*, OCTOPARSE (Aug. 16, 2021), <https://www.octoparse.com/blog/5-anti-scraping-techniques-you-may-encounter> [https://perma.cc/K628-25K3].

52. Wu, *supra* note 51; *What Are the Methods Used Against Web Scraping?*, SCRAPINGBOT (Feb. 14, 2020), <https://www.scraping-bot.io/anti-scraping-methods> [https://perma.cc/R4ZT-LMHP].

53. Jim Snell, *Use of Online Data in the Big Data Era: Legal Issues Raised by the Use of Web Crawling and Scraping Tools for Analytics Purposes*, BLOOMBERG L. (Dec. 23, 2013, 11:00 PM), <https://news.bloomberglaw.com/us-law-week/use-of-online-data-in-the-big-data-era-legal-issues-raised-by-the-use-of-web-crawling-and-scraping-tools-for-analytics-purposes> [https://perma.cc/9Q8W-GUWS].

54. Finley, *supra* note 5.

55. Dastin, *supra* note 37.

certain airline companies, like Ryanair, have adamantly refused to let any travel websites scrape their flight data.⁵⁶

When technical safeguards for their data fail, companies can (and do) pursue litigation. Companies regularly pursue several causes of action against scrapers of their data, including (1) traditional property claims such as trespass to chattels; (2) violations of the Computer Fraud and Abuse Act (“CFAA”) and state computer crimes statutes such as the California Computer Data Access and Fraud Act (“CDAFA”); (3) breach of terms and breach of contract claims under the website’s terms of use; (4) claims under the Copyright Act for copyright infringement and violations of the Digital Millennium Copyright Act (“DMCA”) for anti-circumvention; and, more recently, (5) data privacy claims under state comprehensive data privacy statutes such as the California Consumer Privacy Act (“CCPA”) and biometrics statutes such as Illinois’ Biometric Information Privacy Act (“BIPA”).⁵⁷

These anti-scraping claims have been increasingly tested over the past two decades. The first court decision to seriously address the practice of scraping was *eBay v. Bidder’s Edge* in 2000.⁵⁸ Bidder’s Edge was an auction aggregation website that scraped data from dozens of online marketplaces, including eBay.⁵⁹ The U.S. District Court for the Northern District of California found that Bidder’s Edge’s scraping amounted to an intentional interference with eBay’s possessory interest in its platform, amounting to a trespass to chattels.⁶⁰ In ruling for eBay, the Northern District of California granted an early precedent against scraping.

Intellectual property considerations were only introduced into scraping jurisprudence three years later in *Ticketmaster Corp. v. Tickets.com*.⁶¹ Tickets.com used a “web crawler” scraping system that reviewed and extracted information from Ticketmaster’s website and offered the tickets for sale on Tickets.com’s own site.⁶² When users

56. *Ryanair Ratchets Up War on Data-Scraping Travel Sites*, DECISIONMARKETING (Sept. 8, 2020, 9:06 AM), <https://www.decisionmarketing.co.uk/news/ryanair-ratchets-up-war-on-data-scraping-travel-sites> [https://perma.cc/3C3J-9LQP].

57. See Riley, *supra* note 10, at 265; Zarish Baig & Kristin L. Bryan, *Scraping By: Data Scraping Litigation Continues to Test Limits of Longstanding Data Privacy Laws*, NAT’L L. REV. (Nov. 30, 2020), <https://www.natlawreview.com/article/scraping-data-scraping-litigation-continues-to-test-limits-longstanding-data-privacy> [https://perma.cc/GT6J-HF7S].

58. See O’Reilly, *supra* note 17, at 278–79.

59. *eBay, Inc. v. Bidder’s Edge, Inc.*, 100 F. Supp. 2d 1058, 1061–62 (N.D. Cal. 2000).

60. *Id.* at 1069–71.

61. See O’Reilly, *supra* note 17, at 279.

62. *Ticketmaster Corp. v. Tickets.com, Inc.*, No. CV997654, 2003 WL 21406289, at *2 (C.D. Cal. Mar. 7, 2003).

clicked on an event to purchase a ticket, however, they would be transferred to Ticketmaster's website via a deep link.⁶³ The court held that the raw factual data of the ticketed events was not copyrightable.⁶⁴ The compilation of the ticketed events was copyrightable, but the court found the momentary storage of the scraped website pages in Tickets.com's database, while culling the data, was fair use.⁶⁵

In the years since, these various anti-scraping claims have met with mixed success. Courts have been progressively more reluctant to accept trespass to chattels arguments without proof of tangible damages.⁶⁶ In perhaps the most important data scraping case to date, *hiQ Labs, Inc. v. LinkedIn Corp.*, the U.S. Court of Appeals for the Ninth Circuit held that LinkedIn could not block scraping of its public data, significantly undermining the viability of CFAA claims.⁶⁷ The Supreme Court, in *Van Buren v. United States*, further undermined the effectiveness of the CFAA for preventing scraping when it held that the CFAA only prohibits unauthorized access, not unauthorized uses.⁶⁸ As was highlighted in *Ticketmaster*, web scraping also frequently only targets raw data, which is not protected by copyright law,⁶⁹ and scholars have also argued that at least some uses of data scraping qualify as permissible fair use or would simply be non-infringing transitory copies.⁷⁰ Yet, despite these hurdles, courts

63. *Id.* at *1–2.

64. *Id.* at *4 (“[T]he existence of the event, its date and time, and its ticket prices, are not subject to copyright.”).

65. *Id.*

66. Harry Rubin & Karolina Ebel, *Guilt-Free Web Scraping? Not So Fast*, BLOOMBERG L. (Oct. 21, 2019, 3:00 AM), <https://news.bloomberglaw.com/privacy-and-data-security/insight-guilt-free-web-scraping-not-so-fast> [<https://perma.cc/9EJH-RY9G>].

67. 938 F.3d 985, 995 (9th Cir. 2019), *vacated*, No. 19-1116, 2021 WL 2405144 (U.S. June 14, 2021). Although *hiQ* was vacated and remanded by the Supreme Court in light of *Van Buren*, it is unlikely that the outcome will change. *See supra* note 14.

68. 141 S. Ct. 1648, 1662 (2021).

69. *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 344 (1991) (“facts are not copyrightable”).

70. *See, e.g.*, Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 U.C. DAVIS L. REV. 893, 935–62 (2019) (arguing that copying journal articles to conduct text and data mining is fair use and that most copies produced during scraping are transitory copies, which are lawful under the Second Circuit's holding in *Cartoon Network LP v. CSC Holdings, Inc. (Cablevision)*, 536 F.3d 121 (2d Cir. 2008)); Edward Lee, *Technological Fair Use*, 83 S. CAL. L. REV. 797, 846 (2010) (advocating for an understanding of fair use that “afford[s] more leeway to developers at the creation and operation stages”); Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. REV. 1607, 1607–16 (2009) (finding the fair use doctrine to be crucial for allowing information dissemination); *see also* Jerome H. Reichman & Ruth L. Okediji, *When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale*, 96 MINN. L. REV. 1362, 1368–70 (2012) (arguing that extant copyright laws create gridlock blocking discovery tools such as data scraping).

have upheld challenges to scraping, especially under breach of contract claims pursuant to the websites' terms of use.⁷¹ Illinois' BIPA has also proven to be a promising ground for scraping suits, including in a pending case involving the facial recognition company Clearview AI following a report that revealed that Clearview sold a facial recognition tool trained on scraped data to law enforcement.⁷² As demonstrated by BIPA, new claims and legal theories continue to develop against the practice of unauthorized scraping. One such area that has so far been largely underexamined in the scraping context is trademark law.

II. POSSIBLE TRADEMARKS CLAIMS AGAINST SCRAPING

Trademark law has the mostly untapped possibility of defining the scope of permitted scraping practices. Trademark claims against web scrapers are far less common than the claims discussed above.⁷³ Yet, in certain contexts, trademark law could be a powerful deterrent to scraping and a promising legal claim. Where scraped information is combined with a trademark or leads to increased consumer confusion as to the relationship between parties, trademark infringement claims may arise.⁷⁴ The significant difference between trademark law and the claims usually brought against scrapers is that the latter focus on the illegality of scraping itself, whereas trademark law would circumscribe certain uses of the scraped information. Placing trademarked names or logos next to scraped data from those companies could raise potential trademark law concerns. This could significantly control certain uses of scraped information that depend on displaying different scraped companies' names for purposes such as online ordering and price comparisons. In particular, websites could pursue anti-scraping use claims for the inclusion of trademarks on others' websites under the federal Lanham Act alleging (1) trademark infringement, (2) false advertising, and (3) dilution.

71. Rubin & Ebel, *supra* note 66.

72. Thornley v. Clearview AI, Inc., 984 F.3d 1241 (7th Cir. 2021); Kashmir Hill, *The Secretive Company That Might End Privacy as We Know It*, N.Y. TIMES (Jan. 20, 2020), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html?utm_source=Memberful&utm_campaign=41977c2de4-daily_update_2020_01_21&utm_medium=email&utm_term=0_d4c7fece27-41977c2de4-111021813 [https://perma.cc/6EM3-Z898].

