Brooklyn Journal of Corporate, Financial & Commercial Law

Volume 16
Issue 1 *SYMPOSIUM A Life Navigating the Securities Markets: A Celebration of Professor Roberta Karmel's Work, Teaching, and Mentorship*

Article 14

12-1-2021

Let the Bots Be Bots: Why the CFAA Must Be Clarified to Prevent the Selective Banning of Data Collection Facilitating Private Social Media Information Monopolization

W. Connor McRory

LET THE BOTS BE BOTS: WHY THE CFAA MUST BE CLARIFIED TO PREVENT THE SELECTIVE BANNING OF DATA COLLECTION FACILITATING PRIVATE SOCIAL MEDIA INFORMATION MONOPOLIZATION

ABSTRACT

In September 2019, the United States Court of Appeals for the Ninth Circuit granted plaintiff-startup hiQ Labs a preliminary injunction allowing it to "bot scrape" off of defendant-social networking service LinkedIn's public profiles without triggering liability under Section 1030(a)(2)(C) of the Computer Fraud & Abuse Act (CFAA) for accessing a website "without authorization." Differing judicial interpretations demonstrate the lack of clarity as to the legality of third-party bot scraping against the website owner's consent, which causes irreparable harm to businesses that rely on such practices to operate, and antitrust issues when website owners like LinkedIn and Facebook can selectively ban third-parties from collecting data on their public websites for their own private gain, while facing no comparable competition. Further, while hiQ Labs received a favorable result in the Ninth Circuit, that decision has since been vacated by the Supreme Court, and hiQ Labs ceased business operations in 2018. Hence, the tumultuous litigation's fatal business impact on hiQ Labs creates a blueprint for how other social media sites can hamstring smaller private companies' bot scraping on their sites through lengthy litigation. Thus, this Note proposes that Congress clarify the CFAA by (1) allowing any public social media profile data viewable without a log-in to be free of CFAA liability; and (2) for public profile data that can only be viewed after passing through a log-in threshold, triggering CFAA liability only if the data being collected isn't accessible after merely creating an account. Therefore, information that can be accessed by anyone with internet access, whether fully public or viewable after easily creating an account on the site, should be deemed public information that private social media companies have no authority to prevent collection of under the CFAA.

INTRODUCTION

Social media has irreversibly changed how society has functioned. For better or worse, in the past fifteen years, the percentage of Americans using at least one social media platform has jumped from 5% to 72%,[1] with an

1. *See Social Media Fact Sheet*, PEW RESEARCH CENTER, (Apr. 7, 2021), https://www.pewresearch.org/internet/fact-sheet/social-media/.

estimated 223 million Americans maintaining a social media presence.[2] With such a large portion of the population using these platforms, and only a handful of dominant platforms,[3] many advocate that private social media companies should be regulated as public utilities, or broken up into "smaller component parts" via antitrust laws – in short, the answer is government intervention.[4] These arguments are often reasonable reactions to salacious social media scandals involving election tampering, misinformation campaigns, hate speech, and data privacy violations.[5] While those scandals rightfully provoke public outrage and warrant a regulatory response, many other issues must be addressed to properly confront the disruptive cultural and economic impacts of social networks.

Particularly, Congress must address how social media companies can control who uses the data on their platforms. Specifically, when data collectors are selectively banned from collecting public data on the platforms for legitimate business purposes, those firms are in jeopardy of going out of business and the social media company has the unique opportunity to use the public data for its own business purposes.[6] Because many of these social media companies lack true competitors in the marketplace, this private policing of data collection on social media platforms "risks the possible creation of information monopolies that would disserve the public interest."[7]

The issue of social media companies selectively banning data collection on their platforms and websites generally has been resolved in the courts by interpreting the anti-hacking federal Computer Fraud and Abuse Act (CFAA).[8] Particularly relevant is the 1996 amended section, 18 U.S.C. 1030(a), which states in relevant part, "(a) Whoever. . . (2) intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains—. . . (C) information from any protected computer;" is subject to civil or criminal liability under the statute,[9] meaning private website owners can bring CFAA claims aimed at enjoining data collectors

2. *See* J. Clement, *Percentage of U.S. population who currently use any social media from 2008 to 2021,* STATISTA (April 14, 2021), https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile/.

3. *See Social Media Fact Sheet*, *supra* note 1 *(see* "Which social media platforms are most common" chart, which lists only the following 11 platforms: Facebook, Pinterest, Instagram, LinkedIn, Twitter, Snapchat, YouTube, WhatsApp, Reddit, TikTok, and Nextdoor).

4. Alex Rochefort, *Regulating Social Media Platforms: A Comparative Policy Analysis*, 25 COMM. L. POL'Y 225, 237 (2020) (see Table 2 of Mr. Rochefort's note, which discusses policy proposals for social media regulation).

5. *See id.* at 226.

6. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 992–93 (9th Cir. 2019) (describing the May 2017 cease-and-desist letter and hiQ's CEO asserting that "[i]f LinkedIn prevails… hiQ would have to 'lay off most if not all its employees, and shutter its operations'").

7. *Id.* at 1005.

8. *See* Andrew Sellars, *Twenty Years of Web Scraping and the Computer Fraud and Abuse Act*, 24 B.U. J. SCI. & TECH. L. 372, 375 (2018) (Sellars states, "Web scraping has proliferated beneath the shadow of the federal antihacking statute, the Computer Fraud and Abuse Act (CFAA).")

9. 18 U.S.C. § 1030(a)(2)(C) (2018).

using bot scraping from collecting public website data.[10] With the CFAA term "without authorization" left undefined,[11] the courts have attempted to find a middle ground between two extreme interpretations in CFAA claims against data collectors: (1) a "bright-line rule" or code-based approach, which only imposes CFAA liability if the data collector hacks or "bypasses a password gate," or (2) consider "without authorization" to impose CFAA liability on any website users violating the website's terms of service, not just hackers, i.e., a contract-based approach.[12] The lack of legislative clarity provides judges with discretion, which if used imprudently, can allow social media companies to selectively ban data analytics firms from collecting public data.

Currently, a prolonged battle reaching the three main levels of the federal judicial system questions the substantial control social media companies have over their platforms' data.[13] The action, *hiQ Labs, Inc. v. LinkedIn Corp.*, involves the professional social networking company LinkedIn Corporation (LinkedIn),[14] a wholly owned subsidiary of Microsoft,[15] and startup hiQ Labs, Inc. (hiQ).[16] In short, LinkedIn attempted to selectively ban data analytics firm hiQ from deploying automated bots (bot scraping) to aggregate and analyze public information from LinkedIn profiles "to provide hiQ's clients with insights about their employees' skills and their likelihood of being poached by competitors, which hiQ calls "people analytics."'[17] While only about 28% of the U.S. population uses LinkedIn,[18] hiQ argues that "LinkedIn has willfully acquired and maintained monopoly power in the relevant markets for professional social networking platforms."[19] Thus, hiQ

10. *See hiQ Labs*, 938 F.3d at 1000 (discussing the CFAA's anti-hacking origin and its 1996 amendment, which added section 1030(a)(2)(C) to the CFAA).

11. EF Cultural Travel BV v. Explorica, Inc., 274 F.3d 577, 582 n.10 (1st Cir. 2001).

12. Michael J. O'Connor, *The Common Law of Cyber-Trespass*, 85 BROOK. L. REV. 421, 422 (2020).

13. *See LinkedIn Corp. v. hiQ Labs Inc.*, SCOTUSBLOG, https://www.scotusblog.com/case-files/cases/linkedin-corp-v-hiq-labs-inc/ (last visited Dec. 24, 2020).

14. *See LinkedIn Defends Certiorari Petition in Bot-Scraping, Unauthorized Access Dispute*, 22-6 MEALEY'S LITIG. REP. CYBER TECH & E-COM. 13 (2020).

15. *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1103 (N.D. Cal. 2017), aff'd and remanded, 938 F.3d 985 (9th Cir. 2019).

16. *Court Rules Startup May Collect Data from LinkedIn Profiles*, ASSOCIATED PRESS (Sept. 9, 2019) https://apnews.com/article/1e1cacd92df74f48846e8bce5237b97d.

17. *LinkedIn Defends Certiorari Petition*, *supra* note 14.

18. *See Social Media Fact Sheet*, *supra* note 1 (see "Which social media platforms are most common" chart, which indicates that the percentage of Americans using LinkedIn has modestly risen from 16% to 28% from 2012 to 2020). 28% of the U.S. population is about 93,171,285 people according to the U.S. Census Bureau. *See U.S. and World Population Clock*, U.S. CENSUS BUREAU, https://www.census.gov/popclock/ (last accessed Sept. 17, 2021) (indicating that the current U.S. population is an estimated 332,754,590).

