The aftermath of the case of Cambridge Analytica

Dimitris C. Zoumpakis [redacted]

Fundamentals of Information Systems CIS100

Dr. Katerina Papanikolaou

Department of Sciences, B.Sc. Computer Information Systems, European University Cyprus

7/11/2023

Table of Contents

In the age of web3.0 and the fourth industrial revolution where information has more value than gold and even oil, many questions arise. One such query of critical importance which comes with the rise of OSNs (Online Social Networks) an the subsequent rise of the individuals digital footprint has troubled everyone from data privacy and cybersecurity specialists to sociologists, constantly making the headlines on major news channels. That is the question of whether the individuals personal and private data can be trusted and kept safe and secure in the hands of giants such as Google and Facebook. The same question was asked in a state of panic when in March of 2018 it was made public that a leak coming from Facebook and Cambridge Analytica had left the personal information of millions of people free for grab. This paper will look at what really happened, the consequences and the aftermath of the Facebook – Cambridge Analytica data harvesting, starting from the original research, how the data was harvested, for what reasons and what led to the leak as well as the results it brought.

For starters let's see how it all started, to finally lead to the data harvesting and later data leak of the private information of more than 80 million people from Cambridge Analytica in the first place. The story begins in 2007 where a researcher at the university of Cambridge who goes by the name of David Stillwell developed an application called "myPersonality". This application encouraged Facebook users to take psychographic tests (the study of people's attitudes and interests) up until 2012 (Rehman, 2019). The goal of this academic research was to test if it was possible to find important information on individuals based on a questionnaire the users were asked to fill and then after compared with what they liked, posted, shared, what age, gender, and place of residence they mentioned on Facebook (Rehman, 2019). They would assess this theory through five personality traits known as the Big Five or OCEAN (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism). During those years the data of more than 6 million volunteers was collected and shared with academic collaborators resulting in numerous scientific publications (Kanter and Kanter, 2018). In 2008 Michael Konsinski joined Stillwell and assessed the data collected on the users from the application. In 2012 Konsinski and his team proved using OCEAN that it was possible to predict important information about a user such as their intelligence, religious affiliations, or their association with a political party, based merely on just an average of 68 Facebook "likes" (Grassegger & Krogerus, 2017).

The result of this research brings us to the second part of the story and perhaps the most important, the work of Cambridge Analytica. CA was a British based

political consulting firm which combined the predictive data analysis, behavioral sciences, and innovative ad tech into one approach (Rathi,2019). It was founded in 2013 as a subsidiary of Strategic Communication Limited (SGL) a company known for its work on the Leave EU and Brexit movements. CA's goal was to "persuade users to vote a certain way by showing different advertisements on the same issue to different people" (Rathi, 2019). The key players of Cambridge Analytica are as follows, firstly Nigel Oakes longtime SLC executive and founder of CA, Robert Mercer partial owner of CA, Alexander Nix who was a co-founder and the CEO, Steve Bannon VP of Cambridge Analytica, Christopher Wylie lead data scientist, Brittany Kaiser the director of business development and last but not least Alexander Kogan a research scientist at the university of Cambridge and owner of Global Science Research (Boerboom, 2020). The initial idea of combining micro-targeting (target audience restricted to only a handful of individuals with a very specific element in common) then targeting individuals not only as voters but also as a personality, leading to the creation of a psychological profile of each voter in a particular region/state, was born out of a conversation between Christopher Wylie and Steve Bannon. They immediately realized that to make their idea work they needed huge amounts of data, so Wylie decided to first approach Konsinski, co-author of the "myPersonality" application and research to gain access to the database of the app. Negotiations would fail as Konsinski would refuse to give access, but sooner than later another researcher and a colleague of Konsinski who goes by the name of Alexandr Kogan offered to replicate the original research (Cadwalladr & Graham-Harrison 2018). Kogan would go on to develop an application called "thisisyourdigitallife" which such as the original research ("myPersonality") featured a personality quiz in the form of a questionnaire with 120 question (Detrow, 2018). Then CA in partnership as mentioned by Rehman with Global Science Research (GSR), a company owned by Kogan paid approximately two hundred seventy thousand Facebook users to take the personality test (Wagner, 2021). For the users to be able to take the test they had to not only be FB users, but also be US voters in order for each profile to be matched to electoral rolls and agree to have their data collected for academic purposes (Rehman, 2019). The application developed by Kogan not only collected the results of the people who took the quiz but also their Facebook user ID which includes information such as gender, birthdate, location, likes and interests and most importantly their list of friends and their user ID, this was made possible through a loophole in Facebook's API, which is what ultimately caused the data breach affecting over 87 million FB users (Wagner, 2021). Nevertheless, the users in the first place never agreed to their information being sold to data firms such as CA and