73. Liu, *supra* note 11, at 32 n.33; Riley, *supra* note 10, at 265 n.126; *see also infra* Part IV.

74. *See* O'Reilly, *supra* note 17, at 275.

A. Trademark Infringement

Section 32 of the Lanham Act, codified as § 1114 of Title 15 of the U.S. Code, provides the standard for infringement of a federally registered trademark.⁷⁵ It establishes that any person who, without the consent of the registrant, uses the registered mark in commerce “in connection with the sale, offering for sale, distribution, or advertising of any goods or services” in a manner that is likely “to cause confusion, or to cause mistake, or to deceive” is liable to the registrant.⁷⁶ To succeed on a Section 32 claim, the plaintiff must prove (1) that it owns the exclusive rights in the mark, (2) the defendant used the mark in commerce, (3) the defendant used the mark “in connection with the sale, offering for sale, distribution, or advertising” of goods or services, and (4) there is a likelihood of consumer confusion as to the source or sponsorship of the defendant’s goods or services.⁷⁷

The threshold requirement for a Section 32 claim is that the owner of the trademark has registered it with the United States Patent and Trademark Office (“USPTO”). The Lanham Act provides detailed instructions on the requirements for registration.⁷⁸ Once the trademark is registered with the USPTO, the owner can pursue actions under Section 32.

The inclusion of a trademark with associated scraped information on a website would likely qualify as use in commerce, the second requirement for trademark infringement. A use in commerce for the purposes of infringement has been defined as the use of a mark “on services when it is used or displayed in the sale or advertising of services and the services are rendered in commerce.”⁷⁹ In general, meeting this threshold requirement is not too demanding.⁸⁰ Using another’s mark on a website that sells goods

75. 15 U.S.C. § 1114.

76. *Id.* § 1114(1).

77. *Rosetta Stone Ltd. v. Google, Inc.*, 676 F.3d 144, 152, 157 (4th Cir. 2012) (citing 15 U.S.C. § 1114(1)(a)).

78. *See generally* 15 U.S.C. §§ 1051–72; *see also id.* §§ 1091–96 (laying out the provisions for the Supplemental Register).

79. *Rescuecom Corp. v. Google Inc.*, 562 F.3d 123, 128 (2d Cir. 2009) (quoting 15 U.S.C. § 1127). The exact definition of the term “use in commerce” was discussed at length by the Second Circuit in *Rescuecom*. The Second Circuit did not rule that the definition of “use in commerce” in Section 45 of the Lanham Act was not applicable to infringement, but it did state, in dicta, that the full definition in Section 45 was intended for defining “use in commerce” for the purposes of registration, not infringement. *Id.* at 133, 138. The court went on to suggest two possible interpretations of the Section 45 definition of “use in commerce” for infringement: (1) barring the definition completely or (2) simply disregard the “bona fide” language at the beginning of the definition. *Id.* at 140.

80. *See* BARTON BEEBE, *TRADEMARK LAW: AN OPEN-SOURCE CASEBOOK* 348 (7th ed. 2020) (in comparison to “use in commerce,” “what has proven to be a far more significant threshold

or services would almost certainly meet this threshold requirement of being a use in commerce.

A more difficult hurdle is proving that the defendant used the mark “in connection with the sale, offering for sale, distribution, or advertising” of goods or services, the so-called “commercial use” requirement.⁸¹ Unlike the mere use in commerce requirement, the commercial use requirement requires a specific nexus between the trademark and a commercial use.⁸² Disseminating an idea about an entity by using their mark for identification does not rise to a commercial use, even if the website is itself a commercial website.⁸³ But the use of a commercial or transactional component (i.e., sale or advertising) with the use of a trademark would qualify as a commercial use.⁸⁴ Particular uses of scraped data could fit this requirement. Displaying the wordmark or logo of a company next to its scraped data for the purposes of a sale or commercial service would presumably qualify as a commercial use. For example, a travel service including a trademarked hotel line as a booking option on its aggregated website would involve a use of the trademark in a commercial context: the sale of a hotel reservation. Even if the consumer did not purchase a reservation to that particular hotel, the hotel would still be offered on a commercial travel website, likely meeting the commercial requirement for trademark infringement.

The final requirement for a trademark claim, likelihood of confusion, is the heart of the infringement inquiry. Instead of the classic likelihood of confusion between two marks, scraping potentially triggers “source,”

requirement for liability in U.S. trademark law [is the use of the trademark in connection with the sale . . . of any goods or services], often called the ‘commercial use’ requirement.”).

81. *Id.* at 350 (quoting 15 U.S.C. § 1114(1)(a)); *see also* *Radiance Found., Inc. v. NAACP*, 786 F.3d 316, 322 (4th Cir. 2015) (“[A]n actionable trademark claim does not simply require that the alleged infringer used in commerce the mark that the trademark holder possesses. It also requires that the infringer’s use be ‘in connection with’ goods or services . . .”).

82. *BEEBE*, *supra* note 80, at 351–52.

83. *Radiance Found., Inc.*, 786 F.3d at 326 (“The provision of mere ‘information services’ without any commercial or transactional component is speech—nothing more.”). The rule is looser for domain names. There, not selling goods or services with the trademark is not enough to protect defendants; the defendant’s use interfering with consumers accessing the true trademarked product or service is enough to meet the commercial use requirement. *See* *PETA v. Doughney*, 263 F.3d 359, 365 (4th Cir. 2001) (“Doughney need not have actually sold or advertised goods or services on the *www.peta.org* website. Rather, Doughney need only have prevented users from obtaining or using PETA’s goods or services . . .”); *see also* *Radiance Found., Inc.*, 786 F.3d at 325–26 (confirming that the holding in *PETA* only applied to domain names).

84. *See* *Radiance Found., Inc.*, 786 F.3d at 326 (“The provision of mere ‘information services’ without any commercial or transactional component is speech—nothing more.”).

“sponsorship,” or “affiliation” confusion instead.⁸⁵ Confusion as to the source of the trademark or confusion as to sponsorship by or affiliation with a trademark owner through the use of its mark have been recognized as actionable consumer confusion.⁸⁶ Like all types of consumer confusion for trademark infringement, purported source, sponsorship, or affiliation confusion must meet the likelihood of confusion test. The likelihood of confusion inquiry is fact-driven, and involves the application of a multi-factor circuit court test for confusion as to the source, affiliation, or sponsorship of the goods or services.⁸⁷ For example, the Second Circuit’s *Polaroid* test looks at eight non-exclusive factors to determine if two marks are confusingly similar: (1) “the strength of [the] mark,” (2) “the degree of similarity between the two marks,” (3) “the proximity of the products,” (4) “the likelihood that the prior owner will bridge the gap,” (5) “actual confusion,” (6) “the defendant’s intent,” (7) “the quality of the defendant’s product,” and (8) “the sophistication of the buyers.”⁸⁸ Including another’s trademark on a product or service, under the right circumstances, can be considered to meet the likelihood of confusion inquiry as to affiliation or sponsorship. For example, in a case before the U.S. Court of Appeals for the Fifth Circuit, *Board of Supervisors for Louisiana State University Agricultural & Mechanical College v. Smack Apparel Co.*, an apparel company printed t-shirts with trademarks of major college athletics teams.⁸⁹ The Fifth Circuit determined that this led to

85. *Bd. of Supervisors for La. State Univ. Agric. & Mech. Coll. v. Smack Apparel Co.*, 550 F.3d 465, 478 (5th Cir. 2008) (quoting *Westchester Media v. PRL USA Holdings, Inc.*, 214 F.3d 658, 663 (5th Cir. 2000)) (“[T]he defendant’s use of the mark ‘creates a likelihood of confusion in the minds of potential customers as to the source, affiliation, or sponsorship’ of the product at issue.”).

86. *Int’l Info. Sys. Sec. Certification Consortium, Inc. v. Sec. Univ., LLC*, 823 F.3d 153, 161 (2d Cir. 2016) (quoting 4 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 23:76 (4th ed. 1997)) (“[T]he modern test of infringement is whether the defendant’s use [is] likely to cause confusion *not just as to source*, but also as to sponsorship, affiliation or connection.”); *Team Tires Plus, Ltd. v. Tires Plus, Inc.*, 394 F.3d 831, 835 (10th Cir. 2005) (“[T]he relevant confusion under trademark law is not limited to confusion of consumers as to the source of the goods, but also includes confusion as to sponsorship or affiliation, such as a consumer’s mistaken belief that a retailer is part of a larger franchising operation.”); *Nike, Inc. v. “Just Did It” Enters.*, 6 F.3d 1225, 1228–29 (7th Cir. 1993) (“[C]ustomer ‘confusion’ need not be restricted to a mistake regarding the source of the goods; the court should also consider whether the customer would believe that the trademark owner sponsored, endorsed or was otherwise affiliated with the product.”).

87. BEEBE, *supra* note 80, at 360, 372, 399; see *Smack Apparel*, 550 F.3d at 478; see also *Savannah Coll. of Art & Design, Inc. v. Sportswear, Inc.*, 872 F.3d 1256, 1264 (11th Cir. 2017) (noting that the court “will have to consider whether SCAD has demonstrated that Sportswear’s use of its word marks is likely to create consumer confusion as to origin, source, approval, affiliation, association, or sponsorship. . . . [A] court usually considers a number of factors in assessing whether an infringing use is likely to cause confusion.”).

