THE EVOLUTION OF FRAUD: ETHICAL IMPLICATIONS IN THE AGE OF LARGE-SCALE DATA BREACHES AND WIDESPREAD ARTIFICIAL INTELLIGENCE SOLUTIONS DEPLOYMENT

Abhishek Gupta
District 3, Concordia University, Montreal, Canada

Abstract – Artificial intelligence is being rapidly deployed in all contexts of our lives, often in subtle yet behavior-nudging ways. At the same time, the pace of development of new techniques and research advancements is only quickening as research and industry labs across the world leverage the emerging talent and interest of communities across the globe. With the inevitable digitization of our lives, increasingly sophisticated and ever larger data security breaches in the past few years, we are in an era where privacy and identity ownership is becoming a relic of the past. In this paper, we will explore how large-scale data breaches coupled with sophisticated deep learning techniques will create a new class of fraud mechanisms allowing perpetrators to deploy "Identity Theft 2.0".

Keywords – Artificial intelligence, data privacy, data security, ethics, fraud

1. INTRODUCTION

With the inevitable deployment of artificial intelligence (AI) at an unprecedented pace touching every conceivable aspect of our lives, it is of paramount importance that we begin considering some of the ethical implications of the development work and deployments being done in AI and how we can work towards mitigating adverse outcomes that might arise from the use of these systems.

This paper concerns itself with analyzing the long-run implications of large-scale data breaches and how, beyond the evident, immediate impacts, this will lead to the emergence of novel fraud techniques challenging societal balance. Specifically, the paper will focus on these negative outcomes being aggravated because of advances in deep learning (a subfield of AI).

Given how many diverse parts of our lives rely on products and services that use some sort of pattern recognition (one of the primary outcomes of applying AI techniques), the consequences of unlocking new hitherto anonymized datasets via the mosaic effect [1] means we will see the rapid rise of sophisticated fraud techniques that will span online, physical and biological characteristics of the target individual.

The idea of an individual's data exhaust [2] has gained traction in the security and privacy industry. We "emit" bits of personal data in numerous interactions, both online and offline. Government services [3], social media [4], grocery purchases [5], credit services [6], the list of places where we leave a digital footprint extends almost indefinitely. As the usage of mobile data increases and with the imminent arrival of 5G technologies, there is an increased risk of even more data flowing through these pipes. Perhaps this will also be supplemented by Internet of Things (IoT) devices, both in residential and industrial settings, which will collect even more information about consumer behavior and transmit them via protocols leveraging faster speeds and increased coverage due to the emergence of 5G technologies [7].

The pattern recognition outcomes as mentioned above are significant for two reasons; firstly, they discover existing correlations between different factors without being explicitly programmed to do so [8]. The second reason is more important because they can discover latent patterns to make decisions that are at once unknown to the developers of the system and opaque given current advances in being able to "explain" the reasoning behind the system taking particular decisions [9].

This paper will begin by explaining some of the key ideas that will be used recurrently to connect seemingly disparate developments in the field of deep learning, product development happening in different parts of the world and most importantly the data breaches that will enable the creation of novel

fraud techniques. Then we will go on to explore briefly some recent data breaches and their current impacts, as assessed by the cybersecurity community. We will tie all of this together to explain fraud scenarios that have the potential to arise as a consequence of piecing together the above and subsequently analyze how advances in deep learning techniques will turbocharge this process. Finally, we will look at some challenges that remain to be solved from an ethical, cybersecurity, regulatory and societal perspective. The paper makes preliminary recommendation to set the stage for future developments in this domain, especially as the pace of advances in AI quickens [10].

2. KEY IDEAS

2.1. Data brokers

Data brokers are companies that shy away from media exposure but are responsible for collecting a vast array of information [11] about individuals via different channels. Often they aggregate information from partners like social media companies, telecommunication firms, retail chains, e-commerce firms, etc. At the same time, these above-mentioned partners are also clients that look to augment the profiles that they have on their customers. Key concerns with the operations of the data brokers are that they collect in-depth information about consumers and their behaviors, often without their consent. The lack of consent is almost less shocking considering that most people are unaware [12] of the existence of these firms and the degree of profiling that they can do on a targeted individual.

In the context of this paper, the data brokers will be important because they often accumulate disparate datasets containing non-overlapping bits of information on an individual which if linked together can form an even richer representation of the individual for targeting purposes.

2.2 Mosaic effect

The mosaic effect results in a deeper analysis by combining several large datasets and making cross-references between them to coalesce previously de-linked information into a single, richer profile [1].

Edward Felten from Princeton University succinctly sums up the process of data collection and analysis today [1] as collecting, merging and analyzing data to infer facts about people. It is the merging process that gets a lot more effective as a consequence of the mosaic effect. During the first cycle of merging data, one is able to create more accurate portraits of users but follow-on cycles can do this in a more refined manner as they know more about the user behavior and identity which allows further merging.

This also has the added benefit of being able to generate smaller, more homogenous groups, i.e., increasing the ability to micro-target individuals while creating fine-grained analysis to inform policy-making (advertising, credit decisions, etc. by the acquiring firms) at a macro-level. On occasion, some of these insights arise without prior knowledge that such an objective could be achieved when data brokers are aggregating and analyzing this information.