turned into a lucrative political tool (Cadwalladr & Graham-Harrison, 2018). In an initial trial of one thousand test takers, GSR was able to collect one hundred sixty thousand profiles, about one hundred sixty per person (Rehman, 2019). Within a matter of months, they had a database of about two hundred seventy million American adult consumers according to Amer and Noujaim (“The Great Hack”). As mentioned by Dave Smith in his research, after the data was given to CA, they used LASSO regression and SVD + Linear/Logistic regression to reach their personality predicting algorithms, which are very effective tools that give a way to eliminate data redundancy, focus on the information with the maximum predictive influence and improved computational efficiency. The target of the regressions was “Facebook pages with the highest predictive influences on someone’s level of a specific “big five” personality traits (Smith, 2018). This brings us to the real debate, which has sparked the most contradictions amongst researchers and professionals of the industry and it’s none other than if Cambridge Analytica used the data to assist in the US Presidential elections of 2016 and the election of Donald Trump as president. Sherr (2018), believes that Cambridge Analytica played a key role in Trump’s win since the Trump campaign hired CA in 2016 to firstly run data operations, secondly identify voters to target with ads, thirdly where to make campaign stops, and fourth and final to help with strategic communications. CA then focused on the thousands of users they knew through their previous analysis were hesitant to vote and managed to define thirty-two types of personalities targeting those believed to be worried and neurotic which made them vulnerable to Trump’s message. They identified most of such individuals in Michigan, Wisconsin, and Pennsylvania, targeting them with a not very well-known FB feature called “Dark Post” (Personalized ads on social media visible only by the person specifically targeted). The ad would be sent to the users Facebook feed at a specific time to only be seen by them, disappearing in a few hours, all this based on the habits and digital fingerprints of each user (Rehman, 2019). This as said by Camelot (2018) helped Trump gain 77000 votes in those three states. On the other hand, based on INSTAR (2018) the data collected by CA was supposedly deleted before they helped on D. Trumps campaign. The accuracy of Cambridge Analytica’s regression models were also analyzed by David Sumpter, using a publicly available dataset, and found that although it worked very well for hardcore democrats or republicans it didn’t reveal anything for the 76% of users who did not put their political allegiance on Facebook (Sumpter, 2018). Another key point on the matter of CA assisting Trump in his presidential election is that Cambridge Analytica’s initial goal was to have Ted Cruz win, represent the Republican party and be elected as the US President, always according to Davies (2018). Nevertheless, it is

difficult to know whether or not the data that was collected and the techniques that were used, actually helped Trump win the election.

Moving on, it is pivotal to understand how Kogan and Cambridge Analytica were able and even allowed to harvest the private information of millions of users in the first place. It is also very important to understand the CA is not the only company at fault for the scandal. That's where Facebook and their, at the time, carelessness come into play since they are the platform that allowed it to even possible. The main reasons were insufficient safeguards against data mining firms, inadequate supervision of developers by FB and because of users excessively agreeing to terms and conditions (Rehman, 2019). More specifically, Facebook "failed to read all of GSR's applications terms and conditions during their app review process" (Smith, 2018). But also, from GSR's side, Kogan stated that he did not read Facebook's App Developer policy, which clearly stated that it's prohibited to share data with another party. Facebook had once again breached data protection laws by failing to keep user's personal information secure and by allowing CA to harvest such data without their consent (Wagner, 2021). In 2015 Facebook would find out about the data gathering from Kogan's app and would demand that all data gathered illegally was to be deleted but also banned the application "thisisyourdigitallife" from its platform. Since Kogan had sold the information to CA, both its collection and transfer were presumably unlawful and violated Facebook's terms of agreement but also the security safeguards principle set by the OECD (Organization for Economic Co-operation and Developments) (Wagner, 2021).

After we understand what really occurred, then we can start painting a picture of what led to the whistle blowers waving the flag and "switching" sides but most importantly the response of the involved parties. Firstly, a mention of who brought the leak to the light is necessary. Christopher Wylie who was the initial whistleblower, and Brittany Kaiser, with the latter's reputation being questionable since it has been said by Wakefield (2019) that her whistleblowing was done more to save herself rather than to expose the company she worked for and their unethical and illegal wrongdoings. It wouldn't be fair to not, at least, mention the name of the reporters that helped uncover the whole scandal, these are Paul-Olivier Dehaye and Carole Cadwalladr. In March of 2018 FB would once again find itself into deep waters since they were informed by Christopher Wylie that the data that had been collected by CA and Kogan's application were not deleted, even though both Kogan and Cambridge Analytica had assured and certified FB that it had been done. In the 17 of March of 2018 Facebook would make their first respond on the issue, in a Facebook