88. *Polaroid Corp. v. Polarad Elecs. Corp.*, 287 F.2d 492, 495 (2d Cir. 1961).

89. *Smack Apparel*, 550 F.3d at 472–73.

potential confusion amongst consumers as to the legal relationship between the colleges and the apparel company.⁹⁰

Source confusion, although less likely, is still possible for scraped data connected with trademarks. “[S]crapers often provide links to third-party content on their own websites,”⁹¹ which would seem, at first glance, to clarify that the content is derived from the trademark owner’s website. However, linking to third-party content can still create a risk of causing confusion “as to the source of the marks being used,” resulting in trademark infringement.⁹² If the link is not prominent or not used, consumers may assume that the mark is being used by the scraper as its own mark. Even deep linking (linking to third-party content contained within a third-party website) can create consumer confusion if the user is not actually required to visit the third-party website.⁹³ However, for many uses of scraped data, such as food delivery, flight bookings, or hotel reservations, consumers understand that the scraper is a third-party intermediary rather than the owner of the trademark.⁹⁴

The more likely trademark infringement claim in response to a scraped trademark or scraped information with the source’s trademark is sponsorship or affiliation confusion. In *Smack Apparel*, the Fifth Circuit found that the inclusion of trademarks on products suggested some sort of connection between the trademark owner and the product.⁹⁵ This could potentially be expanded to include the use of trademarks on websites to suggest some sort of affiliation or connection between the website and the brand. Indeed, consumer confusion has been shown with delivery apps that list restaurants with which they do not actually have an affiliation.⁹⁶ For example, Michelin-starred San Francisco restaurant Kin Khao, despite not offering takeout—let alone delivery—received a call from an upset

90. *Id.* at 483–84.

91. Perry J. Viscounty, Jennifer L. Barry & James Field, *Spiders, Crawlers and Bots, Oh My: The Basics of Website Scraping*, INTELL. PROP. TODAY 32, 34 (Oct. 2012), <https://www.lw.com/thoughtLeadership/basics-of-web-scraping-IP> [<https://perma.cc/B2JU-XH5V>].

92. *Id.*

93. *Id.*

94. See Benjamin Edelman, *Mastering the Intermediaries*, HARV. BUS. REV. (June 2014), <https://hbr.org/2014/06/mastering-the-intermediaries> [<https://perma.cc/VDA9-RAHW>] (describing how consumers use aggregation platforms, suggesting that they know they are third-party intermediaries rather than the end-line retailers).

95. *Smack Apparel*, 550 F.3d at 483–84.

96. Caleb Pershan, *‘We Don’t Even Do Takeout’: Why, Then, Is This Restaurant on Seamless?*, EATER (Jan. 29, 2020, 1:59 PM), <https://www.eater.com/2020/1/29/21113416/grubhub-seamless-kin-khao-online-delivery-mistake-doordash> [<https://perma.cc/B2B3-JT29>].

customer asking about the status of their order.⁹⁷ Kin Khao was not alone; other restaurants were similarly contacted about orders placed through food delivery apps or even had food delivery couriers appear at their restaurants to pick up unauthorized orders.⁹⁸ The debacle led to a California law that expressly prohibits food delivery apps from listing restaurants without first obtaining permission from them.⁹⁹ Sponsorship and affiliation confusion is not without critics, as some scholars have criticized the seeming overuse of this type of consumer confusion.¹⁰⁰ However, suggestions for limiting sponsorship or affiliation confusion to “confusion that is actually relevant to purchasing decisions”¹⁰¹ would still appear to apply to potential scraping uses. As demonstrated by the Kin Khao example, posting trademarked names and logos next to scraped data has led to actual consumer confusion.

While Section 32 only applies to federally registered trademarks, Section 43(a)(1)(A), codified as § 1125(a) of Title 15 of the U.S. Code, protects all marks for goods or services, whether they are registered or not.¹⁰² Section 43(a)(1)(A) will protect unregistered trademarks as long as they could qualify for registration as marks under the Lanham Act.¹⁰³ To establish a claim under Section 43(a)(1)(A), the trademark holder must demonstrate that the defendant first uses (1) “in commerce any word, term, name, symbol, or device, or any combination thereof,” (2) “or any false designation of origin,” (3) “false or misleading description of fact, or” (4) “false or misleading representation of fact.”¹⁰⁴ The trademark holder must then show that such a use “is likely to cause confusion . . . as to the

97. *Id.*

98. *Id.*

99. CAL. BUS. & PROF. CODE § 22599 (West, Westlaw through Ch. 435 of 2021 Reg. Sess.); see also Jay Barmann, *Uproar Over Grubhub’s Kin Khao Debacle Leads to New Law Prohibiting ‘Non-Consensual’ Restaurant Listings*, SFIST (Jan. 4, 2021), <https://sfist.com/2021/01/04/uproar-over-doordash-and-sf-ghost-kitchens-leads-to-new-law-prohibiting-non-consensual-restaurant-listings> [https://perma.cc/AEL3-PV8J]. New Hampshire enacted a similar law, to take effect at the end of 2023, and another similar bill is currently pending in New York. H.B. 593-FN, 2021 Reg. Sess. (N.H. 2021); S. A784, 2021–2022 Gen. Assemb. (N.Y. 2021).

100. See, e.g., Mark A. Lemley & Mark McKenna, *Irrelevant Confusion*, 62 STAN. L. REV. 413, 417–22 (2010) (criticizing several court decisions that found likelihood of sponsorship or affiliation confusion where it was highly unlikely, such as *Anheuser-Busch, Inc. v. Balducci Publications and Mutual of Omaha Ins. Co. v. Novak*).

101. *Id.* at 416. But see Matthew B. Kugler, *The Materiality of Sponsorship Confusion*, 50 U.C. DAVIS L. REV. 1911, 1921–27 (2017) (arguing that the actual impact of such a “materiality” requirement would not be as broad as expected).

102. 15 U.S.C. § 1125.

103. *Two Pesos, Inc. v. Taco Cabana, Inc.*, 505 U.S. 763, 768 (1992).

104. 15 U.S.C. § 1125(a)(1).

affiliation” of the two entities.¹⁰⁵ These factors are practically identical to those for a Section 32 analysis for trademark infringement. Indeed, courts often apply them interchangeably.¹⁰⁶ Therefore, the analysis for scraping for trademark infringement under Section 43(a) is effectively the same as for Section 32 trademark infringement analyzed above.¹⁰⁷ Claims for uses of scraped data that infringe trademarks could thus be litigated by both owners of federally registered trademarks as well as non-registered but registerable trademarks under this same framework.

B. False Advertising

In addition to Section 43(a) including a claim for infringement of an unregistered trademark, Section 43(a)(1)(B) also includes a false advertising claim.¹⁰⁸ A claim for false advertising requires the plaintiff to establish:

- (1) A false or misleading statement of fact about a product;
- (2) Such statement either deceived, or had the capacity to deceive a substantial segment of potential consumers;
- (3) The deception is material, in that it is likely to influence the consumer’s purchasing decision;
- (4) The product is in interstate commerce; and
- (5) The plaintiff has been or is likely to be injured as a result of the statement at issue.¹⁰⁹

The threshold question for a false advertising claim is whether the false advertising is a literally false statement or a statement that is misleading, the answer to which dictates what must be proven in regard to the other elements of the test. A literally false statement is one that is incorrect on its face.¹¹⁰ In many circuits, if a false statement is proven,

105. *Id.* at § 1125(a)(1)(A).

106. BEEBE, *supra* note 80, at 337.

107. *See supra* Part II(A).

108. 15 U.S.C. § 1125(a)(1). False advertising claims can be broader than just those under Section 43(a)(1)(B) of the Lanham Act and can include claims under the FTC Act and state and common law actions. *Id.* §§ 41–58.

109. *Pizza Hut, Inc. v. Papa John’s Int’l, Inc.*, 227 F.3d 489, 495 (5th Cir. 2000).

110. *Schering-Plough Healthcare Prods., Inc. v. Schwarz Pharma, Inc.*, 586 F.3d 500, 513 (7th Cir. 2009) (“A ‘literal’ falsehood is bald-faced, egregious, undeniable, over the top.”); *Time Warner*

deception is presumed,¹¹¹ as are materiality and harm to the plaintiff in many cases.¹¹² By comparison, a misleading statement requires more to show that it was false advertising. The plaintiff must present evidence that consumers were actually deceived by the statement, both to prove deception and materiality.¹¹³ For demonstrating injury, the plaintiff must show that a “statistically significant part of the commercial audience holds the false belief” that was communicated.¹¹⁴

Common uses of scraped data could give rise to both literally false and misleading statements. A false statement could be possible where a website incorrectly states information about the entity it scraped from. For example, in the food delivery service examples above,¹¹⁵ food delivery apps included incorrect menu items for restaurants and incorrectly stated that they offered delivery.¹¹⁶ Published scraped data is directly taken from the other entity’s website, but there is still a duty to maintain the information, such as menu and delivery options and hours. Given that statements are directly taken from websites in the context of scraping, however, misleading statements may be a more likely assertion. While the plaintiff would need to show evidence of consumer perception, this could likely be demonstrated in at least some cases. Misleading statements about

Cable, Inc. v. DIRECTV, Inc., 497 F.3d 144, 158 (2d Cir. 2007) (quoting Novartis Consumer Health, Inc. v. Johnson & Johnson-Merck Consumer Pharm. Co., 290 F.3d 578, 587 (3d Cir. 2002)) (“[O]nly an *unambiguous* message can be literally false.”).