Two examples illustrate the power of de-anonymization (the ability to identify an individual in an anonymized dataset) by combining different data points: Netflix movie ratings and AOL data search combination.

In the Netflix example [13], essentially after an anonymized dataset was released as part of a competition by Netflix to incentivize researchers to come up with more efficient recommendation systems, some researchers were able to cross-reference rating scores on rare movies with those on Internet Movie Database (IMDB) which has public ratings to identify the user and consequently reveal inferred information about the users, such as sexual orientation and potential political affiliations.

In the case of AOL [14], in an anonymized dataset that was released, a closer analysis of the actual search queries helped to localize the user to a specific region and then combining that with some easily accessible public information repositories, it became possible to identify users within that anonymized dataset.

Data brokers already leverage the mosaic effect to create rich consumer profiles that they can then sell to clients but following the large-scale data breaches, it will now be possible to "unlock" previously disjointed datasets via attribute groups that act as unique keys to link them together.

3. RECENT DATA BREACHES

This section will briefly look at the largest data breaches that have happened in recent history, looking at the specific bits of information leaked and a high-level view of potential consequences.

3.1. Yahoo

The most recent statement from Yahoo confirmed that all 3 billion accounts [15] with them were compromised, with information such as names, email addresses, dates of birth, encrypted passwords (for some fraction of the accounts the encryption was not state of the art) being leaked, as well as security questions and answers. What is troubling here is that, for example, weakly encrypted passwords and security question-answer pairs are usually similar for users across different web services, and this opens up the risk of users being compromised elsewhere.

Furthermore, this only helps to create more robust rainbow tables [16] which can be used as an easy mechanism to break through other online services that have less than ideal cybersecurity practices.

3.2. Adult Friend Finder

Approximately 412 million accounts were compromised in this leak which represented about 20 years' worth of information cutting across 6 databases [17]. The long temporal nature of this leak provides a unique window to the hacker community to find tidbits of information in terms of both old and new perspectives on the online identities of the users.

This also comes with the additional burden of revealing unsavory details of someone's online life with very tangible real-world consequences. This represents an almost 13 times larger breach compared to the Ashley Madison breach [18] which led to some suicides [19].

3.3. eBay

In this case, in 2014 all 145 million accounts [20] were compromised leading to the release of names, dates of birth, email addresses, physical addresses and salted-hashed passwords (though eBay declined to comment on the algorithm used to do so). What was of particular concern here is the leak of physical address information which adds another crucial link to the databases maintained by data brokers. Even though financial information was not leaked, being able to link the physical address, specifically the zip code gives away information of income ranges among other things via publicly available datasets from the national statistics databases.

3.4. Equifax

Equifax is one of the largest credit bureaus and the most recent large-scale compromise, more than 140 million accounts were hacked and birth dates, addresses, social security numbers (SSN) and in some cases driver license information was leaked. In about 200,000 cases consumers also had their credit card information exposed. Data brokers and other malevolent entities can pick up this information along with some of the ones mentioned before to create a pretty comprehensive profile of individuals who might have been present in all of the above-mentioned compromises. More importantly, the linking of SSN to other bits of personally identifiable information (PII) prepares unique keys that are invaluable in de-anonymizing large existing datasets.

4. EVOLUTION OF FRAUD

If we start to collect all the disparate pieces of information spread across the sample of large-scale compromises mentioned here, there are tremendous "unlocked" datasets that can now be fed into different pattern recognition systems to create the next generation of fraud techniques, i.e. Identity Theft 2.0.

Some of the recent research projects coming out of AI research labs across the world act as prescient reminders of a somewhat dystopic future if combined with the information leaks that we discussed above. As an example, Lyrebird [21] from Montreal, Canada uses "less than one minute" voice samples from an individual to generate voice output of them saying any statement that you input. Lyrebird is not perfect yet because it requires high-quality input samples with an array of sentences but as the technique improves, it would be possible to use voice samples over a telephone connection to collect enough data.

A research paper from the University of Washington [22] highlighted how they could take an audio sample and then synthesize video output of the person saying, with fairly realistic facial expressions and head movements, anything said in the audio. In its current iteration, this is limited in its application because of the need for high-quality video input for training purposes but again as techniques improve this could be applied more broadly.

For both of the above examples, it is already possible to synthesize fake audio and video output for public figures because it is easy to find high-quality training samples for them. In the case of synthesizing audio, anti-spoofing measures are not adequate. These synthesized samples can be used to impersonate users and trick authentication systems, potentially giving attackers access to unauthorized systems [23].

With the rising popularity and easy accessibility of genetic testing from firms like 23&Me [24], data compromises could be disastrous in terms of adding richer information to the open pool. [25] More so, this creates opportunities for targeted attacks on individuals, particularly those who are in the public domain. "Cloudhopping" is a technique where hackers compromise a low-hanging fruit on a particular cloud provider and are then able to "hop" onto the cloud infrastructure of a more valuable target. These techniques further increase the chances of information leaks, especially in targeted attack scenarios.