19. hiQ Labs, Inc. v. LinkedIn Corp., 485 F. Supp. 3d 1137, 1147 n.3 (N.D. Cal. 2020). LinkedIn lacks any serious competitors in the professional social networking market. *See* Jörgen Sundberg, *Why Doesn't LinkedIn Have Any Serious Competitors?*, UNDERCOVER RECRUITER, https://theundercoverrecruiter.com/linkedin-competitors/ (last visited Sept. 9, 2021).

argues there are no other platforms where it can collect the data to make its people analytics products.[20]

Bot scraping is "the retrieval of content posted on the World Wide Web through the use of a program other than a web browser or an application programming interface. . . done through a computer script that will send tailored queries to websites. . . to extract material across an array of websites or a large collection of material from a specific website."[21] The technology is practiced by investment advisers,[22] academic researchers, journalists, and more.[23] Bot scraping can be used for many arguably positive purposes, including to "preserve websites, help identify and extract data for analysis, enable consumers to find deals and discounts in online services,"[24] and arguably negative purposes, including enabling "an invasion of one's sense of privacy. . . facilitat[ing] copyright infringement at scale. . . or help[ing] people cheat in online trivia games."[25]

Some could argue hiQ isn't a sympathetic party and LinkedIn is within its rights as a private company to selectively ban bot scraping.[26] However, the facts of this case demonstrate the pitfalls of providing private companies with such legal cover.[27] For instance, LinkedIn attempted to ban hiQ while launching a similar people analytics product using similar bot scraping methods.[28] Since May 2017, when LinkedIn sent hiQ a cease-and-desist letter that intended to ban hiQ from bot scraping, the very existence of hiQ's business has been at the mercy of the courts.[29] Because the existence of hiQ and its employees' jobs were on the line as soon as this dispute began,[30] hiQ

20. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 993 (9th Cir. 2019).

21. Sellars, *supra* note 8, at 373–74.

22. *See* Peter D. Greene, Benjamin Kozinn, & Robert J. Menendez, *Is the Internet Public? A Review of the Ninth Circuit's Decision in hiQ Labs, Inc. v. LinkedIn Corporation,* LOWENSTEIN SANDLER LLP (Sept. 16, 2019), https://www.lowenstein.com/news-insights/publications/client-alerts/is-the-internet-public-a-review-of-the-ninth-circuit-s-decision-in-hiq-labs-inc-v-linkedin-corporation-investment-management.

23. *See* Sellars, *supra* note 8, at 372–73.

24. *Id.* at 374.

25. *Id.* at 374–75.

26. *See* Reply Brief for Petitioner, *LinkedIn Corp. v. hiQ Labs, Inc.* (No. 19-1116), at 9, https://www.supremecourt.gov/DocketPDF/19/19-1116/147933/20200716141411953_19-1116%20-%20Cert%20Reply.pdf (where LinkedIn argues that "any use by LinkedIn of member data is limited by LinkedIn's User Agreement and Privacy Policy").

27. *See hiQ Labs*, 938 F.3d at 1005 (where the court states that allowing "LinkedIn free rein to decide…who can collect and use data… publicly available to viewers… risks the possible creation of information monopolies that would disserve the public interest").

28. *Id.* at 991–992. The similar product is Talent Insights, which LinkedIn launched in October 2017, strikingly resemblant to hiQ's Keeper product. *See* Eric Owski, *LinkedIn's New Talent Solution Gives You the Insights to Get Ahead,* LINKEDIN TALENT BLOG (Oct. 4, 2017), https://business.linkedin.com/talent-solutions/blog/product-updates/2017/announcing-linkedin-talent-insights.

29. *See hiQ Labs*, 938 F.3d at 993 (describing the May 2017 cease-and-desist letter).

30. *See id.* (hiQ's CEO asserted that "[i]f LinkedIn prevails… hiQ would have to "lay off most if not all its employees and shutter its operations."').

claims LinkedIn engaged in irreparable harm to its business and tortious interference of contracts between hiQ and its clients.[31] Further, LinkedIn's anti-competitive behavior prompted hiQ to argue multiple antitrust-related claims.[32]

The Ninth Circuit Court of Appeals shares the business competition concerns. In September 2019, the Ninth Circuit Court of Appeals affirmed a Northern District of California decision granting hiQ a preliminary injunction enjoining LinkedIn from selectively banning hiQ bots.[33] Most strikingly, Judge Marsha S. Berzon's opinion expresses fears of treating social media companies as the owners of the data on their platforms.[34] Judge Berzon noted, "giving companies like LinkedIn free rein to decide, on any basis, who can collect and use data—data that the companies do not own, that they otherwise make publicly available to viewers, and that the companies themselves collect and use—risks the possible creation of information monopolies that would disserve the public interest."[35] Proclaiming that LinkedIn does not own the data on its platform prompts a policy question: if anyone with a computer can collect a piece of data, should anyone be able to own it?[36]

While the opinion of Judge Berzon is impactful, it is reversible and only affirms a preliminary injunction.[37] It is not a permanent enjoinment of LinkedIn's destructive anti-competitive behavior.[38] Nor does it hold the weight of statutory law. In fact, the Supreme Court vacated the Ninth Circuit's 2019 holding and remanded the case back to the Ninth Circuit to reevaluate,[39] and a myriad of motions between the parties continue in the Northern District of California before Judge Edward M. Chen.[40] Even worse, hiQ ceased operations in 2018, so hiQ's success in the litigation is rendered moot.[41]

This Note argues that (1) the relevant section of the CFAA lacks legislative clarity on the definition of "without authorization" in regards to

31. *See id.*

32. *See hiQ Labs*, 485 F. Supp. 3d at 1144.

33. *See hiQ Labs*, 938 F.3d at 1005.

34. *See id.*

35. *Id.*

36. *See id.* at 1002 (where the opinion states, regarding "public LinkedIn profiles," that "the "breaking and entering" analogue invoked so frequently during congressional consideration has no application, and the concept of "without authorization" is inapt.").

37. *See id.* at 1005.

38. *See* Reply Brief, *supra* note 26, at 10 (where LinkedIn indicates that the Supreme Court "regularly reviews decisions affirming (or denying) preliminary injunctions when, as here, the decision definitively construes a federal statute").

39. *See* Aaron Dilbeck, *Justices' CFAA Ruling Shows Contract Safeguards Insufficient*, LAW360 (June 25, 2021), https://www.law360.com/texas/articles/1396550/justices-cfaa-ruling-shows-contract-safeguards-insufficient.

40. *See hiQ Labs*, 485 F. Supp. 3d at 1143.

41. *See* Wendy Davis, *LinkedIn Says Analytics Company in Scraping Fight Quietly Shuttered In 2018*, MEDIA POST (Sept. 14, 2021), https://www.mediapost.com/publications/article/366916/linkedin-says-analytics-company-in-scraping-fight.html.

bot scraping on public websites; (2) the current *hiQ v. LinkedIn* federal litigation exemplifies the damaging effects such ambiguity can have on business competition; and (3) regardless of this particular dispute's resolution, Congress must clarify the relevant section of the CFAA to both mitigate future irreparable harm to businesses and as part of its broader antitrust policy pursuits against private technology's information monopolies that threaten not just business competition, but basic democratic values, for which this Note recognizes but does not cover.

Part I of this Note focuses on the historical background of the CFAA, the relevant amended section of the CFAA, and how a lack of legislative clarity has allowed for differing interpretations in federal courts, specifically in terms of bot scraping on public websites, including social media platforms. Further, Part II provides a detailed background of the *hiQ v. LinkedIn* case to demonstrate the harmful effects of the relevant CFAA section's ambiguity and predicts the case's outcome.

Finally, Part III proposes how Congress should amend the relevant section of the CFAA to avoid lengthy litigation pertaining to the data collection of public websites, prevent social media companies from selectively banning third-parties from use of public information on their platforms, and as part of a larger antitrust scheme that has begun to take shape,[42] to hold private technology companies accountable for their anti-competitive practices. Specifically, Congress should clarify the CFAA by (1) allowing any public social media profile data viewable without a log-in to be free of CFAA liability; and (2) for public profile data that can only be viewed after passing through a log-in threshold, triggering CFAA liability only if the data being collected isn't accessible after merely creating an account. Therefore, information that can be accessed by anyone with a computer, whether fully public or viewable after easily creating an account on the site, should be deemed public information that private social media companies have no authority to prevent collection of under the CFAA.

I. CFAA HISTORY AND ITS APPLICATION TO BOT SCRAPING

While the *LinkedIn* case involves interpretations of other legal issues, including California's Digital Millennium Copyright Act (DMCA) and the California common law of trespass,[43] the case, and therefore this Note, centers on common law interpretations of the CFAA's "without authorization" language[44] and its relation to legal issues including tortious interference of contract, irreparable harm to business, and anti-competitive business practices.[45]

42. Matthew Perlman, *Google Hit With Landmark Antitrust Suit,* LAW360 (Oct. 20, 2020), https://www.law360.com/corporate/articles/1321121.

43. *See hiQ Labs*, 938 F.3d at 992; *see* Cal. Penal Code § 502(c).

44. 18 U.S.C. § 1030(a)(2)(C).

45. *See hiQ Labs*, 938 F.3d at 993–95.