111. S.C. Johnson & Son, Inc. v. Clorox Co., 241 F.3d 232, 238 (2d Cir. 2001) (quoting Castrol, Inc. v. Quaker State Corp., 977 F.2d 57, 62 (2d Cir. 1992) (stating that under Section 43(a)(1)(B), courts “will presume irreparable harm where plaintiff demonstrates a likelihood of success in showing literally false defendant’s comparative advertisement which mentions plaintiff’s product by name”).

112. BEEBE, *supra* note 80, at 705–06.

113. *DIRECTV*, 497 F.3d at 158 (citing *Am. Home Prods. Corp. v. Johnson & Johnson*, 577 F.2d 160, 166 (2d Cir. 1978)); *Pizza Hut*, 227 F.3d at 497 (quoting *Am. Council of Certified Podiatric Physicians and Surgeons v. Am. Bd. Podiatric Surgery, Inc.*, 185 F.3d 606, 616 (6th Cir. 1999)) (With respect to materiality, “if the statements at issue are either ambiguous or true but misleading, the plaintiff must present evidence of actual deception. . . . [P]roof of actual deception requires proof that ‘consumers were actually deceived by the defendant’s ambiguous or true-but-misleading statements.’”); *see also Pizza Hut*, 227 F.3d at 502 (Statements are material where “they had a tendency to influence the purchasing decisions of, the consumers to which they were directed.”). While the Fifth Circuit determined that the court should assume materiality where there is a literally false statement, *Pizza Hut*, 227 F.3d at 497 (“With respect to materiality . . . the court will assume that the statements actually misled consumers.”), the Second Circuit stated that materiality must be separately proven for both literally false and misleading statements, *DIRECTV*, 497 F.3d at 153 n.3.

114. *Johnson & Johnson * Merck Consumer Pharms. Co. v. Smithkline Beecham Corp.*, 960 F.2d 294, 298 (2d Cir. 1992).

115. Pershan, *supra* note 96.

116. *Id.*; *see also* Edward Ongweso Jr., *Food Delivery Apps Are a Confusing Mess*, VICE (Jan. 28, 2020, 11:17 AM), <https://www.vice.com/en/article/wxejmq/food-delivery-apps-are-a-confusing-mess> [<https://perma.cc/V8Z2-J2T6>] (describing examples of the incorrect information in food delivery apps).

the availability of certain dishes or delivery offerings could influence consumers' decisions on where to order dinner in the future.¹¹⁷ Such uses of scraped data with companies' names or logos deceiving consumers would seem to fall within the very heart of false advertising.¹¹⁸

C. Dilution

Owners of famous marks can also file suit for dilution of their mark, either by blurring or by tarnishment, under Section 43(c).¹¹⁹ A mark is famous "if it is widely recognized by the general consuming public of the United States as a designation of source of the goods or services of the mark's owner."¹²⁰ In determining whether a mark is famous, courts holistically consider all relevant factors, including (1) "[t]he duration, extent, and geographic reach of advertising and publicity of the mark, whether advertised or publicized by the owner or third parties"; (2) "[t]he amount, volume, and geographic extent of sales of goods or services offered under the mark"; (3) "[t]he extent of actual recognition of the mark"; and (4) "[w]hether the mark [is] registered"¹²¹ Examples of famous marks include Coca-Cola,¹²² Nike's "Just Do It,"¹²³ and Google.¹²⁴

The owner of a famous mark can receive an injunction against the user of their mark if that entity "at any time after the owner's mark has become famous, commences use of a mark or trade name in commerce that is likely to cause dilution by blurring or dilution by tarnishment of the famous mark, regardless of the presence or absence of actual or likely confusion, of competition, or of actual economic injury."¹²⁵ This requires the plaintiff to prove:

117. Ed Currington, *Why 3rd-Party Delivery Platforms Are Problematic*, DIGIT. REST. (Apr. 19, 2019), <https://thedigitalrestaurant.com/food-delivery-service-apps-problem-for-restaurants/> [<https://perma.cc/XY3F-HP5S>] (noting that when any delivery problems occur, the restaurant, not the delivery service, is blamed).

118. *See, e.g.,* Schering-Plough Healthcare Prods., Inc. v. Schwarz Pharma, Inc., 586 F.3d 500, 512 (7th Cir. 2009) ("The purpose of the false-advertising provisions of the Lanham Act is to protect sellers from having their customers lured away from them by deceptive ads (or labels, or other promotional materials).").

119. 15 U.S.C. § 1125(c).

120. *Id.* § 1125(c)(2)(A).

121. *Id.*; *Parts.com, LLC v. Yahoo! Inc.*, 996 F. Supp. 2d 933, 940 (S.D. Cal. 2013).

122. *Coca-Cola Co. v. Purdy*, 382 F.3d 774, 780 (8th Cir. 2004).

123. *Nike, Inc. v. Caldwell*, No. 91240394, 2020 WL 3027610, at *3–4 (T.T.A.B. Apr. 28, 2020).

124. *Google Inc. v. ACPS Glob., Inc.*, No. 91198926, 2012 WL 2364349, at *5 (T.T.A.B. June 11, 2012).

125. 15 U.S.C. § 1125(c)(1).

(1) that the plaintiff owns a famous mark that is distinctive; (2) that the defendant has commenced using a mark in commerce that allegedly is diluting the famous mark; (3) that a similarity between the defendant's mark and the famous mark gives rise to an association between the marks; and (4) that the association is likely to impair the distinctiveness of the famous mark or likely to harm the reputation of the famous mark.¹²⁶

Impairment of the mark can take place either through blurring or tarnishment. Dilution by blurring is "association arising from the similarity between a mark or trade name and a famous mark that impairs the distinctiveness of the famous mark."¹²⁷ Dilution by tarnishment, on the other hand, is "association arising from the similarity between a mark or trade name and a famous mark that harms the reputation of the famous mark."¹²⁸

Uses of scraped data with trademarks could lead to dilution by tarnishment. Including scraped information on a website does not suggest a diminishing of the distinctiveness of the mark, as it continues to use it as a mark that signals the trademark owner, making dilution by blurring inappropriate. Dilution by tarnishment, on the other hand, occurs when a trademark is linked to inferior products or services or is portrayed in an unfavorable light.¹²⁹ For example, the Second Circuit deemed the use of the term "New York \$lot Exchange" for a slot machine to tarnish the New York Stock Exchange's trademark.¹³⁰ Dilution by tarnishment would seem to require that the trademark is used for source identification, or, in other words, as a mark.¹³¹ In the above example, the name of the slot machine was used to identify it. But a mere critical statement, such as labeling a company as "evil," would not qualify as tarnishment because

126. *Louis Vuitton Malletier S.A. v. Haute Diggity Dog, LLC*, 507 F.3d 252, 264–65 (4th Cir. 2007); *see also* *Parts.com, LLC v. Yahoo! Inc.*, 996 F. Supp. 2d 933, 940 (S.D. Cal. 2013) (using a five element test that is effectively the same, but splits fame and distinctiveness into two separate elements).

127. 15 U.S.C. § 1125(c)(2)(B).

128. *Id.* § 1125(c)(2)(C).

129. *Deere & Co. v. MTD Prods., Inc.*, 41 F.3d 39, 43 (2d Cir. 1994) ("‘Tarnishment’ generally arises when the plaintiff’s trademark is linked to products of shoddy quality, or is portrayed in an unwholesome or unsavory context likely to evoke unflattering thoughts about the owner’s product.”).

130. *N.Y. Stock Exch., Inc. v. N.Y., N.Y. Hotel, LLC*, 293 F.3d 550, 558 (2d Cir. 2002).

131. Barton Beebe, *A Defense of the New Federal Trademark Antidilution Law*, 16 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1143, 1172 (2006) ("The plaintiff must show, first, that the defendant is tarnishing the plaintiff’s mark by means of something that consumers perceive as a designation of source of the defendant’s goods"); Stacey L. Dogan & Mark A. Lemley, *The Trademark Use Requirement in Dilution Cases*, 24 SANTA CLARA HIGH TECH. L.J. 541, 545, 557 (2008) ("The TDRA clearly includes a trademark use requirement [for dilution].").

that statement is not being used as a trademark itself.¹³² In comparison, using scraped data and a trademark together serve as a mark, identifying the information as connected to the trademark owner. Like with false advertising claims, inaccurate or misleading statements could harm the reputation of the trademark owner. It could also suggest an association between the trademark owner and the scraper,¹³³ which could harm the trademark owner if consumers would think less of the trademark owner due to the association.

