Carter Brainerd

Professor McCarthy

ENGW1111

29 September 2019

Privacy in an Increasingly Connected Age: What Happens When There's a Leak?

It seems like every other week there's a report about a major data breach on the news[1]. In today's modern age, people are moving more and more of their lives online[2]. Because there are more people online, there is more personal data these companies are storing and therefore should be keeping secure. However, what happens when internet companies' data protection is lacking? Moreover, what happens if malicious actors or foreign governments get their hands on millions of people's personal data? Using past data breaches like Facebook, Equifax, and Yahoo as examples, the results are clear: if millions of people's leaked personal data is used in AI or political campaigns, people's lives could be at risk of drastic harmful change. Because of the enormous risk of data breaches, governments around the world should increase the severity of punishments for inadequate data security.

Before data breaches are discussed in detail, we first must first define what data breaches are and how they happen. In short, a data breach is a security incident in which information is accessed, and in some cases made public, without authorization[3]. The most common ways attackers get access to this data without authorization are exploiting system vulnerabilities, weak

---

[1] Weisbaum, Herb. "Data Breaches Happening at Record Pace, Report Finds." NBCNews.com. NBCUniversal News Group, July 24, 2017. https://www.nbcnews.com/business/consumer/data-breaches-happening-record-pace-report-finds-n785881.

[2] "Demographics of Social Media Users and Adoption in the United States," Pew Research Center: Internet, Science & Tech, June 12, 2019, https://www.pewinternet.org/fact-sheet/social-media/.

[3] NortonOnline, "What Is a Data Breach?," Official Site, accessed September 19, 2019, https://us.norton.com/internetsecurity-privacy-data-breaches-what-you-need-to-know.html.

passwords, and targeted malware attacks[4]. There are some measures individuals can take to further protect their personal data like using more secure passwords and using a VPN, but when a company's security is lacking, there's not much an individual person can do. Attackers are clever though, there's always a cat and mouse game between companies and malicious actors and the companies are always trying to stay one step ahead of attackers, even though its usually the other way around.

Ever since there has been data people want to keep private, there has been people that want to get it and get it without anyone knowing. This has, unfortunately, been happening at a record pace[5] and in larger numbers. In fact, just in the last decade, there have been 1,624,000,000 people affected by just four data breaches: five hundred million from Yahoo in 2016 and 2014[6], one hundred and forty seven million from Equifax in 2017[7], eighty seven million from Facebook/Cambridge Analytica in 2016[8], and a whopping eight hundred and ninety million from First American Financial Corporation in 2019[9]. However, the total number of affected people does not take into account duplicate entries, so the number is likely smaller. For an estimate of how many total accounts have been breached, the website haveibeenpwned.com keeps a record

---

[4] Ibid
[5] Weisbaum, Herb. "Data Breaches Happening at Record Pace, Report Finds." NBCNews.com. NBCUniversal News Group, July 24, 2017.
[6] Joseph Cox, "Yahoo 'Aware' Hacker Is Advertising 200 Million Supposed Accounts on Dark Web," Vice, August 1, 2016, https://www.vice.com/en_us/article/aeknw5/yahoo-supposed-data-breach-200-million-credentials-dark-web.
[7] "Equifax Data Breach Settlement," Federal Trade Commission, September 9, 2019, https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement.
[8] Paul Chadwick, "How Many People Had Their Data Harvested by Cambridge Analytica? | Paul Chadwick," The Guardian (Guardian News and Media, April 16, 2018), https://www.theguardian.com/commentisfree/2018/apr/16/how-many-people-data-cambridge-analytica-facebook.
[9] Nicole Perlroth and Stacy Cowley, "Security Gap Leaves 885 Million Mortgage Documents Exposed," The New York Times (The New York Times, May 24, 2019), https://www.nytimes.com/2019/05/24/technology/data-leak-first-american.html.

of 405 major major data breaches and has collected over eight billion accounts* from data breaches[10]. All of these breaches occur through the aforementioned methods: exploiting system vulnerabilities, weak passwords, and targeted malware attacks. Most cybersecurity defense today is reactive. Whenever a new threat emerges, the companies usually try their best to fix the vulnerability as soon as possible. To beat attackers to the punch, companies have started hiring private security researchers, or "white hat hackers" to find and report vulnerabilities before the bad guys do.