A. BRIEF HISTORY OF THE CFAA

In 1986, with the prevalence of computers still in its relative infancy, Congress passed the CFAA.[46] At the time, the statute aimed to respond to the harmful trend of hacking government computers.[47] According to a 1986 Senate Report, the CFAA intended to deter "intention[al] trespassing into someone else's computer files."[48] U.S. Senator Jeremiah Denton explained that the CFAA "makes it clear that unauthorized access to a Government computer is a trespass offense, as surely as if the offender had entered a restricted Government compound without proper authorization."[49] Hence, Congress intended that courts interpret the CFAA as an "anti-intrusion statute,"[50] not a "misappropriation" statute.[51] While largely intentioned as a criminal statute to deter sinister hackers, civil liability also attaches to the CFAA, since the statute permits a private party "who suffers damage or loss by reason of a violation of [the statute]" to bring a civil action."[52]

As the use of computers increased, Congress amended the CFAA in 1996.[53] The 1996 amendment expanded section 1030(a)(2) to protect private information, not just government computers.[54] In relevant part, this section states, "(a) Whoever. . . (2) intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains—. . . (C) information from any protected computer. . ." is subject to civil or criminal liability under the statute.[55] In passing the amended section, the Senate Judiciary Committee stated that the intention was "to increase protection for the privacy and confidentiality of computer information."[56] As Professor Orin S. Kerr states, "an authentication requirement, such as a password gate, is needed to create the necessary barrier that divides open spaces from closed spaces on the Web."[57] This historical background demonstrates a statutory intent to safeguard password-protected spaces on the internet from intrusion.

---

46. *See* United States v. Nosal (Nosal I), 676 F.3d 854, 858 (9th Cir. 2012).

47. *See hiQ Labs*, 938 F.3d at 1000 n. 11; *see also* Orin S. Kerr, *Norms of Computer Trespass*, 116 COLUM. L. REV. 1143, 1161 n. 91 (2016) (stating that the World Wide Web was invented in 1990 by Tim Berners-Lee, i.e., after the CFAA was enacted by Congress).

48. S. REP. 99-432, 1, *9, 1986 U.S.C.C.A.N. 2479.

49. *hiQ Labs*, 938 F.3d at 1000.

50. *Id.*

51. *See Nosal I*, 676 F.3d at 863 (quoting Shamrock Foods Co. v. Gast, 535 F.Supp.2d 962, 965 (D. Ariz. 2008)).

52. Jeffrey Neuburger, *Important Developments (Including Supreme Court Review) in the Interpretation of the Computer Fraud and Abuse Act,* PROSKAUER ROSE LLP (Apr. 22, 2020), https://newmedialaw.proskauer.com/2020/04/22/important-developments-including-supreme-court-review-in-the-interpretation-of-the-computer-fraud-and-abuse-act/, (quoting 18 U.S.C. § 1030(e)(6) (2018)).

53. *See* Orin S. Kerr, *Vagueness Challenges to the Computer Fraud and Abuse Act*, 94 MINN. L. REV. 1561, 1566–67 (2010).

54. *See hiQ Labs*, 938 F.3d at 1001 (citing S. Rep. No. 104-357, at 7).

55. 18 U.S.C. § 1030(a)(2)(C).

56. S. REP. NO. 104–357, at 7 (1995).

57. Kerr, *supra* note 47, at 1161.

As is seen later, when merely a login threshold is required to access data, like on Facebook, courts are left to interpret whether such an "ephemeral gate"[58] grants social media companies the authority to selectively ban data collectors for violation of its terms and conditions. Since use of some websites only requires one to create an account with a username and password,[59] can CFAA liability extend to the bot scraping of data easily accessible after signing in? Is collection of data after creating an account considered an intrusion of closed spaces on the internet when anyone with a computer can make an account? The thorniness of these issues further exemplifies the necessity of legislative clarity.

B. THE PRACTICE OF BOT SCRAPING

Bot scraping is increasingly common and utilized for several different purposes.[60] Professor Andrew Sellars describes bot scraping as:

> [T]he retrieval of content posted on the World Wide Web through the use of a program other than a web browser or an application programming interface (API). . . [i]n most cases. . . done through a computer script that will send tailored queries to websites to retrieve specific pieces of content. These requests are often sent in an automatically generated series of requests, in order to extract material across an array of websites or a large collection of material from a specific website.[61]

Bot scraping can be used for many positive purposes, including to "preserve websites, help identify and extract data for analysis, aggregate information from disparate sources. . . map out unexplored networks of servers and websites. . . lowering startup information barriers, [and] enable consumers to find deals and discounts in online services."[62] Bot scraping has also led to nefarious outcomes, including enabling "an invasion of one's sense of privacy, expos[ing] content that a website host wished instead to remain hidden, facilitat[ing] copyright infringement at scale, enable[ing] new forms of surveillance, or help[ing] people cheat in online trivia games."[63] The technique can be used by investment advisers,[64] academic researchers, journalists, and historians.[65]

---

58. O'Connor, *supra* note 12, at 445.

59. This includes Facebook. *See hiQ Labs*, 938 F.3d at 1002 (citing Facebook, Inc. v. Power Ventures, Inc., 844 F. Supp. 2d 1025, 1028 (N.D. Cal. 2012)).

60. *See* Sellars, *supra* note 8, at 374 (Sellers states that the bot scraping "technique has countless applications").

61. *Id.* at 373–74.

62. *Id.* at 374.

63. *Id.* at 374–75.

64. *See* Greene, *supra* note 22.

65. Professor Sellars outlines multiple colorful examples of bot scraping in action, including (1) a faculty and student group called the Environmental Data & Governance Initiative that collected scientific data on federal government websites in early 2017, fearing that the incoming Trump Administration would remove environmental data contradicting its political agenda, only to uncover

Widespread use of bot scraping compounded with its dubious legality breeds confusion in the courts. Largely, outcomes of CFAA cases applicable to bot scraping typically strike a middle ground between two extreme interpretations: (1) consider "without authorization" to impose CFAA liability on any website user violating the site's terms of service, not just hackers, or (2) a "bright-line rule" which only imposes CFAA liability if website user "bypasses a password gate" without the website owner's consent.[66] This ambiguity on rulings controlling the practice of bot scraping is unfavorable as its prevalence continues to rise.[67]

C. COMMON LAW INTERPRETATIONS OF THE CFAA'S "WITHOUT AUTHORIZATION" LANGUAGE IN BOT SCRAPING CASES

In a series of cases, courts have struggled to consistently interpret the CFAA's "without authorization" language in cases controlling the practice of bot scraping.[68] The main question is whether website owners have the authority under the CFAA to block data collectors from deploying automated bots to collect data on public websites, including social media platforms.[69] Two competing interpretations are the contract-based approach and code-based approach.[70] The contract-based approach argues that CFAA liability should be triggered when the website user violates the website owner's terms of service, while the code-based approach interprets the CFAA's "without authority" language to mean "the system owner would need to use technical mechanisms designed to limit access," such as a password requirement, before the website owner gains its discretionary authority to ban the user from

that the National Park Service actually removed nearly 100 documents regarding efforts to reduce carbon emissions, prompting the Service to re-post the documents; (2) journalist Alexis Madrigal who collected all 76,897 "microgenres" that Netflix had created for its users, including genres as specific as "Critically-Acclaimed Crime Movies from the 1940s," then interviewed Netflix personnel who created the microgenre system for an article published with *The Atlantic* in 2014; and (3) historian and founder of "Archive Team" Jason Scott who uses scraping tools to preserve websites before they wind-down, including Geocities, Friendster, and Miiverse. *See* Sellars, *supra* note 8, at 372–73.

66. O'Connor, *supra* note 12, at 422.

67. *See* Sellars, *supra* note 8, at 376 (where Sellers states that "both web scraping and lawsuits about web scraping have become more common").

68. *See id*. at 376–77 (where Sellars states that "practical advice on the legality of web scraping is hard to come by, and rarely extends beyond a rough combination of "try not to get caught" and "talk to a lawyer"').

69. *See id.* at 375 (where Sellars states, "For those who do not want their websites scraped, the CFAA presents a possible remedy through its broad prohibition against obtaining information by accessing a computer without authorization or by exceeding one's authorized access."); *see* Dorothy Atkins, *Israeli Ad Data Scraper Wants Facebook Access Restored,* LAW360 (Oct. 26, 2020), https://www.law360.com/articles/1322868/israeli-ad-data-scraper-wants-facebook-access-restored (which covers the ongoing Facebook, Inc. v. BrandTotal Ltd. et al case in the Northern District of California and compares it to the "similar case" of hiQ v. LinkedIn).

70. *See generally* O'Connor, *supra* note 12, at 431–47.

the website.[71] The competing approaches are particularly relevant to the *hiQ v. LinkedIn* case. In its Petition for a Writ of Certiorari filed in the Supreme Court of the United States, LinkedIn argued that the contract-based approach ascribed in the First Circuit case *EF Cultural Travel BV v. Zefer Corp.* should be adopted, in order to reverse the Ninth Circuit decision that implemented the code-based approach.[72]

1. Contract Approach Requires Website Owner's Consent

Some circuit courts have favored the contract-based approach and ruled that CFAA liability is triggered if the website user violates the website owner's terms of service.[73] If applied to the *hiQ v. LinkedIn* case, LinkedIn could selectively prevent hiQ from bot scraping under the guise that hiQ violated LinkedIn's User Agreement.[74] When applied, this contractual approach of the CFAA's "without authorization" language allows terms of service to "logically bind users just as much as employment agreements."[75] This contradicts the common sense notion that "virtually no one reads or understands" websites' terms of service agreements.[76]

LinkedIn argues that the First Circuit's contract-based approach in *EF Cultural* creates a circuit split.[77] However, it bases the alleged circuit split on mere dicta and ignores the drastically different method that the bot scraping party in *EF Cultural* used to gain access to the website.[78] In that case, the First Circuit held that violations of a confidentiality agreement or other contractual restraints could give rise to a claim for unauthorized access under the CFAA.[79] Notably, the court said, "If EF wants to ban scrapers, let it say so on the webpage or a link clearly marked as containing restrictions."[80]

71. *See* EF Cultural Travel BV v. Zefer Corp., 318 F.3d 58, 62 (1st Cir. 2003) (The court states, "A lack of authorization could be established by an explicit statement on the website restricting access. Many webpages contain lengthy limiting conditions, including limitations on the use of scrapers."); Patricia L. Bellia, *Defending Cyberproperty*, 79 N.Y.U. L. REV. 2164, 2212 (2004).