III. THE VIABILITY OF NOMINATIVE FAIR USE

The most applicable defense for improper trademark use with scraped data is nominative fair use. The Lanham Act provides for a number of defenses to trademark infringement.¹³⁴ But scrapers will often try to defend against claims of trademark infringement by arguing that their use of the trademark qualifies as nominative fair use.¹³⁵ Nominative fair use provides that a trademark can be used in a descriptive way to identify the trademark owner's goods or services.¹³⁶ An entity engaging in nominative fair use would use the trademark in a competitive or comparative manner that does not falsely imply sponsorship or endorsement by the trademark holder.¹³⁷

At present, there is a circuit split on the test used for nominative fair use under the Lanham Act between the Ninth and Third Circuits, as well as separate interpretations by the Second and Fifth Circuits. The Ninth Circuit established its nominative fair use test in *New Kids on the Block v. News America Publishing, Inc.*¹³⁸ In *New Kids on the Block*, the plaintiff boy band brought Lanham Act claims against newspapers who had created 900-numbers that allowed fans to participate in telephonic polls about the

132. Beebe, *supra* note 131, at 1172.

133. O'Reilly, *supra* note 17, at 277 (“[A]ggregators’ sites present information from various sites and may also include other corporations’ logos, which might suggest a relationship between the two entities where none exists.”).

134. 15 U.S.C. § 1115(b) (enumerating eight defenses to Section 32 infringement, many of which also apply to Section 43 claims). Nominative fair use is only explicitly referred to in the Lanham Act as a defense to dilution, 15 U.S.C. § 1125(c)(3)(A), but has been interpreted to be a defense to trademark infringement, *New Kids on the Block v. News Am. Publ’g, Inc.*, 971 F.2d 302, 308–09 (9th Cir. 1992).

135. Viscounty et al., *supra* note 91, at 34.

136. See *New Kids on the Block*, 971 F.2d at 308.

137. See *id.*

138. 971 F.2d 302.

band and its members.¹³⁹ The Ninth Circuit recognized that a trademark is a “limited property right,” and that there is a valid trademark defense “where the mark is used only ‘to describe the goods or services of [a] party, or their geographic origin.’”¹⁴⁰ The classic trademark fair use case is where a defendant uses the plaintiff’s trademark to refer to the defendant’s goods or services.¹⁴¹ A nominative fair use instead deliberately uses the trademark to refer to the plaintiff.¹⁴² The Ninth Circuit held that a defendant is entitled to a nominative fair use defense where (1) “the product or service in question must be one not readily identifiable without use of the trademark;” (2) “only so much of the mark or marks may be used as is reasonably necessary to identify the product or service;” and (3) “the user must do nothing that would, in conjunction with the mark, suggest sponsorship or endorsement by the trademark holder.”¹⁴³ In *New Kids on the Block*, the newspapers met this test, as there was no substitute term for New Kids on the Block, the band was only referenced “to the extent necessary to identify them as the subject of the polls,” and nothing in the newspapers’ use suggested a sponsorship or endorsement by New Kids on the Block.¹⁴⁴

Thirteen years later, the U.S. Court of Appeals for the Third Circuit established its own, different test for nominative fair use in *Century 21 Real Estate Corp. v. Lendingtree, Inc.*¹⁴⁵ In that case, Lendingtree used several of Century 21’s trademarks on its website, including statements that Lendingtree’s service allowed users to access Century 21 brokers and that Lendingtree was “affiliated with” Century 21 brokers.¹⁴⁶ The Third Circuit acknowledged that a separate test was needed for nominative fair use cases, but disagreed with the Ninth Circuit’s test removing consideration of likelihood of confusion and placing the burden entirely on the defendant.¹⁴⁷ The Third Circuit instead established a two-prong test, where (1) the plaintiff must prove likelihood of confusion under the

139. *Id.* at 304.

140. *Id.* at 306 (quoting 15 U.S.C. § 1115(b)(4)).

141. *Id.* at 308; *see also* 15 U.S.C. § 1115(b)(4).

142. *New Kids on the Block*, 971 F.2d at 308.

143. *Id.*

144. *Id.* at 308–09.

145. 425 F.3d 211 (3d Cir. 2005).

146. *Id.* at 215.

147. *Id.* at 220–21 (“[W]e do not accept the legal basis or advisability of supplanting the likelihood of confusion test entirely. . . . [T]he approach of the Court of Appeals for the Ninth Circuit would relieve the plaintiff of the burden of proving the key element of a trademark infringement case—likelihood of confusion—as a precondition to a defendant’s even having to assert and demonstrate its entitlement to a nominative fair use defense.”).

relevant factors from the traditional likelihood of confusion test and (2) the defendant must then show that its use is fair, using a three-part test largely derived from that of the Ninth Circuit.¹⁴⁸ The three-part fairness test in the Third Circuit is:

(1) that the use of plaintiff's mark is necessary to describe both the plaintiff's product or service and the defendant's product or service; (2) that the defendant uses only so much of the plaintiff's mark as is necessary to describe plaintiff's product; and (3) that the defendant's conduct or language reflect the true and accurate relationship between plaintiff and defendant's products or services.¹⁴⁹

Since *Lendingtree*, the Fifth and Second Circuits have devised their own interpretations of the nominative fair use test. In *International Information Systems Security Certification Consortium, Inc., v. Security University, LLC*, the Second Circuit rejected having nominative fair use be an affirmative defense, as the Third Circuit did.¹⁵⁰ It instead held that "in nominative use cases, district courts are to consider the Ninth Circuit and Third Circuit's nominative fair use factors, in addition to the *Polaroid* factors."¹⁵¹ The Fifth Circuit has stated that the nominative fair use inquiry should be done in conjunction with the likelihood of confusion inquiry,¹⁵² suggesting a greater emphasis on likelihood of confusion than in *New Kids on the Block*.

These four tests set up a dichotomy between the Second, Fifth, and Ninth Circuits and the Third Circuit. The former hold that if a use creates a likelihood of confusion, it cannot be nominative fair use.¹⁵³ The Third Circuit holds that nominative fair use is an affirmative defense to infringement (and a likelihood of confusion).¹⁵⁴ However, all of the tests effectively incorporate a direct consideration of likelihood of confusion as to affiliation or sponsorship by including the third element of the Ninth Circuit's test, prohibiting a suggestion of "sponsorship or endorsement"

148. *Id.* at 222, 225–26.

149. *Id.* at 222.

150. 823 F.3d 153, 167 (2d Cir. 2016).

151. *Id.* at 168.

152. *Bd. of Supervisors for La. State Univ. Agric. & Mech. Coll. v. Smack Apparel Co.*, 550 F.3d 465, 489 (5th Cir. 2008). The *Smack Apparel* court heavily cites to the earlier Fifth Circuit decision in *Pebble Beach Co. v. Tour 18 I Ltd.*, but that decision relied primarily on the Ninth Circuit test from *New Kids on the Block*. *Pebble Beach Co. v. Tour 18 I Ltd.*, 155 F.3d 526, 545–46 (5th Cir. 1998).

153. *Int'l Info. Sys. Sec. Certification Consortium, Inc. v. Sec. Univ., LLC*, 823 F.3d 153, 168 (2d Cir. 2016); *Smack Apparel*, 550 F.3d at 489; *New Kids on the Block v. News Am. Publ'g, Inc.*, 971 F.2d 302, 308 (9th Cir. 1992).

154. *Century 21 Real Est. Corp. v. Lendingtree, Inc.*, 425 F.3d 211, 222, 225–26 (3d Cir. 2005).

between the products or services, or, in the language of the Third Circuit, “reflect[ing] the true and accurate relationship between plaintiff and defendant’s products or services.”¹⁵⁵

It is this sponsorship or affiliation language that could be the greatest hurdle for a nominative fair use defense to overcome in the context of scraping. Uses of the trademarks alongside scraped data are often for the purpose of describing or referring to the trademark owner, which would seem to fit the goal of nominative fair use quite well. But, as explained in Part II, posting scraped data alongside a trademark could impermissibly suggest an affiliation or sponsorship by the trademark owner.¹⁵⁶ If this were the case, the use would not qualify as nominative fair use under the third prong of the Ninth Circuit test, as well as elements under the Second and Fifth Circuit tests. The Third Circuit’s requirement that a nominative fair use “reflect the true and accurate relationship between plaintiff and defendant’s products or services”¹⁵⁷ would also seemingly be violated by an improper suggestion of affiliation or sponsorship.

On the other hand, a plausible nominative fair use argument could delay a ruling on a Lanham Act case. Nominative fair use arguments are typically not ruled on at the motion to dismiss stage because they inherently entail a detailed factual inquiry.¹⁵⁸ This would require the plaintiff to go through (often expensive) discovery, possibly deterring potential claimants.

IV. ANTI-SCRAPING TRADEMARKS CLAIMS THUS FAR

Despite the robust legal issues around scraping and trademarks as described in Parts II and III, jurisprudence on these issues has been severely lacking. As of January 2021, there have only been seven judicial opinions addressing scraping and trademark usage in any detail.¹⁵⁹

155. *Id.* at 222; *New Kids on the Block*, 971 F.2d at 308.