With big internet companies storing petabytes of data on its users, leaking it not only degrades the view of the company in the public eye, but it also can be used to influence or harm those affected. One prime example of this shady use of data is from the political consulting group Cambridge Analytica. Cambridge Analytica combined data mining, data brokerage, and data analysis with strategic communication in political elections[11]. They have had a major influence of elections in India in 2019, Kenya in 2013 and 2017, Malta in 2018, Brexit, and, most notably, the United States presidential election in 2016[12]. It was revealed that in the years preceding the 2016 US elections, Cambridge Analytica had harvested the data of 87 million[13] Facebook users without its consent and with Facebook's knowledge and permission[14].

---

[10] Troy Hunt, Have I Been Pwned?, n.d., https://haveibeenpwned.com/.
[11] David Ingram, "Factbox: Who Is Cambridge Analytica and What Did It Do?," Reuters (Thomson Reuters, March 20, 2018), https://www.reuters.com/article/us-facebook-cambridge-analytica-factbox/factbox-who-is-cambridge-analytica-and-what-did-it-do-idUSKBN1GW07F.
[12] Devjyot Ghoshal, "Mapped: The Breathtaking Global Reach of Cambridge Analytica's Parent Company," Quartz (Quartz, April 2, 2018), https://qz.com/1239762/cambridge-analytica-scandal-all-the-countries-where-scl-elections-claims-to-have-worked/.
[13] "An Update on Our Plans to Restrict Data Access on Facebook," Facebook Newsroom, April 4, 2018, https://newsroom.fb.com/news/2018/04/restricting-data-access/.
[14] Matthew Rosenberg, Nicholas Confessore, and Carole Cadwalladr, "How Trump Consultants Exploited the Facebook Data of Millions," The New York Times (The New York Times, March 17, 2018), https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.
* "Accounts" meaning a single online account. This is not the number of people affected. There are many instances of multiple accounts being tied to one person.

Cambridge Analytica was hired by Donald Trump's campaign in 2016 to help him win the general presidential election[15] and were ultimately successful (although there are doubts about how much influence Cambridge Analytica actually had[16]). Cambridge mainly developed personalized online advertising campaigns influence American voters who were deemed "influenceables"[17]. They determined who fell into this category through a machine learning model called "Singular Value Decomposition". This is roughly the same machine learning model Netflix uses to recommend movies. But instead of recommending movies, Cambridge Analytica created targeted ads to influence American voters to vote for Trump. Cambridge Analytica figured out who was influenceable by using the data improperly collected from Facebook to perform psychometric grouping on voters[18]. It was discovered later that Cambridge Analytica gave this data to Russia to further help the Trump campaign[19]. Most laws concerning protection and handling of digital data are many years behind our data collection ability. Governments around the world, but the US in particular, must modernize their laws and force them to be revised every few years in order to keep giant internet companies in check.

There are a plethora of historical examples of what happens when peoples personal data gets into malicious hands, notably the Facebook/Cambridge Analytica scandal. Its ramifications

---

[15] Alex Altman, "Silent Partners", Time, (Time, October 10, 2016)
[16] Kris-Stella Trump, "Four and a Half Reasons Not to Worry That Cambridge Analytica Skewed the 2016 Election," The Washington Post (WP Company, March 23, 2018), https://www.washingtonpost.com/news/monkey-cage/wp/2018/03/23/four-and-a-half-reasons-not-to-worry-that-cambridge-analytica-skewed-the-2016-election/.
[17] The Great Hack, Netflix (Netflix, n.d.), https://www.netflix.com/title/80117542.