72. *See* Reply Brief, *supra* note 26, at 2 (quoting EF Cultural Travel BV v. Zefer Corp., 318 F.3d 58, 62–63 (1st Cir. 2003)).

73. *See EF Cultural*, 318 F.3d at 62 (The court states, "A lack of authorization could be established by an explicit statement on the website restricting access. Many webpages contain lengthy limiting conditions, including limitations on the use of scrapers."); *see* United States v. John, 597 F.3d 263, 269 (5th Cir. 2010).

74. *See hiQ Labs*, 938 F.3d at 992 (an argument LinkedIn made in its cease-and-desist letter).

75. O'Connor, *supra* note 12, at 433.

76. *Nosal I*, 676 F.3d at 861 (pointing out that many violations of a website's terms of service are violated habitually and unknowingly, since "Google forbade minors from using its services" and "Facebook makes it a violation of the terms of service to let anyone log into your account" and eHarmony requires that users "not provide inaccurate, misleading or false information to eHarmony or to any other user").

77. Reply Brief, *supra* note 26, at 2 (citing *EF Cultural*, 318 F.3d at 62–63).

78. *Compare hiQ Labs*, 938 F.3d at 991 (where the LinkedIn profiles hiQ bot scraped were entirely public), *with EF Cultural*, 318 F.3d at 63 (where the defendant used confidential information to gain access to the website).

79. *See EF Cultural*, 318 F.3d at 62–63.

80. *Id.* at 63.

However, the defendant bot scraping in that case gained access to the website using confidential information from a former plaintiff employee obtained in violation of a confidentiality agreement signed by former employees of the plaintiff website owner.[81] Thus, the defendant gained access using confidential information and hiQ further argues *EF Cultural* is distinguishable because it "stands only for the uncontroversial proposition that an injunction may be enforced against a third-party."[82]

Further, the Eleventh Circuit applied the contract-based approach in *U.S. v. Rodriguez*, finding that defendant violates the CFAA and "exceeds authorized access" when violating policies governing authorized use of databases.[83] Highly distinguishable from the *hiQ v. LinkedIn* case, the criminal defendant in *U.S v. Rodriguez* was a TeleService representative for the Social Security Administration who accessed "sensitive personal information" of seventeen different individuals on the Administration's database for non-work purposes in violation of the Administration's policy "prohibit[ing] an employee from obtaining information from its databases without a business reason."[84]

2. Code-Based Approach Requires an Authentication Gate

Since 2012, the Ninth Circuit Court has inconsistently led a code-based approach,[85] including in *hiQ v. LinkedIn*.[86] This approach interprets the CFAA's "without authority" language to mean that "the system owner would need to use technical mechanisms designed to limit access,"[87] such as a password requirement, before the website owner gains its discretionary authority to ban the user from the website.[88]

This approach first took form as mere dicta in 2012, allowing lower courts to maintain the broad contractual approach.[89] In two cases involving defendant access past an authentication gate, the Ninth Circuit permitted

81. The confidential information, provided by co-defendant Explorica's Chief Information Officer and plaintiff EF Cultural's former Vice President of Information Strategy Philip Gormley, included "a description of how EF's website was structured and identified the information that Explorica wanted to have copied" and "'codes'" identifying in computer shorthand the names of EF Cultural's gateway and destination cities." All of these actions violated a confidentiality agreement between Gormley and plaintiff EF Cultural. *See id.* at 62.

82. Br. In Opp'n for Resp't at 3, LinkedIn Corp. v. hiQ Labs, Inc., No. 19-1116 (U.S. June 25, 2020) [https://www.supremecourt.gov/DocketPDF/19/19-1116.pdf].

83. U.S. v. Rodriguez, 628 F.3d 1258, 1260, 1263 (11th Cir. 2010).

84. *Id.* at 1260.

85. O'Connor, *supra* note 12, at 444.

86. *See hiQ Labs*, 938 F.3d at 1005 (where the court states that allowing "LinkedIn free rein to decide…who can collect and use data… publicly available to viewers… risks the possible creation of information monopolies that would disserve the public interest").

87. *See* Bellia, *supra* note 71, at 2212.

88. *See* O'Connor, *supra* note 12, at 444.

89. *See id.*; *see Nosal I*, 676 F.3d at 863 (where the court stated that the CFAA's "general purpose is to punish hacking–the circumvention of technological access barriers–not misappropriation of trade secrets").

website owners' cease-and-desist letters to authorize bans of data collectors, even when website owner Facebook "placed pages behind an authentication gate that any person could circumvent with a free, easy-to-create account."[90] It wasn't until the *hiQ* case in 2019 that the Ninth Circuit could transform its prior dicta into a substantive holding in favor of the data collector because no authentication gate was involved – the data hiQ collected was entirely public.[91]

The Ninth Circuit's struggle to adequately and consistently apply the code-based approach[92] and past contract-based interpretations of the CFAA creates ambiguities that glaringly favor parties with deep pockets such as LinkedIn.[93] LinkedIn's reference to the *EF Cultural* case exemplifies how even an alleged circuit split can discredit the CFAA's true purpose as an anti-intrusion statute.[94] The ongoing *hiQ v. LinkedIn* case provides a crystal-clear example of why this confusion is so damaging to the nature of business in a competitive marketplace and incentivizes social media companies to create "information monopolies."[95]

II. THE *HIQ V. LINKEDIN* DISPUTE

The ongoing *hiQ v. LinkedIn* case is a real-world example of social media companies using their powers to hamstring competitors.[96] While the law interpreted in this case is important, the facts demonstrate the problem with information monopolies and why legislative clarity concerning the CFAA's "without authorization" language is needed.[97] Without clarity, judicial interpretation differs, leading to years-long litigation that puts the very existence of hiQ's business in an utter state of limbo.[98] In this case, hiQ's attorneys continue to litigate this issue, despite the company ceasing operations in 2018.[99] Even with a favorable decision by the Ninth Circuit in 2019, the Supreme Court vacated the decision in light of its interpretation of the *Van Buren* case, even though the interpretation narrowed the scope of the CFAA's applicability.[100]

90. *Nosal I*, 676 F.3d. at 445 (citing U.S. v. Nosal (Nosal II), 844 F.3d 1024, 1033–41 (9th Cir. 2016). Adding more dicta without substantive results, the Ninth Circuit stated that "a violation of the terms of use of a website—without more—cannot establish liability under the CFAA." *See* Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1067, 1067 n. 1 (9th Cir. 2016).

91. *See* O'Connor, *supra* note 12, at 445 (citing *hiQ Labs*, 938 F.3d at 1001).

92. *See id.*

93. Reply Brief, *supra* note 26, at 2 (quoting *EF Cultural*, 318 F.3d at 62-63).

94. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 1000–01 (9th Cir. 2019) (quoting Kerr, *supra* note 47, at 1161).

95. *Id.* at 1005.

96. *See id.*

97. *See* 18 U.S.C. § 1030(a)(2)(C).

98. *See hiQ Labs*, 938 F.3d at 993.

99. *See* Davis, *supra* note 41.

100. *See* Dilbeck, *supra* note 39.

A. FACTUAL BACKGROUND

LinkedIn Corporation was founded in 2002.[101] As the "world's largest professional network,"[102] LinkedIn enables users to post about their career developments and network with other professionals.[103] hiQ Labs, Inc. was founded in 2012[104] – it can still be described as a "startup."[105] Using the practice of bot scraping, hiQ primarily collects data from public LinkedIn profiles and analyzes the large data sets to create "people analytics" products called "Keeper" and "Skill Mapper" for its clients.[106] These people analytics products are analyzed data sets created "to provide hiQ's clients with insights about their employees' skills and their likelihood of being poached by competitors."[107]

hiQ's people analytics products are created only by harvesting and analyzing data on LinkedIn's public profiles.[108] Particularly, the Keeper product uses public LinkedIn profiles to determine which employees are "at the greatest risk of being recruited away," while the Skill Mapper assesses employees' skills to "identify skill gaps in their workforces so that they can offer internal training in those areas, promoting internal mobility and reducing the expense of external recruitment."[109] Some employees of hiQ clients have complained that, while the information is public, bot scraping their LinkedIn profile data for purposes of identifying which employees will be poached or lack certain skills feels like they're "under this surveillance, and they're already at risk of being ratted out to their employers with this system."[110] Nonetheless, these services have proven useful, since hiQ previously landed clients including eBay, Capital One, and GoDaddy.[111]

Before the litigation, LinkedIn was no stranger to hiQ and the products it created using data from LinkedIn's public profiles.[112] In attending hiQ-hosted

101. *See* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1103 (N.D. Cal. 2017), aff'd and remanded, 938 F.3d 985 (9th Cir. 2019).