156. *See supra* Part II(A).

157. *Century 21*, 425 F.3d at 222.

158. *Int’l Council of Shopping Ctrs., Inc. v. Info Quarter, LLC*, 17-CV-5526, 2018 U.S. Dist. LEXIS 152974, at *15 (S.D.N.Y. Sept. 7, 2018) (citing various cases in the U.S. District Court for the Southern District of New York refusing to address nominative fair use arguments at the motion to dismiss stage).

159. *Register.com, Inc. v. Verio, Inc.*, 356 F.3d 393 (2d Cir. 2004); *Sw. Airlines Co. v. Roundpipe, LLC*, 375 F. Supp. 3d 687 (N.D. Tex. 2019); *Alan Ross Mach. Corp. v. Machinio Corp.*, No. 17-cv-3569, 2018 U.S. Dist. LEXIS 113012 (N.D. Ill. July 9, 2018); *DHI Grp., Inc. v. Kent*, No. H-16-1670, 2017 WL 8794877 (S.D. Tex. Apr. 21, 2017); *Craigslist Inc. v. 3Taps Inc.*, 942 F. Supp. 2d 962 (N.D. Cal. 2013); *Citizens Info. Assocs., LLC v. JustMugshots.com*, No. 1-12-CV-573, 2013 U.S. Dist. LEXIS 194547 (W.D. Tex. Feb. 26, 2013); *Cvent, Inc. v. Eventbrite, Inc.*, 739 F. Supp. 2d

Reviewing these cases displays the legal landscape so far and elucidates the areas that will need greater clarity in the future.

The first case to address scraped data and trademarks was the early scraping case of *Register.com v. Verio, Inc.* in 2004.¹⁶⁰ Verio was a data aggregator that scraped information from Internet registries to compile a database of website registration information.¹⁶¹ Like in *Ticketmaster*,¹⁶² the use of bots for scraping was not copyright infringement because the scraped information was public information not owned by anyone.¹⁶³ However, unlike in *Ticketmaster*, Verio used the scraped information to call and email the domain name registrants and offer them Verio's services.¹⁶⁴ In these phone conversations and messages, Verio explicitly referred to the responders' registration with Register.com and provided misleading information about their "recently registered domain name," which the court found could have led—and, at least in a few situations, did lead—recipients to believe that Verio was affiliated with Register.com.¹⁶⁵ The district court issued an injunction based on these actions violating the Lanham Act, and the Second Circuit upheld the injunction, finding that Verio's actions could be misleading as to the relationship between Register.com and Verio.¹⁶⁶

It would be six years before a court addressed the interplay between scraping and trademarks again. In *Cvent, Inc. v. Eventbrite, Inc.*, the plaintiff alleged that events ticketing website Eventbrite scraped aggregated event information off of its website and copied the information onto its own website.¹⁶⁷ Cvent filed a complaint alleging, *inter alia*, the rare trademark infringement claim of reverse passing off.¹⁶⁸ The court found that Cvent's trademarks claim was not that Eventbrite had "passed off its *ideas* as its own, but rather that Eventbrite ha[d] re-branded and re-packaged its *product* . . . and sold it as its own," distinguishing it from the

927 (E.D. Va. 2010). An eighth case, *Facebook, Inc. v. Power Ventures, Inc.*, also contained a trademark infringement claim, but the parties stipulated to dismiss the trademark infringement claim prior to the summary judgment stage, so it was not discussed in the decision. No. 08-CV-05780, 2017 WL 3394754, at *2 (N.D. Cal. Aug. 8, 2017).

160. 356 F.3d 393 (2d Cir. 2004).

161. *Id.* at 396–97.

162. *Ticketmaster Corp. v. Tickets.com, Inc.*, No. CV997654, 2003 WL 21406289, at *4 (C.D. Cal. Mar. 7, 2003).

163. *Verio*, 356 F.3d at 418 ("WHOIS information cannot be copyrighted").

164. *Id.* at 396.

165. *Id.* at 405.

166. *Id.* at 405–06.

167. 739 F. Supp. 2d 927, 930 (E.D. Va. 2010).

168. *Id.*

facts in the Supreme Court decision in *Dastar Corp. v. Twentieth Century Fox Film Corp.*,¹⁶⁹ which had limited the breadth of reverse passing off claims to prevent perpetual copyrights through trademark enforcement actions.¹⁷⁰ Given this finding, the *Cvent* court refused to dismiss *Cvent*'s Lanham Act claim.¹⁷¹ This decision was not without controversy, however, as noted intellectual property law scholars, such as Rebecca Tushnet, remarked that the Eastern District of Virginia misinterpreted the division between copyright and trademarks that was established in *Dastar*, which should have blocked the Lanham Act claim in this case as the "product" was either uncopyrightable facts or only copyrightable expression in how the facts were compiled on *Cvent*'s site, but not Eventbrite's.¹⁷²

The next trademark case also involved a reverse passing-off claim. In *Citizens Information Associates, LLC v. JustMugshots.com*, defendant JustMugshots.com allegedly scraped over fourteen million records from website BustedMugshots.com to create its own website with police photographs and related arrest records.¹⁷³ The plaintiff alleged that JustMugshots.com was engaging in reverse passing-off, violating the Lanham Act.¹⁷⁴ However, unlike in *Verio* and *Cvent*, the data JustMugshots.com was scraping was not only publicly available, but was public domain mugshots and arrest-record information.¹⁷⁵ Like in *Cvent*, the case turned on the court's application of *Dastar*, but this time, the court found that protecting public domain information via trademarks was barred under *Dastar* and dismissed the claim.¹⁷⁶

Later that same year, the United States District Court for the Northern District of California issued its opinion in *Craigslist Inc. v. 3Taps Inc.*¹⁷⁷ In *Craigslist*, defendant 3Taps aggregated and republished content scraped from Craigslist's platform, effectively replicating the entire Craigslist

169. *Id.* at 936.

170. *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23, 37 (2003) (holding that allowing claims where a defendant merely used a plaintiff's creation without attribution "would be akin to finding that § 43(a) created a species of perpetual patent and copyright, which Congress may not do.").

171. *Cvent*, 739 F. Supp. 2d at 936.

172. Rebecca Tushnet, *Website Scraping Claim Survives Dastar Defense*, 43(B)LOG (Sept. 18, 2010, 8:26 PM), <https://tushnet.blogspot.com/2010/09/website-scraping-claim-survives-dastar.html> [<https://perma.cc/9TUDU-VXRB>].

173. No. 1-12-CV-573, 2013 U.S. Dist. LEXIS 194547, at *3-4 (W.D. Tex. Feb. 26, 2013).

174. *Id.* at *6.

175. *Id.* at *7.

176. *Id.*

177. 942 F. Supp. 2d 962 (N.D. Cal. 2013).

website under 3Taps' name.¹⁷⁸ Defendants Padmapper and Lovely, meanwhile, provided online real estate listings, including Craigslist listings they had received from 3Taps.¹⁷⁹ Craigslist filed a complaint alleging, *inter alia*, trademark infringement, false designation of origin, and dilution of a famous mark.¹⁸⁰ Unlike the reverse passing off claims in *Cvent* and *Citizens Information*, Craigslist alleged that 3Taps had suggested that Craigslist was the source of or had endorsed 3Taps' site.¹⁸¹ The Northern District of California held that these regular "passing off" claims were not prohibited by *Dastar*.¹⁸² The court went on to find that Craigslist had plausibly alleged its trademarks claims, citing examples such as 3Taps saying "craigslist data, better than craigslist!" and Padmapper stating that it was "Bringing Craigslist Back" to its website.¹⁸³ The court therefore refused to dismiss Craigslist's trademark and dilution claims.¹⁸⁴

Next, in *DHI Group, Inc. v. Kent*, the plaintiffs used a scraping program to bypass defendant Oilpro's security measures and download data from its website in order to populate oil professionals' resume information on their own platform.¹⁸⁵ Oilpro alleged that the plaintiffs had infringed its trademark by including it on their website along with the scraped data.¹⁸⁶ As Oilpro's trademark was unregistered, the court analyzed the claim under Section 43(a), and not under Section 32, which only applies to registered trademarks.¹⁸⁷ When the plaintiffs republished Oilpro's data on its own website, it also republished the Oilpro trademark.¹⁸⁸ The court found that the allegation that the trademark was distinctive and the image of the Oilpro mark on plaintiffs' publication was sufficient to survive a motion to dismiss.¹⁸⁹ The plaintiffs raised a fair use defense, but the court declined to address the fair use defense before the summary judgment stage.¹⁹⁰

178. *Id.* at 966.

179. *Id.* at 966–67.

180. *Id.* at 967.

181. *Id.* at 978.

182. *Id.* at 979–80.

183. *Id.* at 980.

184. *Id.*

185. No. H-16-1670, 2017 WL 8794877, at *2 (S.D. Tex. Apr. 21, 2017).

186. *Id.* at *9.

187. *Id.* at *9–10.

188. *Id.* at *10.

189. *Id.* at *11.