post by Vice-President and Deputy General Counsel Paul Grewal who wrote “The claim that this is a data breach is completely false. Kogan requested and gained access to information from users who chose to sign up and gave their consent, no systems were infiltrated, no passwords or sensitive pieces of information were stolen or hacked” (Grewal, 2018). On the same day the CSO (Chief Security Officer), tweeted that “Kogan did not break into any systems, bypass any technical controls, or use a flaw in our software. He misused the data after he gathered it” (Johnston, 2018). Both statements are correct since no systems were infiltrated and no passwords or sensitive information were stolen (Wong, 2019). So, by Information Security standards this was not a data breach but a privacy harm which are different, since the latter involves data that is collected or used in ways that consumers didn’t consent to or knew about (Wagner, 2021). Only two hundred seventy thousand individuals downloaded the app and gave their consent, the rest eighty-seven million were neither notified nor gave their consent. Facebook would also violate the security safeguard principle set by OECD, as they would fail to implement sufficient security safeguards to prevent app developers from harvesting user data from friends or other users as stated by Wagner.

Even after all that, what really changed. For starters the Federal Trade Commission, in an attempt to make users confident with their rights against Facebook, issued a five billion US dollars fine for as stated by the FTC “deceiving users about their ability to control the privacy of their personal information”, a data security/privacy penalty over twenty times more than any other company has ever been imposed globally, and even though FB’s revenue in 2018 was fifty-five point eight billion US dollars, it surely did send a message to the rest of the OSN’s. They also demanded a new privacy compliance system which included two-factor authentication and tools that would help FTC (Federal Trade Commission) monitor FB (Wakefield, 2019). Additionally, FB would lose billions in share value. Cambridge Analytica filed for bankruptcy in 2018, with the concern of if they ever deleted the data, they had bought from Kogan and GSR still unclear and questioned.

In conclusion, the academic research turned into a cyber-weapon by Cambridge Analytica and Kogan’s GSR which left the private information of more than eighty millions of users available to all through illegal and unethical techniques, in order to assist in elections and political advertising and had as a result the biggest fine in history against the leading OSN Facebook, made organizations and most importantly governments and lawmakers rethink and reshape outdated laws and develop strong regulations and privacy principles in order to

stop the current complexities of data usage/transfer and increase data privacy. Such as the General Data Protection Regulation by the European Union, the OECD, the CCPA (California Consumer Privacy Act) which are only examples of the change happening. Nonetheless the harm has also made the users rethink if they should post or even be blamed for the information they put online, since as consumers they all must understand the risks and consequences of the digital world. Finally the Cambridge Analytica-Facebook scandal has raised the question of even if you can collect the data, should you?

References:

- Rehman, I 2019. “Facebook-Cambridge Analytica data harvesting: What you need to know”. Library Philosophy and Practice
- Cadwalladr, C., & Graham-Harrison, E. 2018. “The Cambridge Analytica Files. *I made Steve Bannon’s psychological warfare tool: meet the data was whistleblower”.* The Guardian.
- Grewal, P. 2018 (March 17). “Suspending Cambridge Analytica and SLC Group from Facebook”. Newsroom.fb.com
- Johnston, C. 2018 (March 19). “Cambridge Analytica’s leak shouldn’t surprise you, but it should scare you”. The Outline
- Sherr, I. 2018 (April 19). “Number of social media users worldwide 2010-2021”. Cnet
- Wagner, P. 2021. “Data Privacy: The Ethical, Sociological, and Philosophical Effects of Cambridge Analytica”
- Boerboom, C. 2020. “Cambridge Analytica: The Scandal on Data Privacy”. Augustana College, Rock Island Illinois
- Davies, A. 2018 (March 17). “Whistleblower Reveal to Channel 4 News Data Grab of 50 million Facebook Profiles by Cambridge Analytica – Data Firm Linked to Trump Win”. Channel 4 news
- Detrow, S. 2018 (March 20). “What did Cambridge Analytica Do During the 2016 Election?” NPR
- FTC 24 July 2019. “FTC sues Cambridge Analytica, Settles with former CEO and App Developer.” *Federal Trade Commission*

- INSTAR 2018 “Robert Mercer, Cambridge Analytica, and the Election of Donald Trump” *Medium*
- Rathi, R. 2019. “Effect of Cambridge Analytica’s Facebook Ads on the 2016 US Presidential Election” *Medium*
- Smith, D. 2018. “Weapons of Micro Destruction: How our “Likes” Hijacked Democracy” *Medium*
- Sumpter, David J. T. 2018. “Outnumbered: from Facebook and Google to Fake News and Filter-Bubbles – the Algorithms that Control our Lives.” Bloomsbury Sigma, chapter 5
- Amer, Karim and Jehane Noujaim, directors. 24 July 2019 *The Great Hack,* Netflix
- Wakefield, J. 2019. “Brittany Kaiser Calls for Facebook Political Ad Ban at Web Summit” BBC News