102. *About LinkedIn*, LINKEDIN, https://about.linkedin.com (last visited Nov. 21, 2020).

103. LinkedIn's platform centers on "professional networking." *See hiQ Labs*, 273 F. Supp. 3d at 1103. In December 2016, Microsoft acquired LinkedIn for $26.2 billion. *See id*. From 2016 to 2019, the Pew Research Center reports that LinkedIn has modestly grown in users from 25% to 27% of the U.S. population. *See* SOCIAL MEDIA FACT SHEET, *supra* note 1 (see "Which social media platforms are most common" chart).

104. *See hiQ Labs*, 938 F.3d at 991.

105. *Court Rules Startup May Collect Data from LinkedIn Profiles*, *supra* note 16.

106. *See hiQ Labs*, 273 F. Supp. 3d at 1104.

107. *LinkedIn Defends Certiorari Petition*, *supra* note 14.

108. *See hiQ Labs*, 938 F.3d at 993 (where hiQ's CEO is quoted saying, "there is no current viable alternative to LinkedIn's member database to obtain data for hiQ's Keeper and Skill Mapper services").

109. *Id*.

110. Transcript of Oral Argument, hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. 2017), (No. C 17-03301), at 13.

111. *See hiQ Labs*, 938 F.3d at 993.

112. hiQ regularly organized and hosted "Elevate" conferences, which were used to network and introduce other businesses to the people analytics products hiQ had created. *See id.* at 991.

conferences,[113] LinkedIn seemed to have an ulterior motive: reconnaissance. In May 2017, LinkedIn sent hiQ a cease-and-desist letter demanding that hiQ cease bot scraping of public LinkedIn profiles, effectively making it impossible for hiQ to create its Skill Mapper and Keeper products.[114] While LinkedIn argued that hiQ could create its products scraping data from professional data on Facebook, hiQ's products must be bot scraped off public websites, and Facebook data "is not generally accessible… and therefore is not an equivalent alternative source of data."[115] Even arguments that hiQ could use publicly available information from other sources such as Google and Facebook proves disingenuous, since competitors like Xing and Viadeo[116] "are too small to rival" LinkedIn, and Workplace by Facebook "is more likely to disrupt team building apps. . . than LinkedIn's global network."[117] Thus, there really is no viable alternative professional social network, which is why the possibility that hiQ couldn't bot scrape off LinkedIn led hiQ to cease operations in 2018.[118]

LinkedIn's letter coincided with its investment into similar data analytics products. In a television interview on CBS in June 2017, LinkedIn CEO Jeff Weiner explained how the company hoped to "leverage all this extraordinary data we've been able to collect by virtue of having 500 million people join the site."[119] In October 2017, LinkedIn announced the 2018 launch of its own people analytics product called "Talent Insights," strikingly similar to hiQ's Keeper product.[120] It is therefore undisputed that LinkedIn took steps to selectively ban hiQ from scraping data off its platform while simultaneously launching a product that collects LinkedIn's public profile data for a nearly identical business purpose.

Beginning in 2015, at least ten LinkedIn employees attended these Elevate conferences, along with one employee receiving the "Elevate Impact Award" at an Elevate conference in 2016. *See id.*

113. *See id.* (where the Court describes the "Elevate" conferences hosted by hiQ and attended by LinkedIn).

114. *See* Letter from Abhishek Bajoria to Mark Weidick (May 23, 2017) [https://www.hiqlabs.com/new-imwithhiq].

115. *hiQ Labs*, 938 F.3d at 993.

116. Neither Xing nor Viadeo are serious competitors to LinkedIn in the U.S. Xing is focused on the German-speaking countries of Austria, Germany, and Switzerland, where it had an estimated 19.5 million users in the Second Quarter of 2020. *See* Evgeniya Koptyug, *Number of Xing users in D-A-CH region from 1st quarter 2013 to 2nd quarter 2020,* STATISTA (Oct. 2, 2020), https://www.statista.com/statistics/360796/xing-dach-members/. As of November 2019, Viadeo only has about 7.5 million users, largely in France. *See 7 European tech startups disrupting social media in 2019,* SILICON CANALS (Nov. 18, 2019), https://siliconcanals.com/news/european-tech-startups-disrupting-social-media/.

117. Sundberg, *supra* note 19.

118. *See* Davis, *supra* note 41.

119. *hiQ Labs*, 938 F.3d at 991–992.

120. *See* Owski, *supra* note 28. LinkedIn describes its Talent Insights product as "a tool that will help you answer questions like: which companies are you losing talent to?" *See LinkedIn Talent Insights Gives You the Real-Time, Accurate Talent Data You Need*, LINKEDIN TALENT BLOG (Sept. 25, 2018), https://business.linkedin.com/talent-solutions/blog/product-updates/2018/linkedin-talent-insights-now-available.

The selective nature of banning hiQ from LinkedIn public profiles serves to prove its anti-competitive intent. LinkedIn continued to allow other companies to bot scrape its public profiles, including Google and Parent-Microsoft's Bing, which weren't creating people analytics products.[121] To make matters worse, LinkedIn's cease-and-desist letter stalled a crucial financing round and several employees jumped ship because of LinkedIn's adverse action.[122] hiQ CEO Mark Weidick described LinkedIn's letter as placing the company in limbo, saying that a LinkedIn victory would prompt hiQ to "lay off most if not all its employees, and shutter its operations."[123] With the litigation ongoing, CTO Dan Miller left hiQ in September 2018 and the company ceased operations the same month.[124] Further, Weidick accepted a General Manager position at SimpleLegal in March 2020.[125]

Following the cease-and-desist letter, to attempt to save its business, hiQ involved the Courts.[126] As of November 2021, the litigation has yet to be resolved.[127] Yet, from a business perspective, LinkedIn's cease-and-desist letter to hiQ clearly fulfilled its goal of sinking a competitor. A company that isn't operating isn't a threat.

B. THE LITIGATION SO FAR

LinkedIn's May 2017 cease-and-desist letter prompted hiQ to immediately seek enjoinment of LinkedIn's proposed ban of bot scraping its public profiles.[128] Since LinkedIn's letter claimed hiQ's bot scraping without LinkedIn's authorization violated Section 1030 of the CFAA,[129] in June 2017, hiQ filed a complaint in the U.S. District Court for the Northern District of California, seeking a declaratory judgment that it hadn't violated the CFAA, along with claims of tortious interference of contract, irreparable

121. *See* Letter from Deepak Gupta, Farella Braun & Martel LLP to Abhishek Bajoria (May 31, 2017) [https://www.hiqlabs.com/new-imwithhiq].

122. *hiQ Labs*, 938 F.3d at 993. hiQ has failed to secure Series B or any other funding since the LinkedIn dispute began in 2017. hiQ's last funding round occurred in June 2016, when it raised $7.4 million from venture capital firm Moonshots Capital in a Venture Round. *See Venture Round – hiQ Labs,* CRUNCHBASE (June 17, 2016), https://www.crunchbase.com/organization/hiq-labs.

123. *hiQ Labs*, 938 F.3d at 993.

124. *See* Dan Miller (@DanielBMiller), LINKEDIN, (last accessed Oct. 24, 2020), https://www.linkedin.com/in/danielbmiller/ (under "Experience" section, Miller lists that he left hiQ Labs in September 2018, where he "[s]pearheaded fight to ensure the right to compete with established search engines through access to public data (*hiQ vs LinkedIn*)"); *see Davis*, *supra* note 41.

125. *See SimpleLegal Appoints Mark Weidick as General Manager*, GLOBENEWSWIRE, (Mar. 16, 2020), https://www.globenewswire.com/news-release/2020/03/16/2001002/0/en/SimpleLegal-Appoints-Mark-Weidick-as-General-Manager.html (hiQ CEO Mark Weidick accepts position at SimpleLegal).

126. *hiQ Labs*, 273 F. Supp. 3d at 1103.

127. *See* Dilbeck, *supra* note 39 (where, as of August 15, 2021, litigation in the Northern District of California and Ninth Circuit remain).

128. *hiQ Labs*, 273 F. Supp. 3d at 1103.