190. *Id.*

The following year, the plaintiff in *Alan Ross Machine Corp. v. Machinio Corp.* alleged that Machinio scraped its website for sales listings on machinery and duplicated those listings on its own website.¹⁹¹ The plaintiff alleged trademark infringement through false endorsement.¹⁹² But the court dismissed the trademark infringement claim because the plaintiff had not used “Alan Ross Machinery” as a trademark to distinctively identify the machinery it sold or its machinery listings.¹⁹³ The court further held that even if “Alan Ross Machinery” were a valid trademark, there was no evidence of likelihood of consumer confusion because the name does not even appear on Machinio’s website.¹⁹⁴ A “click to contact seller” button or link to Alan Ross’s website next to each listing “does not imply that Alan Ross has endorsed Machinio or has any specific affiliation with Machinio to support a false endorsement claim.”¹⁹⁵ The court then went further, stating that “[e]ven if Machinio had listed Alan Ross’s name on its website, that does not suggest a consumer would be misled into believing that Alan Ross had endorsed Machinio’s website.”¹⁹⁶

The seventh and final case involves Southwest Airlines. Southwest Airlines pursued anti-scraping litigation alleging trademark violations as early as 2004,¹⁹⁷ but the Lanham Act claims only made it to discussion in a court decision in 2019. In *Southwest Airlines Co. v. Roundpipe, LLC*, Southwest alleged that Roundpipe had scraped flight cost information from its website and reproduced it, along with the Southwest name and logos, on its own website.¹⁹⁸ The court held that Southwest had stated plausible claims for, *inter alia*, unfair competition, trademark infringement, and dilution under the Lanham Act because it showed valid trademark registrations and how the unauthorized use of the Southwest mark on Roundpipe’s website tarnished the Southwest trademarks and likely caused consumer confusion.¹⁹⁹

These seven decisions provide important baseline principles for which uses of scraped data are permissible and which are prohibited. The Lanham Act claims were only dismissed in two out of seven cases.²⁰⁰ The

191. No. 17-cv-3569, 2018 U.S. Dist. LEXIS 113012, at *2–3 (N.D. Ill. July 9, 2018).

192. *Id.* at *5.

193. *Id.* at *6, *9.

194. *Id.* at *7.

195. *Id.* at *8.

196. *Id.*

197. *Sw. Airlines Co. v. Farechase, Inc.*, 318 F. Supp. 2d 435, 437 (N.D. Tex. 2004).

198. 375 F. Supp. 3d 687, 691–92 (N.D. Tex. 2019).

199. *Id.* at 706.

200. *Alan Ross*, 2018 U.S. Dist. LEXIS 113012, at *6–9; *Citizens Info. Assocs., LLC v.*

Supreme Court opinion in *Dastar* served as an early roadblock to Lanham Act claims for scraping, leading to dismissal in one case and a controversial retention in another,²⁰¹ but the *Craigslist* court provided a clear path forward by differentiating regular infringement and dilution claims from reverse passing off, which was largely blocked by *Dastar*.²⁰² Trademark infringement claims under Section 32 or Section 43(a) were preserved in five cases,²⁰³ and dilution claims were raised and preserved in two cases.²⁰⁴ A nominative fair use defense was only raised in one case, but the court did not decide on it, preserving the defense for the summary judgment stage.²⁰⁵ False advertising claims were not raised in any of the seven cases. The *Alan Ross* court was the most protective of scraping uses, arguing that even if the wordmark were on the website with the scraped information, that does not suggest that there would be consumer confusion as to endorsement.²⁰⁶ The courts in *Verio*, *Cvent*, *Craigslist*, *DHI Group*, and *Southwest Airlines*, on the other hand, found that trademark infringement claims, particularly as to confusion over sponsorship or affiliation, as well as dilution claims, could proceed.²⁰⁷ Overall, existing precedent demonstrates that trademark claims can be successful in the scraping context, although the case law is still scant for trademark infringement and dilution and non-existent for false advertising.

V. FUTURE SCRAPING LITIGATION

While the seven cases in Part IV provide some guidance, the legal possibilities and restrictions around trademark usage and scraping are still largely unsettled due to the dearth of precedent. However, since 2020, six potentially significant cases have been filed in courts around the United

JustMugshots.com, No. 1-12-CV-573, 2013 U.S. Dist. LEXIS 194547, at *7 (W.D. Tex. Feb. 26, 2013).

201. *Cvent, Inc. v. Eventbrite, Inc.*, 739 F. Supp. 2d 927, 936 (E.D. Va. 2010) (retained, to the extent Cvent's Lanham Act claim was an alternative to its copyright claim); *Citizens Info.*, 2013 U.S. Dist. LEXIS 194547, at *7 (dismissed).

202. *Craigslist Inc. v. 3Taps Inc.*, 942 F. Supp. 2d 962, 979–80 (N.D. Cal. 2013).

203. *Register.com, Inc. v. Verio, Inc.*, 356 F.3d 393, 405–06 (2d Cir. 2004); *Cvent*, 739 F. Supp. 2d at 936; *Craigslist*, 942 F. Supp. 2d at 980; *DHI Group, Inc. v. Kent*, No. H-16-1670, 2017 WL 8794877, at *11 (S.D. Tex. Apr. 21, 2017); *Sw. Airlines*, 375 F. Supp. 3d at 705.

204. *Craigslist*, 942 F. Supp. 2d at 980; *Sw. Airlines*, 375 F. Supp. 3d at 705.

205. *DHI Group*, 2017 WL 8794877, at *11.

206. *Alan Ross Mach. Corp. v. Machinio Corp.*, No. 17-cv-3569, 2018 U.S. Dist. LEXIS 113012, at *8 (N.D. Ill. July 9, 2018).

207. *Verio, Inc.*, 356 F.3d at 405–06; *Cvent*, 739 F. Supp. 2d at 936; *Craigslist*, 942 F. Supp. 2d at 980; *DHI Group, Inc.*, 2017 WL 8794877, at *11; *Sw. Airlines*, 375 F. Supp. 3d at 705.

States that allege Lanham Act violations in conjunction with scraping practices. The outcomes of these cases could prove pivotal for defining the utility of trademark infringement, false advertising, and dilution claims in relation to scraping. These six cases can be succinctly broken up into two groups: (1) suits by Southwest Airlines and (2) food delivery app cases.

Southwest Airlines has previously tried to use trademark law to limit the uses of scraped data.²⁰⁸ In January 2021, Southwest sued travel booking website Kiwi.com, which is based in the Czech Republic, for scraping flight data from Southwest's website.²⁰⁹ Kiwi allegedly scraped flight and pricing data from Southwest's website and then populated this information on its own website.²¹⁰ On Kiwi.com, Southwest's name and iconic "Heart" logo were allegedly listed next to relevant flights.²¹¹ Southwest grounded its complaint, in large part, on claims under the Lanham Act for trademark infringement, trademark dilution, and false designation of origin.²¹² Southwest alleged that the use of the Southwest name and heart mark "has caused and is likely to cause confusion, or to cause mistake, or to deceive as to the affiliation, connection or association of Kiwi with Southwest, or as to the origin, sponsorship or approval of Kiwi's goods and services by Southwest."²¹³ Southwest also alleged that Kiwi's use of its mark would dilute the mark, as Kiwi offers deceptive and poor quality services and includes change fees and hidden fees while Southwest's brand is associated with "no change fees" and "no hidden fees."²¹⁴ On March 26, 2021, the court rejected Kiwi's motion to dismiss for improper venue.²¹⁵ On August 10, 2021, the court also rejected Kiwi's motion to dismiss for lack of personal jurisdiction.²¹⁶

In July, 2021, Southwest filed a related lawsuit against Skiplagged, Inc., an airfare search engine that is known for selling "hidden city" tickets, in which "the passenger's intended final destination is not the

208. See *Sw. Airlines*, 375 F. Supp. 3d at 705; *Sw. Airlines Co. v. Farechase, Inc.*, 318 F. Supp. 2d 435, 437 (N.D. Tex. 2004).

209. Complaint, *Sw. Airlines Co. v. Kiwi.com, Inc.*, No. 3:21-cv-00098 (N.D. Tex. Jan. 14, 2021).

210. *Id.* ¶¶ 5, 31.

211. *Id.* ¶¶ 5, 33–34.

212. *Id.* ¶¶ 63–87.

213. *Id.* ¶ 65.

214. *Id.* ¶¶ 82, 85.

215. Order, *Sw. Airlines Co. v. Kiwi.com, Inc.*, No. 3:21-cv-00098 (N.D. Tex. Mar. 26, 2021).

216. Order, *Sw. Airlines Co. v. Kiwi.com, Inc.*, No. 3:21-cv-00098 (N.D. Tex. Aug. 10, 2021).

ticketed final destination, but rather an intermediate or connecting city.”²¹⁷ In that lawsuit, Southwest alleged that Skiplagged connected to or scraped Kiwi’s website and displayed this unauthorized data on its own website.²¹⁸ Southwest even alleged that Skiplagged and Kiwi collaborated to sell “hidden city” tickets on Southwest.²¹⁹ Like in its lawsuit against Kiwi, Southwest alleged a variety of Lanham Act claims—trademark infringement, false designation of origin, and dilution—for displaying Southwest’s trademarked name and logo, along with tort claims.²²⁰

These two Southwest cases have the potential to pick up where *Southwest Airlines Co. v. Roundpipe, LLC* left off. The Northern District of Texas could provide greater clarity on which uses of trademarks are allowed in combination with scraped data. In particular, Southwest alleges trademark infringement claims on the basis of sponsorship or affiliation confusion, which, as discussed in Part II(A), is the most probable ground for trademark infringement in relation to scraping. Dilution claims in relation to scraping have not been addressed in any serious detail by courts thus far, also creating the possibility for significant guidance from the court on such claims.

