DATA SCRAPING & THE DOWNFALL OF PUBLIC ANONYMITY

Daniel B. Rankin[*]

Some open-internet advocates have applauded the Ninth Circuit's recent decision in *hiQ Labs v. LinkedIn*,[1] which held that scraping data off a public website didn't violate the Computer Fraud & Abuse Act. The decision was thought consistent with the notion that access to public data on the web is inherently authorized.[2] While the Ninth Circuit's decision appears to further ensconce that norm, it may have wrested from companies a tool that could safeguard our privacy.

* * *

Law enforcement's facial-recognition capabilities have traditionally been tethered to "government-provided images, such as mug shots and driver's license photos."[3] This limited scale confined law enforcement to searching through a narrow class of potential suspects. Lack of data (if that could ever be a problem these days) fettered the ability to swiftly produce accurate matches.

But an up-and-coming tech start-up, Clearview AI, is changing that. Clearview AI's product, a facial-recognition app, has been eagerly purchased and used by hundreds of law-enforcement agencies.[4] The game-changing feature of Clearview's app is its database of roughly three billion images. Clearview has compiled all these images by scraping the data off popular websites such as Facebook, Instagram, and YouTube.[5] Clearview app users can thus take a picture of anyone, upload it, and within seconds, see all the photos of that person that appear on the web.[6]

Apart from being a "valuable crime-solving tool,"[7] the concerns for abuse of this powerful technology may be warranted. If used as intended, Clearview's app could destroy public anonymity. No longer could anyone walk down the street and remain a nameless stranger. Anytime your face is captured by a surveillance camera

---

[*] Daniel B. Rankin, Law Clerk, Supreme Court of Texas. The views expressed in this article do not necessarily reflect the views of any justice on the Supreme Court of Texas.

[1] *E.g.*, Orin Kerr, *Scraping a Public Website Doesn't Violate the CFAA, Ninth Circuit (Mostly) Holds*, VOLOKH CONSPIRACY (Sept. 9, 2019), https://reason.com/2019/09/09/scraping-a-public-website-doesnt-violate-the-cfaa-ninth-circuit-mostly-holds/.

[2] *See* Orin Kerr, *Norms of Computer Trespass*, 116 COLUM. L. REV. 1143, 1162 (2016); *see also* Kerr, *supra* note 1 (saying that the Ninth Circuit's holding "presumes a right to open access under the CFAA unless there is some technological measure placed on access").

[3] Kashmir Hill, *The Secretive Company That Might End Privacy as We Know It*, N.Y. TIMES (Jan. 1, 2020), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

[4] *Id.*

[5] *Id.*

[6] *Id.*

[7] *Id.* (paraphrasing a statement by a Clearview investor).

or smartphone, your entire web presence can be reached and analyzed.[8] This, of course, allows app users to collect a trove of information on the person being searched: name, birthday, home address, friends, whereabouts, place of work, and more.

The wellsprings of Clearview's vast database—social-media companies—have attempted to stop Clearview from scaping their data.[9] LinkedIn, for example, sent Clearview a cease-and-desist letter, stating that scraping users' data violated its policies.[10] But are there any legal consequences for violating a site's terms of service, particularly when the violator is put on notice from the company itself?

Under the Computer Fraud & Abuse Act (CFAA), the federal anti-hacking statute, there are criminal penalties for accessing a computer "without authorization."[11] In addition to imposing potential felony liability, the CFAA also creates a private right of action, which permits hacking victims to sue the hacker and obtain damages and equitable relief.[12] For Clearview, the question comes down to this: What counts as accessing a computer "without authorization"? This deceivingly simple question has engendered much confusion. The Ninth Circuit, the circuit with a relatively heavy legal–tech docket, has been a leading voice on this issue.

In 2016, in *Facebook v. Power Ventures*, the question before the Ninth Circuit was whether Power Ventures could be held liable under the CFAA when, despite a cease-and-desist letter from Facebook, it accessed Facebook users' profiles with the users' permission.[13] Facebook, in seeking to hold Power Ventures civilly liable, sought a broad reading of the CFAA. Facebook pointed out, too, that what Power Ventures was doing violated Facebook's terms of service.[14] The Ninth Circuit agreed, holding that Power Ventures' access was "unauthorized" because Facebook's letter withdrew Power Venture's permission to access Facebook's servers.[15] Facebook could thus seek damages and injunctive relief under the CFAA for Power Ventures' unauthorized data scraping.[16]

---

[8] Clearview's app also allows the user performing the search to find not only pictures that the targeted person posts but also pictures of that targeted person that other people post, even if the targeted person was not tagged in the photo. *See id.* ("[T]he app helped identify . . . a person who was accused of sexually abusing a child whose face appeared in the mirror of someone else's gym photo . . . .").
[9] Jon Porter, *Facebook and LinkedIn are latest to demand Clearview stop scraping images for facial recognition tech*, THE VERGE (Feb. 6, 2020), https://www.theverge.com/2020/2/6/21126063/facebook-clearview-ai-image-scraping-facial-recognition-database-terms-of-service-twitter-youtube.
[10] *Id.*
[11] 18 U.SC. § 1030(a) (2018).
[12] *Id.* § 1030(g) (2018).
[13] Facebook v. Power Ventures, 844 F.3d 1058, 1062 (9th Cir. 2016).
[14] *Id.*
[15] *Id.* at 1067.
[16] *See* 18 U.S.C. § 1030(g) (2018) (creating a civil cause of action and a damages remedy for victims of computer hacking).