129. *See* Letter from Bajoria, *supra* note 114, at 2.

harm to business, and anti-competitive business practices, among other state claims.[130] After converting hiQ's action to a motion for a preliminary injunction, in August 2019, Judge Edward M. Chen granted the motion, ruling that hiQ "unquestionably faces irreparable harm in the absence of an injunction, as it will likely be driven out of business," while LinkedIn "has presented no evidence of harm, financial or otherwise resulting from hiQ's activities."[131] LinkedIn appealed, and Judge Marsha S. Berzon of the Ninth Circuit Court of Appeals affirmed in September 2019.[132]

In affirming the decision, the Ninth Circuit found that hiQ met the four required elements of a preliminary injunction: "(1) likelihood of success on the merits; (2) likelihood to suffer irreparable harm in the absence of preliminary relief; (3) that the balance of equities tip in its favor; and (4) that an injunction serves the public interest."[133] In determining the likelihood of success issue, Judge Berzon solely addressed LinkedIn's defense for its banning actions under the CFAA.[134] In rejecting its sister circuits' interpretations of the CFAA as a "misappropriation statute," Judge Berzon interpreted the CFAA as an anti-intrusion statute, stating, "the legislative history of section 1030 thus makes clear that the prohibition on unauthorized access is properly understood to apply only to private information. . . an authentication requirement, such as a password gate, is needed to create the necessary barrier that divides open spaces from closed spaces on the Web."'[135] Contrarily, when sister circuits interpret the CFAA as a "misappropriation statute," any violation of contractual restraints would create a claim of unauthorized access under the CFAA.[136] Applying the anti-intrusion interpretation of the CFAA to the facts, the Ninth Circuit found that hiQ didn't violate the CFAA, since public LinkedIn profiles are open to the general public, thus the "'breaking and entering' analogue invoked so frequently during congressional consideration has no application, and the concept of 'without authorization' is inapt."[137]

In addressing the other three elements, the Ninth Circuit found that: (1) there was likelihood of irreparable harm because, without an injunction, hiQ wouldn't be able to create its people analytics products and may go out of business;[138] (2) the balance of equities favored hiQ because the importance of hiQ staying in business outweighs possible privacy interests of scraping LinkedIn users' profiles, since the users have chosen to make their profiles

130. *See LinkedIn Defends Certiorari Petition*, *supra* note 14.
131. *hiQ Labs*, 273 F. Supp. 3d at 1107.
132. *See hiQ Labs*, 938 F.3d at 1005.
133. *See id.* at 992 (citing Winter v. Nat. Res. Def. Council, Inc., 555 U.S. 7, 20 (2008)).
134. *See id.* at 995.
135. *Id.* at 1000–01 (quoting Kerr, *supra* note 47, at 1161).
136. *See id.* at 1001 (citing *EF Cultural*, 274 F.3d at 583–84).
137. *Id.* at 1002.
138. *See id.* at 993.

public;[139] and (3) an injunction serves the public interest because allowing "companies like LinkedIn free rein to decide, on any basis, who can collect and use data—data that the companies do not own, that they otherwise make publicly available to viewers, and that the companies themselves collect and use—risks the possible creation of information monopolies that would disserve the public interest."[140]

However, the differing Ninth Circuit anti-intrusion interpretation and First Circuit misappropriation interpretation of the CFAA creates an alleged "circuit split."[141] Thus, LinkedIn filed a petition before the Supreme Court of the United States to overturn the Ninth Circuit opinion.[142] In LinkedIn's Reply Brief, it argued that the Ninth Circuit has created a circuit split because the First Circuit's *EF Cultural BV v. Explorica, Inc.* opinion states, "under the CFAA, "[a] lack of authorization could be established by an explicit statement… restricting access," and that if a "public website" operator "wants to ban scrapers, let it say so."[143] Therefore, if the Supreme Court adopted the *EF Cultural* interpretation, LinkedIn and other social media platforms could selectively ban firms like hiQ from accessing public data under the CFAA.

Meanwhile, the filings of amended complaints, motions, and orders continue before Judge Edward M. Chen of the Northern District of California. In February 2020, hiQ filed an amended complaint alleging three antitrust claims under the federal Sherman Act: "(1) Monopolization. . . in the markets for professional social networking platforms and people analytics services"; (2) "Attempted monopolization. . . in the market for people analytics services"; and (3) "Unreasonable restraint of trade" by "enter[ing] into contracts or combinations that have the effect of unreasonably restraining trade."[144] Following LinkedIn's motion to dismiss against hiQ's opposition, in September 2020, Judge Chen felt hiQ failed to allege the "parameters of the people analytics market," and provided hiQ four weeks to file a second amended complaint that properly alleged the existence of such a market needed for its antitrust claims.[145] Further, Judge Chen denied LinkedIn's motion to dismiss hiQ's claims for tortious interference and damages outright, finding that LinkedIn may still be subject to the antitrust liability underlying those claims.[146] In conditionally denying hiQ's antitrust

139. *See id.* at 995.
140. *Id.* at 1005.
141. *LinkedIn Defends Certiorari Petition*, *supra* note 14.
142. *See LinkedIn Corp. v. hiQ Labs Inc.*, *supra* note 13.
143. Reply Brief, *supra* note 26, at 2 (quoting *EF Cultural*, 318 F.3d at 62-63).
144. hiQ Labs, Inc. v. LinkedIn Corp., 485 F. Supp. 3d 1137, 1143 (N.D. Cal. 2020).
145. *hiQ Labs, Inc. v. LinkedIn Corp*., No. 17-CV-03301-EMC, 2020 WL 5408210 (N.D. Cal. Sept. 9, 2020), at 1148, 1155.
146. *See* Julia Arciga, *Judge Trims Suit Accusing LinkedIn Of Holding Monopoly,* LAW360 (Sept. 10, 2020), https://www.law360.com/competition/articles/1308822/judge-trims-suit-accusing-linkedin-of-holding-monopoly.

claims, acknowledging that LinkedIn is a leading professional social network, Judge Chen pointed out other sources of public information, "such as Google and Facebook or other industry directories and sources."[147] In the most recent decision at the District Court level in April 2021, Judge Chen merely denied hiQ's motion to dismiss LinkedIn counterclaims for breach of contract, misappropriation, and trespass to chattels.[148] As is apparent, the litigation demonstrates no clear end in sight.

Theoretically, Google and Facebook seem like reasonable alternatives to LinkedIn. However, setting aside the legal definitions, users primarily use LinkedIn for professional networking. Users likely use LinkedIn to connect with colleagues, potential employers or potential employees in their professional industry.[149] Users likely add their complete employment history, skills relevant to their professional pursuits, and only interact with posts relevant to the professional world instead of leisurely activities.[150] An alternative social networking site that could be used for those purposes doesn't exist in the United States. It's unlikely that LinkedIn users use YouTube, Twitter, Instagram, or Facebook the same way as they use LinkedIn.[151] Yet, Judge Chen deemed Google and Facebook as viable alternatives for hiQ to collect the data needed for its products.[152]

Thus, hiQ was left simultaneously litigating before the District Court and awaiting the Ninth Circuit's reevaluation of its previous opinion in light of the Supreme Court's *Van Buren* decision.[153] While the LinkedIn petition was pending, the Supreme Court reviewed an Eleventh Circuit interpretation of the CFAA in the case *U.S. v. Van Buren*.[154] In *Van* Buren, which regards the criminal liability underlying the CFAA, the Eleventh Circuit affirmed the conviction of a police officer who used police databases for personal and financial gain.[155] However, on June 3, 2021, the Supreme Court reversed the conviction, interpreting the CFAA narrowly, holding that "exceeds

147. *hiQ Labs*, 485 F. Supp. 3d at 1149.

148. *See* Zarish Baig and Kristin L. Bryan, *hiQ LinkedIn Data Scraping CFAA Ruling Delayed Pending SCOTUS Decision,* NAT'L L. REV. (Apr. 26, 2021), https://www.natlawreview.com/article/hiq-linkedin-data-scraping-cfaa-ruling-delayed-pending-scotus-decision.

149. *See* Dave Johnson, *'What is LinkedIn?': A Beginner's Guide to the Popular Professional Networking and Career Development Site*, BUS. INSIDER (Sept. 6, 2019), https://www.businessinsider.com/what-is-linkedin (providing a "Beginner's Guide" on how LinkedIn works, in case the reader is unfamiliar).

150. *See id.*

151. *See id.* (The article states, "But unlike most social networks, LinkedIn is a professional networking site, designed to help people make business connections, share their experiences and resumes, and find jobs.").

152. *hiQ Labs*, 485 F. Supp. 3d at 1149.

153. *See id.* (the current ongoing litigation at the district court-level); LinkedIn Corp. v. hiQ Labs, Inc., No. 19-1116, 2021 WL 2405144, at *1.

154. *See Van Buren v. United States*, SCOTUSBLOG (last visited Sept. 17, 2021), https://www.scotusblog.com/case-files/cases/van-buren-v-united-states/.

155. *See* U.S. v. Van Buren, 940 F.3d 1192, 1208 (11th Cir. 2019), *cert. granted*, (U.S. Apr. 20, 2020) (No. 19-783).

authorized access" means "the act of entering a part of the system to which a computer user lacks access privileges."[156] Notably, the Supreme Court applies this holding to civil cases as well.[157] Therefore, on June 14, 2021, the Supreme Court vacated the Ninth Circuit 2019 *hiQ* holding and remanded the case back down to the Ninth Circuit "for further consideration in light of *Van Buren*."[158]

Beyond the abstract legal and procedural arguments, there are short-term, real world, drastic consequences arising out of this case. hiQ could ultimately obtain injunctive relief once the case is fully remanded back to the District Court, a vindication of its legal position. However, hiQ has been litigating this case for over four years[159] – the company has only existed for nine years[160] and hasn't been in operation since 2018.[161] During this time, the prospect of LinkedIn banning hiQ from collecting "publicly available" data on a platform with over 500 million profiles caused hiQ to go out of business.[162]

There are also long-term, real world, drastic consequences to the outcome of this case. Even if LinkedIn eventually lost, the litigation sunk a competitor,[163] while LinkedIn created a comparable product.[164] Therefore, LinkedIn and other large platforms can take similar measures to siphon off smaller, startup competitors' clients, gaining clients that are spooked away from a competitor's legal troubles, which would obstruct a needed round of funding for said competitors like hiQ.[165] LinkedIn has effectively provided a blueprint to other social media companies who want to selectively thwart third-party data collection for any purpose it sees fit. LinkedIn's victory sets a precedent that accommodates social media companies' information monopolies over the public data on their platforms. Regardless, serious antitrust issues arise while differing court interpretations forecast prolonged litigation with never-ending appeals and motions. With shareholders, quarterly earnings reports, and all corporations' "pathological pursuit of profit and power,"[166] LinkedIn and others with similarly advantageous

156. *See* Dilbeck, *supra* note 39, *quoting* U.S. v. Van Buren, 940 F.3d 1192, 1208 (11th Cir. 2019), *rev'd and remanded*, No. 19-783, 2021 WL 2229206, at *9 (June 3, 2021).