[18] Matthew Hindman, "How Cambridge Analytica's Facebook Targeting Model Really Worked – According to the Person Who Built It," The Conversation, August 27, 2019, http://theconversation.com/how-cambridge-analyticas-facebook-targeting-model-really-worked-according-to-the-person-who-built-it-94078.
[19] Massimo Calabresi, "Russia's US Social Media Hacking: Inside the Information War," Time (Time), accessed September 20, 2019, https://time.com/4783932/inside-russia-social-media-war-america/.

help shape the current political stage of the United States. That, however, are the results of only one data breach (and not a huge one at that). Since Cambridge Analytica used the data effectively, they had effective results. If an intelegent AI was trained with the masses of breached data, however, what would the consequences be? The methods Cambridge Analytica used to identify who was influenceable is, in the grand scheme of things, a pretty simple machine learning algorithm. More powerful ones and complex ones like OpenAI's GPT-2, however, have people scared, even its creators[20].

GPT-2 is a "large transformer-based language model with 1.5 billion parameters."[21] Its goal is a very simple one: predict the next word in 40 GB of text. However, the AI wrote so well and so much like a human[22] that when the company that created it, OpenAI, released it to the world they were concerned about malicious use so they did not release a fully trained model[23].

If malicious actors were persistent enough and had the willpower to do so, they could collect peoples personal data from multiple breaches and train GPT-2 to generate fake news articles that seem very real but are used to influence the public to follow some agenda. While this speculation of malicious AI influencing humans via fake articles is all somewhat speculative according to its creators, it is nonetheless scary, and a very real possibility.[24] However, in the present day, data breaches still pose a threat to the public even without AI usage. Use of leaked personal data by malicious actors can still cause reputational damage through leaked private and

---

[20] Alec Radford, "Better Language Models and Their Implications," OpenAI (OpenAI, July 3, 2019), https://openai.com/blog/better-language-models/.
[21] Ibid
[22] James Vincent, "OpenAI's New Multitalented AI Writes, Translates, and Slanders," The Verge (The Verge, February 14, 2019), https://www.theverge.com/2019/2/14/18224704/ai-machine-learning-language-models-read-write-openai-gpt2.
[23] Alec Radford. "Better Language Models and Their Implictions." OpenAI.
[24] Alex Hern, "New AI Fake Text Generator May Be Too Dangerous to Release, Say Creators," The Guardian (Guardian News and Media, February 14, 2019), https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction.

personal details, enable identity fraud through leaked social security numbers, and incite monetary theft through leaked bank credentials.[25] Each one of these consequences of data breaches is extremely harmful to individuals, but when used together, many peoples lives can be ruined. With the threat of harsher punishment from governments, companies will be forced to check every nook and cranny of their network to ensure nobody's lives get ruined from a data breach.

The amount of data companies store in individuals grows every day. But with this growth of their respective data hordes comes the need and the responsibility to keep the data safe and secure. While some companies are excellent at protecting their users' data and implement modern protection methods, other companies like Facebook, Equifax, and Yahoo have suffered very major data breaches in the past due to their poor data protection policies. The consequences of these data breaches have the potential to be catastrophic to individual people's lives and the companies responsible are not being held accountable enough. In order to force companies to take action in securing their data, governments around the world should increase fines for when data breaches happen.

[25] Annabelle Graham, "The Damaging after-Effects of a Data Breach," IT Governance Blog, May 29, 2019, https://www.itgovernance.co.uk/blog/the-damaging-after-effects-of-a-data-breach.