The Ninth Circuit substantially narrowed the precedential scope of *Power Ventures*, however, in *hiQ Labs v. LinkedIn*.[17] The dispute between hiQ and LinkedIn began when hiQ, a data-analytics company, scraped data off the public-facing parts of LinkedIn.[18] Like Facebook, LinkedIn sent hiQ a cease-and-desist letter, telling it to stop accessing and copying data from its server.[19] In response, hiQ sued LinkedIn, seeking injunctive relief and a declaratory judgment that what it was doing did not run afoul of the CFAA.[20] The question before the Ninth Circuit, then, was much like the one in *Power Ventures*.

Yet the decision came out much differently. In distinguishing *Power Ventures*, the panel held that hiQ's scraping of LinkedIn's data, even after the cease-and-desist letter, did not violate the CFAA.[21] Unlike *Power Ventures*, the panel said, hiQ was scraping data that "was available to anyone with a web browser," and thus did not need a username and password.[22] According to the panel, this distinction required a different result because legislative history made clear that the CFAA was meant to protect "information delineated as private through use of a permission requirement of some sort."[23] The panel therefore concluded that "when a computer network generally permits public access to its data, a user's accessing that publicly available data will not constitute access without authorization under the CFAA."[24]

With *hiQ* on the books, where does the practice of data scraping stand under the CFAA? According to Professor Orin Kerr, "*HiQ Labs* now places a critical limit on *Power Ventures*."[25] "[T]he cease-and-desist letter," Kerr says, now "only controls access to rights to non-public data."[26] So if someone wants to scrape data off a public website, they can likely do so without fear of having the broad CFAA hammer dropped on them.[27]

HiQ was a win for the open internet. But how about for privacy?

After getting word that Clearview was scrapping data to compile its vast image database, notable companies, such as Facebook, LinkedIn, Twitter, Google,

---

[17] 938 F.3d 985 (9th Cir. 2019).
[18] *Id.* at 992.
[19] *Id.*
[20] *Id.*
[21] *Id.* at 1003–04.
[22] *Id.* at 1002.
[23] *Id.* at 1001.
[24] *Id.* at 1003.
[25] Orin Kerr, *Scraping a Public Website Doesn't Violate the CFAA, Ninth Circuit (Mostly) Holds*, VOLOKH CONSPIRACY (Sept. 9, 2019), https://reason.com/2019/09/09/scraping-a-public-website-doesnt-violate-the-cfaa-ninth-circuit-mostly-holds/.
[26] *Id.*
[27] *See id.* ("Putting the cases together, the Ninth Circuit law right now seems to go like this. You can scrape a public website, and you can violate terms of service, without violating the CFAA.")

and YouTube, sent Clearview cease-and-desist letters.[28] Some of the letters informed Clearview that data scraping violated their websites' terms of service.[29] This face-off has its fair share of irony: "Although we've long criticized these platforms for profiting off our data," Rebeca Heilweil of Open Sourced writes, "we're now potentially reliant on these companies to defend us from a dystopian world of facial recognition."[30]

This newfound (and ironic) reliance, however, could be misplaced. Under *Power Ventures*, it appeared social-media companies could leverage the CFAA to prevent Clearview-types from collecting our data. But the Ninth Circuit's decision in *hiQ* dashes that hope. Clearview scrapes data off of public-facing websites—precisely what hiQ did to LinkedIn.[31] So social-media companies, equipped with their cease-and-desist letters, are mostly[32] powerless to prevent Clearview from collecting our posted images. And if they're powerless to stop that, then our privacy will be further eroded.

Hello open internet, but goodbye public anonymity.

---

[28] Rebeca Heilweil, *The world's scariest facial recognition software, explained*, Vox (Feb. 11, 2020), https://www.vox.com/recode/2020/2/11/21131991/clearview-ai-facial-recognition-database-law-enforcement.

[29] Alfred Ng & Steven Musil, *Clearview AI hit with cease-and-desist letters from Google, Facebook over facial recognition collection*, CNET (Feb. 5, 2020), https://www.cnet.com/news/clearview-ai-hit-with-cease-and-desist-from-google-over-facial-recognition-collection/.

[30] Heilweil, *supra* note 28.

[31] *hiQ Labs, Inc. v. LinkedIn Corp*, 938 F.3d 985, 992 (9th Cir.).

[32] While reliance on the CFAA may prove unfruitful, social-media companies could still make use of state-law equivalents of the CFAA, which are oftentimes worded differently. *See, e.g.*, TEX. PEN. CODE ANN. § 33.02(a) (West 2015) ("A person commits an offense if the person knowingly accesses a computer, computer network, or computer system without the effective consent of the owner."); CAL. PEN. CODE § 502(c)(2) (West 2020) (imposing liability on anyone who "[k]nowingly accesses and without permission takes, copies, or makes use of any data from a computer, computer system, or computer network, or takes or copies any supporting documentation, whether existing or residing internal or external to a computer, computer system, or computer network."). Notably, the Ninth Circuit held in *Power Ventures* that Power Ventures had also violated the California Penal Code provision above when it scraped Facebook's data after receiving the cease-and-desist letter. *Power Ventures*, 844 F.3d at 1062.