157. U.S. v. Van Buren, 940 F.3d 1192, 1208 (11th Cir. 2019), *rev'd and remanded*, No. 19-783, 2021 WL 2229206, at *10 (June 3, 2021).

158. LinkedIn Corp. v. hiQ Labs, Inc., No. 19-1116, 2021 U.S. LEXIS 2997 (June 14, 2021).

159. *See LinkedIn Defends Certiorari Petition*, *supra* note 14 (stating that the cease-and-desist letter was sent in May 2017).

160. *See hiQ Labs*, 938 F.3d at 991 (stating hiQ was founded in 2012).

161. *See* Davis, *supra* note 41.

162. *hiQ Labs*, 485 F. Supp. 3d at 1142; *see* Davis, *supra* note 41.

163. *see* Davis, *supra* note 41.

164. *See LinkedIn Talent Insights, supra* note 120

165. *See hiQ Labs*, 938 F.3d at 993 (hiQ CEO stated that the cease-and-desist letter stalled a round of funding).

166. *See generally* JOEL BAKAN, THE CORPORATION: THE PATHOLOGICAL PURSUIT OF PROFIT AND POWER (Simon & Schuster, Inc.) (2005).

positions are practically required to exploit their control of the data on their platforms until their powers are checked.

Naturally, this confusion invites clarity from Congress because the length of litigation and an allegedly unresolved circuit split[167] incentivizes social media companies to engage in anti-competitive selective bot scraping bans, then file motion-after-motion and appeal-after-appeal under the premise that they believe the bot scraping on their public platforms violates the CFAA.

III. HOW CONGRESS CAN CLARIFY THE CFAA

Currently, the term "without authorization" in the CFAA remains undefined, despite the 1996 Amendment significantly broadening Section 1030(a)(2) of the CFAA "from prohibiting unauthorized access that obtains certain sensitive information to prohibiting unauthorized access that obtains any information."[168] As social media platforms became ubiquitous and conducting internet searches of friends' and acquaintances' public profiles became commonplace,[169] Congress neglected to explain whether social media companies could selectively deny access to such public information for any reason it sees fit. If an authentication gate is the only relevant factor, as the Ninth Circuit interprets "without authorization" in the *hiQ* case,[170] then Congress must say so. The differing judicial interpretations currently enable LinkedIn to plausibly point to the CFAA as legal cover for selective banning practices, while legislative clarity would make such arguments dead on arrival.

A. PRACTICAL ISSUES WITH BROADENING AND NARROWING THE CFAA

An amendment to Section 1030(a)(2) of the CFAA should practically consider how information is viewed on each major social media platform and whether such information is or isn't behind an authentication gate. For instance, LinkedIn users have the discretion to make their public profiles viewable by simply googling the user's name, seeing the person's LinkedIn profile as a search result, then clicking on that result to view the person's profile, without the searcher ever creating a LinkedIn account.[171] Since information that appears through this simple Google search requires no

167. *See* Reply Brief, *supra* note 26, at 2 (quoting *EF Cultural*, 318 F.3d at 62–63).

168. *See EF Cultural*, 274 F.3d at 582 n.10; O'Connor, *supra* note 12, at 425.

169. *See Does Googling Your Friends Make Me a Stalker?* THE BOLD ITALIC (Feb. 12, 2014), https://thebolditalic.com/does-googling-your-friends-make-me-a-stalker-the-bold-italic-san-francisco-be24c2c6832d (discussing how "we create such immersive profiles of ourselves for the world to see that it's almost impossible to remain anonymous").

170. *See* O'Connor, *supra* note 12, at 445 n. 170.

171. *See hiQ Labs*, 938 F.3d at 990 (explaining how LinkedIn users can both make their profiles visible to non-LinkedIn members or can limit certain profile data to LinkedIn members or even just the users' "connections").

passage through an authentication gate, Congress could specify that data collected by this means creates no liability under the CFAA, which would essentially codify the since-vacated *hiQ v. LinkedIn* decision in the Ninth Circuit.[172]

However, Congress could extend data collection freedom further under the CFAA, as some parties to CFAA litigation have argued.[173] For instance, what if the data collector creates a LinkedIn account for the express purpose of viewing information on LinkedIn profiles that couldn't be viewed without an account?[174] Is the creation of an account and subsequent sign-in to the account considered an authentication gate that provides the social media company the discretionary authority under the CFAA to ban that user? While the courts didn't address this issue in the *hiQ* case, the Ninth Circuit said yes, at least when the user bot scraped Facebook profile data.[175] In *Facebook v. Power Ventures*, the Court held that the CFAA granted Facebook the authority to ban bot scrapers, even though "Facebook placed pages behind an authentication gate that any person could circumvent with a free, easy-to-create account."[176]

More Facebook litigation ensued, proving that the *Facebook v. Power Ventures* decision failed to settle the CFAA definition of "without authorization" within the context of Facebook cases, let alone social media platforms generally.[177] In October 2020, Facebook sued Israeli-based companies BrandTotal Ltd. and Unimania Inc. for violating Facebook's terms of services by collecting user profile data for marketing purposes.[178] In that case, BrandTotal used internet browser extensions to scrape password-protected data from the browser, irrespective of who was using the computer at a given time, so it's unclear whether individual Facebook users were consistently consenting to BrandTotal's scraping of their data.[179] Given the Ninth Circuit precedent in the *hiQ* case,[180] attorneys for BrandTotal argued that because Facebook users agree to share data with them, BrandTotal doesn't "have a service to provide" and would go out of business without a temporary restraining order against Facebook.[181] BrandTotal further asserted

172. *See id.*; LinkedIn Corp. v. hiQ Labs, Inc., No. 19-1116, 2021 WL 2405144, at *1.

173. *See* Atkins, *supra* note 69 (where BrandTotal and Unimania argue that "users agreed to share their data with the companies and that the data collected was "innocuous" information that is not sensitive or tied to individuals").

174. LinkedIn users can limit certain data to just LinkedIn members. *See* hiQ Labs, 938 F.3d at 990.

175. *See generally* Facebook, Inc. v. Power Ventures, 844 F.3d 1058 (9th Cir. 2016).

176. O'Connor, *supra* note 12, at 445, (citing *Power Ventures*, 844 F.3d at 1067 n.1).

177. *See* Atkins, *supra* note 69.

178. *See* Sarah Perez & Zach Whittaker, *Facebook Sues Two Companies Engaged In Data Scraping Operations,* TECHCRUNCH (Oct. 1, 2020), https://techcrunch.com/2020/10/01/facebook-sues-two-companies-engaged-in-data-scraping-operations/.

179. *See* Atkins, *supra* note 69.

180. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985, 990 (9th Cir. 2019).

181. *See* Atkins, *supra* note 69.

that "Facebook has a monopoly over the data," so it "should not have to comply with Facebook's terms and conditions."[182] Further complicating this case is Facebook's consent decree with the Federal Trade Commission, "which required the social media giant to take steps to protect user data from third-parties."[183]

Facebook doesn't just wrangle with private companies and government agencies on bot scraping issues, but academic institutions as well. On October 16, 2020, Facebook sent a letter to the New York University Ad Observatory threatening legal action if the Observatory didn't cease scraping data from political ads on Facebook.[184] Instead of a private data collector, the Ad Observatory is an NYU Engineering School research project with over 6,000 volunteers which, among other things, uncovered that "Facebook was not labeling all political ads to show who had paid for them as its own disclosure rules require."[185] Andy Sellars[186] outlined the Ad Observatory's practices:

> Facebook calls what the Ad Observatory does "scraping," but that's not what this is. The data source, the "Ad Observer," is a browser plugin. Researches often use plugins like these to collect information for data science and algorithmic accountability projects. It isn't a script serially visiting websites in some independent or autonomous way from a human at a browser. It is a small piece of software, installed by the data subject, that extracts data from the local copy of a webpage that is inherently made when a person loads website onto a computer… A plugin more naturally gives an opportunity for data subject consent, or revocation of consent. You can, as the Ad Observer does, explain to the data subject what you plan to do before you do it, and give the data subject an unambiguous way to manifest their consent… scraping data off of sites is rarely done with this level of informed user choice.[187]

Thus, the type of data collection by NYU Ad Observatory doesn't raise the same user consent concerns that Judge Spero raised in the *BrandTotal* case.[188] In defending its right to selectively ban the NYU Ad Observatory,

182. *Id.*

183. *Id.*

184. *See* Kim Lyons, *Facebook Wants the NYU Ad Observer To Quit Collecting Data About Its Ad Targeting,* THE VERGE (Oct. 23, 2020), https://www.theverge.com/2020/10/23/21531232/facebook-nyu-ads-politics-data-election.