Works Cited

1. Weisbaum, Herb. “Data Breaches Happening at Record Pace, Report Finds.” NBCNews.com. NBCUniversal News Group, July 24, 2017. https://www.nbcnews.com/business/consumer/data-breaches-happening-record-pace-report-finds-n785881.
2. “Demographics of Social Media Users and Adoption in the United States.” Pew Research Center: Internet, Science & Tech, June 12, 2019. https://www.pewinternet.org/fact-sheet/social-media/.
3. NortonOnline. “What Is a Data Breach?” Official Site. Accessed September 19, 2019. https://us.norton.com/internetsecurity-privacy-data-breaches-what-you-need-to-know.html.
4. Ibid
5. Weisbaum “Data Breaches Happening at Record Pace, Report Finds.”
6. Cox, Joseph. “Yahoo 'Aware' Hacker Is Advertising 200 Million Supposed Accounts on Dark Web.” Vice, August 1, 2016. https://www.vice.com/en_us/article/aeknw5/yahoo-supposed-data-breach-200-million-credentials-dark-web.
7. “Equifax Data Breach Settlement.” Federal Trade Commission, September 9, 2019. https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement.
8. Chadwick, Paul. “How Many People Had Their Data Harvested by Cambridge Analytica? | Paul Chadwick.” The Guardian. Guardian News and Media, April 16, 2018. https://www.theguardian.com/commentisfree/2018/apr/16/how-many-people-data-cambridge-analytica-facebook.
9. Perlroth, Nicole, and Stacy Cowley. “Security Gap Leaves 885 Million Mortgage Documents Exposed.” The New York Times. The New York Times, May 24, 2019. https://www.nytimes.com/2019/05/24/technology/data-leak-first-american.html.
10. Hunt, Troy. Have I Been Pwned?, n.d. https://haveibeenpwned.com/.
11. Ingram, David. “Factbox: Who Is Cambridge Analytica and What Did It Do?” Reuters. Thomson Reuters, March 20, 2018. https://www.reuters.com/article/us-facebook-cambridge-analytica-factbox/factbox-who-is-cambridge-analytica-and-what-did-it-do-idUSKBN1GW07F.
12. Ghoshal, Devjyot. “Mapped: The Breathtaking Global Reach of Cambridge Analytica's Parent Company.” Quartz. Quartz, April 2, 2018. https://qz.com/1239762/cambridge-analytica-scandal-all-the-countries-where-scl-elections-claims-to-have-worked/.
13. “An Update on Our Plans to Restrict Data Access on Facebook.” Facebook Newsroom, April 4, 2018. https://newsroom.fb.com/news/2018/04/restricting-data-access/.
14. Rosenberg, Matthew, Nicholas Confessore, and Carole Cadwalladr. “How Trump Consultants Exploited the Facebook Data of Millions.” The New York Times. The New York Times, March 17, 2018. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.
15. Altman, Alex, “Silent Partners”, Time. Time, October 10, 2016.

16. Trump, Kris-Stella. “Four and a Half Reasons Not to Worry That Cambridge Analytica Skewed the 2016 Election.” The Washington Post. WP Company, March 23, 2018. https://www.washingtonpost.com/news/monkey-cage/wp/2018/03/23/four-and-a-half-reasons-not-to-worry-that-cambridge-analytica-skewed-the-2016-election/.
17. The Great Hack, Netflix (Netflix, n.d.), https://www.netflix.com/title/80117542.
18. Hindman, Matthew. “How Cambridge Analytica's Facebook Targeting Model Really Worked – According to the Person Who Built It.” The Conversation, August 27, 2019. http://theconversation.com/how-cambridge-analyticas-facebook-targeting-model-really-worked-according-to-the-person-who-built-it-94078.
19. Calabresi, Massimo. “Russia's US Social Media Hacking: Inside the Information War.” Time. Time. Accessed September 20, 2019. https://time.com/4783932/inside-russia-social-media-war-america/.
20. Radford, Alec. “Better Language Models and Their Implications.” OpenAI. OpenAI, July 3, 2019. https://openai.com/blog/better-language-models/.
21. Ibid
22. Vincent, James. “OpenAI's New Multitalented AI Writes, Translates, and Slanders.” The Verge. The Verge, February 14, 2019. https://www.theverge.com/2019/2/14/18224704/ai-machine-learning-language-models-read-write-openai-gpt2.
23. Radford, Alec. “Better Language Models and Their Implictions.” OpenAI.
24. Hern, Alex. “New AI Fake Text Generator May Be Too Dangerous to Release, Say Creators.” The Guardian. Guardian News and Media, February 14, 2019. https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction.
25. Graham, Annabelle. “The Damaging after-Effects of a Data Breach.” IT Governance Blog, May 29, 2019. https://www.itgovernance.co.uk/blog/the-damaging-after-effects-of-a-data-breach.