Proskauer

Trove of Online LinkedIn User Data Fuels LinkedIn's Anti-Scraping Position

New Media and Technology Law Blog on **April 13, 2021**

Last week, the Italian data protection authority (the "GPDP") opened an investigation after reports that a dataset allegedly containing data compiled from 500 million LinkedIn profiles and other websites was available for sale on a hacker forum. Apparently, this data represents more than two-thirds of LinkedIn's estimated 740 million users. The hacker reportedly posted approximately two million records visibly online as evidence of the dataset, and offered to sell the rest for an undisclosed bitcoin payment.

According to a statement by LinkedIn, the company investigated the posting and determined that it is "an aggregation of data from a number of websites and companies," including publicly viewable LinkedIn member profile data that apparently was scraped from LinkedIn's site. LinkedIn stated that it was not a data breach because no private member profile data was included in the dataset it was able to review. LinkedIn stated that such scraping of data violated its terms.

The posting of this scraped data immediately reminds us of the ongoing scraping dispute between LinkedIn and data analytics start-up hiQ, Inc. ("hiQ"). The principal issue in the case concerns the scope of Computer Fraud and Abuse Act (CFAA) liability associated with web scraping of publicly available social media profile data. In a prior ruling, the Ninth Circuit affirmed the lower court's order granting a preliminary injunction barring LinkedIn from blocking hiQ from accessing and scraping publicly available LinkedIn member profiles.

The current incident and the hiQ-LinkedIn dispute are distinguishable in at least one major way – one involves an unknown party that scraped and posted for sale a massive trove of public LinkedIn profile data and the other involves a start-up company that scraped specific public LinkedIn profile data germane to its clients and crunched it internally for data analytic products about employee mobility and other tendencies of LinkedIn users. Still, both evoke similar big picture issues about potential data privacy and other issues associated with the scraping of public social media data. Throughout the litigation and most recently in its Counterclaims filed against hiQ in the case, LinkedIn has outlined the data privacy considerations and the potential loss of user trust it would suffer if scraping of public user profiles, which violates its site's terms, is freely permitted:

> "[O]nce a scraper has scraped a member's data, the members have no recourse to stop the scraper from copying, archiving, and forever keeping any information.... The member cannot stop the scraper from, among other things, using that data to target spam, selling or inadvertently exposing that data to scammers.... In short, once data has been scraped, member data can end up in any number of databases controlled and used for any purpose."

hiQ has countered this narrative by suggesting, in its Motion to Strike, that LinkedIn is merely attempting to "pose as a defender of user privacy by characterizing hiQ as a 'scraper'" in a veiled attempt to "shut down fair competition" for creating data analytic products from its users' data and that LinkedIn has demonstrated no user harm from hiQ's activities regarding data that users have already made public.

Regardless of which side of the foregoing argument one takes, the current scraping attack calls to mind the data privacy concerns LinkedIn has expressed about indiscriminate scraping and may give LinkedIn added momentum in the pending dispute.

This current development also evokes some of the issues that are being debated surrounding the practices of Clearview AI, Inc., an entity that scraped billions of photographs from public websites without consent and created a commercial service that allowed certain entities to upload a photograph to instantly identity the person depicted via its facial recognition matching technology. Even though the images scraped by Clearview AI had been publicly posted by users, the State of Vermont, a civil liberties organization and a class of users, among others, have brought various privacy and consumer protection-related suits against Clearview AI over its data collection and usage.

Given the unsettled area of law surrounding data scraping, there are no clear answers in this area yet.

View Original

Related Professionals

- **Jeffrey D. Neuburger**
  Partner