

<https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm>

How Cambridge Analytica turned Facebook 'likes' into a lucrative political tool

The algorithm used in the Facebook data breach trawled through personal data for information on sexual orientation, race, gender - and even intelligence and childhood trauma

Carole Cadwalladr and Emma Graham-Harrison

Sat 17 Mar 2018 13.02 GMT

The algorithm at the heart of the Facebook data breach sounds almost too dystopian to be real. It trawls through the most apparently trivial, throwaway postings - the "likes" users dole out as they browse the site - to gather sensitive personal information about sexual orientation, race, gender, even intelligence and childhood trauma.

A few dozen "likes" can give a strong prediction of which party a user will vote for, reveal their gender and whether their partner is likely to be a man or woman, provide powerful clues about whether their parents stayed together throughout their childhood and predict their vulnerability to substance abuse. And it can do all this without an need for delving into personal messages, posts, status updates, photos or all the other information Facebook holds.

Some results may sound more like the result of updated online sleuthing than sophisticated data analysis; "liking" a political campaign page is little different from pinning a poster in a window.

But five years ago psychology researchers showed that far more complex traits could be deduced from patterns invisible to a human observer scanning through profiles. Just a few apparently random "likes" could form the basis for disturbingly complex character assessments.

When users liked "curly fries" and Sephora cosmetics, this was said to give clues to intelligence; Hello Kitty likes indicated political views; "Being confused after waking up from naps" was linked to sexuality.

These were just some of the unexpected but consistent correlations noted in a paper in the Proceedings of the National Academy of Sciences journal in 2013. "Few users were associated with 'likes' explicitly revealing their attributes. For example, less than 5% of users labelled as gay were connected with explicitly gay groups, such as No H8 Campaign," the peer-reviewed research found.

The researchers, Michal Kosinski, David Stillwell and Thore Graepel, saw the dystopian potential of the study and raised privacy concerns. At the time Facebook "likes" were public by default.

"The predictability of individual attributes from digital records of behaviour may have considerable negative implications, because it can easily be applied to large numbers of people without their individual consent and without them noticing," they said.

"Commercial companies, governmental institutions, or even your Facebook friends could use software to infer attributes such as intelligence, sexual orientation or political views that an individual may not have intended to share."

To some, that may have sounded like a business opportunity. By early 2014, Cambridge Analytica chief executive Alexander Nix had signed a deal with one of Kosinski's Cambridge colleagues, lecturer

Aleksandr Kogan, for a private commercial venture, separate from Kogan's duties at the university, but echoing Kosinski's work.

The academic had developed a Facebook app which featured a personality quiz, and Cambridge Analytica paid for people to take it, advertising on platforms such as Amazon's Mechanical Turk.

The app recorded the results of each quiz, collected data from the taker's Facebook account - and, crucially, extracted the data of their Facebook friends as well.

The results were paired with each quiz-taker's Facebook data to seek out patterns and build an algorithm to predict results for other Facebook users. Their friends' profiles provided a testing ground for the formula and, more crucially, a resource that would make the algorithm politically valuable.



Aleksandr Kogan

To be eligible to take the test the user had to have a Facebook account and be a US voter, so tens of millions of the profiles could be matched to electoral rolls. From an initial trial of 1,000 "seeders", the researchers obtained 160,000 profiles - or about 160 per person. Eventually a few hundred thousand paid test-takers would be the key to data from a vast swath of US voters.

It was extremely attractive. It could also be deemed illicit, primarily because Kogan did not have permission to collect or use data for commercial purposes. His permission from Facebook to harvest profiles in large quantities was specifically restricted to academic use.

And although the company at the time allowed apps to collect friend data, it was only for use in the context of Facebook itself, to encourage interaction. Selling that data on, or putting it to other purposes, - including Cambridge Analytica's political marketing - was strictly barred.

It also appears likely the project was breaking British data protection laws, which ban sale or use of personal data without consent. That includes cases where consent is given for one purpose but data is used for another.

The paid test-takers signed up to T&Cs, including collection of their own data, and Facebook's default terms allowed their friends' data to be collected by an app, unless they had changed their privacy settings. But none of them agreed to their data possibly being used to create a political marketing tool or to it being placed in a vast campaign database.

Kogan maintains everything he did was legal and says he had a "close working relationship" with Facebook, which had granted him permission for his apps.

Facebook denies this was a data breach. Vice-president Paul Grewal said: "Protecting people's information is at the heart of everything we do, and we require the same from people who operate apps on Facebook. If these reports are true, it's a serious abuse of our rules."

The scale of the data collection Cambridge Analytica paid for was so large it triggered an automatic

shutdown of the app's ability to harvest profiles. But Kogan told a colleague he "spoke with an engineer" to get the restriction lifted and, within a day or two, work resumed.

Within months, Kogan and Cambridge Analytica had a database of millions of US voters that had its own algorithm to scan them, identifying likely political persuasions and personality traits. They could then decide who to target and craft their messages that was likely to appeal to them for those individuals - a political approach known as "micro-targeting".

Facebook announced on Friday that it was suspending Cambridge Analytica and Kogan from the platform pending information over misuse of data related to this project.

Facebook denies that the harvesting of tens of millions of profiles by GSR and Cambridge Analytica was a data breach.

It said in a statement that Kogan "gained access to this information in a legitimate way and through the proper channels", but "did not subsequently abide by our rules" because he passed the information onto third parties.