The remaining four complaints are all by restaurants against food delivery services, filed between May 2020 and February 2021. In May 2020, a class action suit was filed against Grubhub alleging that Grubhub had falsely stated that non-partner restaurants were either closed or not accepting online orders.²²¹ Plaintiff CO Craft alleged that these actions constituted false advertising.²²² The suit preliminary settled in February 2021,²²³ but the settlement is being challenged by Illinois plaintiffs who were later included in the suit under a revised definition of the class, as discussed below.²²⁴

Similarly, in September 2020, a class action suit was filed by Missouri restaurant Lona’s Lil Eats against DoorDash.²²⁵ Lona’s is not one of

217. Complaint at ¶¶ 1–10, *Sw. Airlines Co. v. Skiplagged, Inc.*, No. 3:21-cv-01722, (N.D. Tex. July 23, 2021).

218. *Id.* ¶¶ 12–16.

219. *Id.* ¶¶ 17–24.

220. *Id.* ¶¶ 117–61.

221. Complaint at ¶¶ 21–33, *CO Craft, LLC v. Grubhub, Inc.*, No. 20-cv-0132733 (D. Colo. May 11, 2020).

222. *Id.* ¶¶ 44–56.

223. Stipulation and [Proposed] Order re: Settlement, *CO Craft, LLC v. Grubhub, Inc.*, No. 1:20-cv-01327 (D. Colo. Feb. 24, 2021).

224. *See infra* notes 234–243.

225. Complaint, *Lona’s Lil Eats, LLC v. Doordash, Inc.*, No. 20-cv-6703 (N.D. Cal. Sept. 24,

DoorDash's partner restaurants.²²⁶ Lona's alleged that DoorDash published false and misleading information by publishing the restaurant's name and menu information on the app, and allowing a customer to go through the process of placing an order, only to be told that (no matter the proximity of the customer) Lona's was "unavailable" due to being "out of the delivery area."²²⁷ Lona's further alleged that in the past, DoorDash included Lona's but said that it was "[c]losed."²²⁸ The class filed a claim of false advertising under the Lanham Act, as well as state false advertising and unfair competition claims.²²⁹ In January, the Northern District of California denied DoorDash's motion to dismiss, finding that Lona's had adequately pled the elements of a false advertising claim.²³⁰

In October 2020, another class action suit was filed against Grubhub in Illinois for allegedly scraping restaurants' information from their websites and listing the restaurants on the Grubhub app without their permission.²³¹ The class alleged trademark infringement as to sponsorship or affiliation by misleading consumers into believing that these restaurants had partnered with Grubhub.²³² In January 2021, the plaintiffs in the case in the U.S. District Court for the District of Colorado amended the definition of their class, which expanded it to include the Illinois plaintiffs.²³³ The Colorado class action preliminarily settled in February 2021, potentially jeopardizing the Illinois plaintiffs' claims by binding the members of the larger class.²³⁴ However, the settlement negotiations began in August 2020, months before the Illinois plaintiffs were added to the class.²³⁵ The claims of the two groups of plaintiffs are also slightly different, as the Colorado plaintiffs wanted Grubhub to only list correct information on its website,²³⁶ whereas the Illinois plaintiffs wanted

2020).

226. *Id.* ¶ 13.

227. *Id.* ¶¶ 12–16.

228. *Id.* ¶ 18.

229. *Id.* ¶¶ 40–68.

230. Order re: Motion to Dismiss, *Lona's Lil Eats, LLC v. Doordash, Inc.*, No. 20-cv-6703 (N.D. Cal. Jan. 18, 2021).

231. Complaint at ¶ 30, *Lynn Scott, LLC v. Grubhub Inc.*, No. 1:20-cv-06334 (N.D. Ill. Oct. 26, 2020).

232. *Id.* ¶¶ 34–35, 106–12 (stating a cause of action under 15 U.S.C. § 1125(a)).

233. Motion to Intervene by Lynn Scott, LLC and The Farmer's Wife, LLC at ¶ 5, *CO Craft, LLC v. Grubhub, Inc.*, No. 1:20-cv-01327 (D. Colo. Mar. 18, 2021).

234. *Id.* ¶ 6.

235. *Id.* ¶ 7.

236. Complaint at ¶ 112, *Lynn Scott, LLC v. Grubhub Inc.*, No. 1:20-cv-06334 (N.D. Ill. Oct. 26, 2020).

Grubhub to not list non-partner restaurants at all.²³⁷ The Illinois plaintiffs have disputed the settlement,²³⁸ but the Northern District of Illinois has stayed the suit while the Colorado case is pending.²³⁹

Finally, in February 2021, a fourth suit was filed against a major food delivery service. This time, the defendant was Postmates. Like the claimants in the previous two cases, Plaintiff Lucky Boy Hamburgers did not sign up to be affiliated or associated with the food delivery service.²⁴⁰ Lucky Boy alleged, however, that Postmates listed Lucky Boy in their app and incorrectly showed that the restaurant was “closed,” while in reality it was open, just not affiliated with Postmates.²⁴¹ Postmates also allegedly included Lucky Boy’s menu on their app, but with incorrect prices.²⁴² Lucky Boy alleged that Postmates had committed trademark infringement, among other claims.²⁴³ The case settled in June 2021, however, before briefs could be filed.²⁴⁴

All four complaints allege similar actions by food delivery services, but took different approaches. CO Craft and Lona’s grounded their arguments under false advertising law, whereas Lynn Scott and Lucky Boy relied on trademark infringement as to sponsorship or affiliation. *Lona’s Lil Eats, LLC v. Doordash, Inc.* is already the first decision to address a Lanham Act false advertising claim in relation to scraping. In refusing to dismiss Lona’s false advertising claim, the court credited the view that a false advertising claim can be adequately pled against certain uses of trademarks with scraped data. Future decisions in either *Lona’s Lil Eats* or *CO Craft* could provide further guidance on false advertising and scraped data. While trademark infringement claims have been decided in relation to scraping, the law is still largely unsettled,²⁴⁵ so the claims in *Lynn Scott*, as well as those in the Southwest cases against Kiwi and Skiplagged, could prove essential to establishing clearer guidance on permissible uses of trademarks in relation to scraped data. While none of

237. Complaint at ¶ 53, *CO Craft, LLC v. Grubhub, Inc.*, No. 1:20-cv-01327 (D. Colo. May 11, 2020).

238. Motion to Intervene by Lynn Scott, LLC and The Farmer’s Wife, LLC, *CO Craft, LLC v. Grubhub, Inc.*, No. 1:20-cv-01327 (D. Colo. Mar. 18, 2021).

239. Memorandum Opinion & Order, *Lynn Scott, LLC v. Grubhub, Inc.*, No. 1:20-cv-06334 (N.D. Ill. Mar. 22, 2021).

240. Complaint at ¶ 18, *Lucky Boy Hamburgers, Inc. v. Postmates Inc.*, No. 2:21-cv-01706 (C.D. Cal. Feb. 24, 2021).

241. *Id.* ¶¶ 18–20.

242. *Id.* ¶ 21.

243. *Id.* ¶¶ 29–36.

244. Notice of Voluntary Dismissal with Prejudice by Plaintiff (FRCP 41(a)(1)(A)(i)), *Lucky Boy Hamburgers, Inc. v. Postmates Inc.*, No. 21-cv-01706 (C.D. Cal. June 2, 2021).

245. *See supra* Part IV.

the answers filed so far by the defendants asserted nominative fair use, if raised, the defense could play a significant role in these cases and establishing precedent for future trademark claims against uses of scraped data.

CONCLUSION

The practice of scraping has become a central business strategy to today's online competitive market. In response, numerous legal claims have been pursued to prevent the practice of scraping itself, including under the CFAA and the Copyright Act. The Lanham Act, on the other hand, has been largely ignored by plaintiffs. While the Lanham Act cannot prevent the act of scraping itself, it has the potential to perhaps significantly limit how scraped data is used.

This Article identified three possible claims against trademark uses with scraped data: trademark infringement, false advertising, and dilution by tarnishment. As the cases discussed in Part V showed, there is a growing adoption by plaintiffs of Lanham Act claims to limit the uses of scraped data. While the case law is scant, this Article hoped to generate understanding of the range of possibilities for companies, attorneys, and judges in considering scraping through the lens of the Lanham Act. Growing uses of these claims will help to better illuminate the legal possibilities and limitations around using the Lanham Act to combat practices around the use of scraped data.

* * *