185. *Id.*

186. Mr. Sellars is the Founding Director of the BU/MIT Technology Law Clinic, a collaboration of Boston University School of Law and the Massachusetts Institute of Technology. *See* Andy Sellars, *Facebook's Threat To The NYU Ad Observatory Is An Attack On Ethical Research,* NIEMAN JOURNALISM LAB AT HARV. (Oct. 29, 2020), https://www.niemanlab.org/2020/10/facebooks-threat-to-the-nyu-ad-observatory-is-an-attack-on-ethical-research/.

187. *Id.*

188. *See* Atkins, *supra* note 69 (where Magistrate Judge Spero raised doubts as to whether a user consistently consents to BrandTotal's use of their data, since "any browser that has the extension collects data regardless of who is using the computer").

Facebook deploys a familiar argument: that the CFAA triggers liability when bot scrapers violate Facebook's terms of service, since "Facebook is not a 'public' website because users log in to see its contents," thus distinguishing itself from public platforms like Twitter, Google, and LinkedIn.[189]

In August 2021, "Facebook shut down accounts belonging to two academic researchers. . . cutting off their ability to study political ads and misinformation on the world's biggest social network. . . accus[ing] the academics of engaging in 'unauthorized scraping' and compromising user privacy on the platform, claims that Facebook's many critics are slamming as a thin pretense for killing the transparency work."[190] In response, Firefox developer Mozilla stated that it "reviewed [the Ad Observatory] twice, conducting both a code review and examining the consent flow" before recommending the browser extension through its storefront, and that Facebook's claims "simply do not hold water."[191] Clearly, the courts haven't settled whether CFAA liability is triggered once the data collector has passed an initial authentication gate, so private companies can act zealously against bot scrapers.

As demonstrated above, even though both BrandTotal and the NYU Ad Observatory passed through Facebook's login threshold and therefore violated Facebook's terms of service, the facts of the two cases are highly distinguishable. While BrandTotal argued that the CFAA allows them to bot scrape in spite of Facebook's terms and conditions,[192] Facebook argued that both private company BrandTotal and the NYU Ad Observatory can't bot scrape Facebook user data because user logins constitute an "authentication" for CFAA purposes.[193] To resolve this, Congress could favor data collectors, even once such companies are scraping data that can only be obtained after passing through a log-in. Since anyone can create a Facebook account, one can argue that the CFAA shouldn't trigger liability for use of a platform that can be easily accessed by anyone with a computer. Such a stance by Congress would extend the rights of data collectors past the precedent set by the *hiQ* and *Facebook v. Power Ventures* decisions.

Given the slippery slope of allowing the bot scraping of profile data that can only be viewed with an account,[194] Congress should consider limiting

189. Sellars, *supra* note 186.

190. Taylor Hatmaker, *Facebook Cuts Off NYU Researcher Access, Prompting Rebuke from Lawmakers*, TECHCRUNCH (Aug. 4, 2021), https://techcrunch.com/2021/08/04/facebook-ad-observatory-nyu-researchers/?guccounter=1.

191. *Id.* In addition, Senator Ron Wyden (D-OR) and Senator Mark Warner (D-VA) expressed their displeasure with Facebook's actions. *See id.*

192. *See* Atkins, *supra* note 69.

193. *See id.*; Sellars, *supra* note 186.

194. In 2016, British data analytics firm Cambridge Analytica bot-scraped Facebook user profiles to influence voters leading up to the 2016 U.S. Presidential election between Donald Trump and Hillary Clinton. *See* Natasha Lomas, *Facebook Staff Raised Concerns About Cambridge Analytica in September 2015, Per Court Filing,* TECHCRUNCH (Mar. 22, 2019),

data collectors' immunity from the CFAA to data that is fully public, like in the *hiQ* case. Such a clarification would be easy for courts to follow, and ambiguous interpretations of "without authorization" would be less likely to arise. However, such a clarification would deem the data collection by the NYU Ad Observatory[195] in violation of the CFAA, despite the public interest in academic research of proliferation of political advertising on Facebook, given scandals including Cambridge Analytica.[196] Andy Sellars argues that NYU Ad Observatory's data collection should be lawfully protected, "especially when it provides us a rare insight into how political advertisements are influencing voters in" meaningful national elections.[197]

Further, allowing a simple log-in to serve as the threshold that provides social media companies the authority to ban data collection allows Facebook to hold an information monopoly over the data on its platform. Facebook greatly exceeds LinkedIn in terms of number of users.[198] If Congress intends to amend the CFAA to ensure public data on social media platforms is not privately owned and can be collected by anyone with a computer, then limiting data collectors' immunity from the CFAA to data that can be viewed without a log-in threshold may protect data collectors' bot scraping of LinkedIn, Google, or Twitter data, but will only refine Facebook's information monopoly. Given the privacy concerns for users, such as in the *BrandTotal* case, judicial discretion may be needed when data collectors gather data on Facebook's site in violation of its terms and services, since user consent can often be obtained in an imperfect manner that raises consumer protection concerns.[199]

B. THE POLICY SOLUTION

The demonstrable complexities of website authentication gates indicate there is no one-size-fits-all solution. In assessing how CFAA clarification can best meet policy goals, Congress must apply the CFAA differently for each social media platform. For profile data that can be viewed without a log-in, such as on Twitter and LinkedIn,[200] website owners would have no authority under the CFAA to ban bot scraping, which would codify the Ninth Circuit's

https://techcrunch.com/2019/03/22/facebook-staff-raised-concerns-about-cambridge-analytica-in-september-2015-per-court-filing/.

195. *See* Sellars, *supra* note 186.

196. *See* Lomas, *supra* note 194.

197. Sellars, *supra* note 186.

198. *See Social Media Fact Sheet*, *supra* note 1 (as of 2019, 27% of U.S. adults use LinkedIn, while 69% of U.S. adults use Facebook).

199. *See* Atkins, *supra* note 69 (explaining that Magistrate Judge Spero raised doubts as to whether a user consistently consents to BrandTotal's use of their data, since "any browser that has the extension collects data regardless of who is using the computer").

200. *See* Sellars, *supra* note 186 (saying that Facebook claims that it "is not a "public" website because users log in to see its contents… a power that is denied to other major platforms like Twitter, Google, or LinkedIn").

2019 holding in the *hiQ* case.[201] For profile data that can only be viewed after passing through a log-in threshold, like Facebook,[202] the issue of user privacy becomes the primary concern. To decide whether the data collector may bot scrape such profile data, website owners would only have authority under the CFAA to ban data collectors if the data being collected isn't accessible after merely passing the log-in gate, since such a hurdle creates no privacy concerns. If data collection involves any further interaction with users, then the social media network would have the authority under the CFAA to ban such activities. Such clarity would avoid subjecting data collectors to unpredictable litigation outcomes, yet protect the privacy interests of the user, which were never at risk in the *hiQ* case.[203]

CONCLUSION

Going forward, clarifying the legality of bot scraping on social media sites is paramount. CFAA ambiguity caused hiQ to suffer through endless litigation with LinkedIn and the termination of its operations,[204] creating a blueprint for other social media companies looking to hold a monopoly over their platforms' data by hamstringing bot scraping competitors. For this reason, Congress should clarify the CFAA by (1) allowing any public social media profile data viewable without a log-in to be free from triggering CFAA liability; and (2) for public profile data that can only be viewed after passing through a log-in threshold, the website owners should only have the authority under the CFAA to selectively ban bot scraping if the data being collected isn't accessible after merely creating an account. Therefore, information that can be accessed by anyone with a computer, whether fully public or viewable after easily creating an account on the site, should be deemed public information that private social media companies have no authority to prevent collection of under the CFAA. The damage done by LinkedIn should prompt lawmakers, regulators, and private companies to extinguish social media platforms' exclusive right to data in public view.

W. Connor McRory[*]

201. *See* hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985 (9th Cir. 2019).

202. *See* Sellars, *supra* note 186.

203. *See hiQ Labs*, 938 F.3d at 995 (saying "even if some users retain some privacy interests in their information notwithstanding their decision to make their profiles public, we cannot… conclude that those interests… are significant enough to outweigh hiQ's interest in continuing its business").

204. *See LinkedIn Defends Certiorari Petition*, *supra* note 14 (stating that the cease-and-desist letter was sent in May 2017); Davis, *supra* note 41.

* B.A. Fordham University, 2018; J.D. Candidate, Brooklyn Law School, 2022. I would like to thank the entire staff of the Brooklyn Journal of Corporate, Financial & Commercial Law, especially Marissa Brown, Daniel Finnegan, and Cindy Chan for their valued feedback and assistance. Thank you to my parents and my family for their ceaseless, unending encouragement; Dina D'cruze for always loving, motivating, and supporting me.