In a particularly far-reaching hack, Cloudflare leaked important tokens via a vulnerability dubbed "Cloudbleed" [26] which affected services from Uber, Fitbit and 1Password among others. These are all services that store highly personalized information about an individual which if combined with pieces of information from other data breaches mentioned in the previous section, via the mosaic effect, essentially enable the creation of a virtual avatar of an individual, replete with genetic and physical world information that can now be deployed to commit fraud using very complete real-world identities.

The most concerning thing here is that some of the techniques that are being used to create a synthetic identity and output for an individual can be automated; telephony systems can be connected via Twilio [27] to collect legitimate audio samples, video samples can be collected via camera feeds from all the places that an individual visits, their smartphone, their laptop, etc. and upon finding hits within the datasets sourced from data brokers, this can be combined to create a rich representation of an individual with sufficient data points to access even government services on their behalf.

5. OTHER ETHICAL CONSEQUENCES

There will be several long-run implications of these large-scale data breaches that are not immediately obvious.

Credit score ratings will be affected, negatively in some cases where disputed transactions and other historical data are tied to the identity of the user. Advertising firms will be able to further micro-target individuals based on these richer, "unlocked" datasets. Partnerships between large firms, public and private, can enable larger data sharing pools and access to each other's rapid AI advances [10].

Where this becomes especially problematic is in countries where one can be persecuted for particular political leanings, sexual orientation, group membership, etc. and this sort of access to data and advanced pattern recognition only enables that.

Even in progressive countries, hate groups can use this type of information to target individuals and engage in doxing behavior (publishing personal information of an individual into the public domain with malicious intent) that can inflict tremendous damage to the individual. [28] Easier access to AI tools via standard libraries and the falling learning curve to deploy basic techniques heralds a certain "democratization" of these methods.

6. AGGRAVATION BY AI ADVANCES

The emergence of Generative Adversarial Networks (GANs) as a way to produce synthetic data [29] poses an increased level of threat when combined with large-scale data breaches that provide a high amount of training data for these systems. Larger and more representative input into the system leads to more realistic synthetic data coming out of the GANs. They feed a cycle of being able to train AI systems into making more effective predictions and outputs. GANs can also be folded into the techniques mentioned in section 4 to further enhance the virtual avatar of an individual.

Cambridge Analytica [30] is known [31] to have been involved in creating advantages for candidates in senatorial races in the USA, the most recent US Presidential election, elections in the UK and South Africa. They use big data and advanced machine learning techniques to provide a full suite of services to enable highly targeted marketing and campaigning. They also have a commercial offering that is geared towards marketers helping them use "data to change user behavior".

The pace of development of deep learning is only hastening [10] and work is coming out of different research labs across the world (as an example, audio generation and video generation) that can potentially

be combined together to create more powerful tools that have unseen consequences. With strong incentives to monetize such advances, there will be an emergence of many novel use cases that have the potential to dramatically alter our notion of ownership of our identity and how we project ourselves online. In a world where we live increasingly larger portions of our lives online, it is crucial that we be able to manage adverse outcomes.

7. CHALLENGES TO BE ADDRESSED

Given what has been discussed so far, there are quite a few challenges that the research community, industry, policy makers and concerned citizens need to think about to come up with solutions to mitigate adverse outcomes. At the highest level, we need to orient the development and deployment of AI-powered solutions towards enhancing human welfare. But more specifically, to be able to safely unlock the benefits from AI development and to have them be distributed in an equitable, transparent and fair manner will be one of the key challenges facing the community.

There are questions around our ability to "regulate" the development of AI; this idea while noble in its intentions greatly underestimates not only the vastness and distributed nature of research at the moment but also the great difficulty in being able to find such a body that can not only assimilate ongoing research from across the world but also have the right jurisdictional powers that have been agreed upon by such a community to act as their regulators.

The European Union (EU) General Data Protection Regulation (GDPR) [32] serves as a great precedent in the field of data privacy and security on how we can elevate the rights of the users and strengthen the ownership that they can have on their PII. Yet we need to wait till May 2018 when it comes into effect to see how organizations react and how user rights actually get strengthened.

In the context of widely differing standards and approaches used in different industries, one approach to mitigate adverse outcomes is to fold decision-making around best practices into existing regulatory entities within that industry. More so, sharing challenges across research communities, different industries and across geographic borders can hasten the pace at which we discover and develop the right mix of policies that preserve the rights of consumers while not stifling the development of AI. Practitioners can self-impose a "Hippocratic Oath" for AI development as a starting point which while non-binding can at least trigger the right conversations around the distribution of responsibility and share the onus in terms of arriving at solutions to help address some of the issues raised in this paper.

On the data side where organizations today cannot only collect an increasing amount of data on any individual via their online, and slowly their physical, activities as well, they start creating an ability to infer a lot of different facts about an individual. Open datasets that have been de-anonymized fall prey to sophisticated statistical techniques as shown in the case of AOL and Netflix that led to the re-identification of individuals which will now remain in the public domain forever. While it is great to have research efforts that help us identify the shortcomings of de-anonymization and especially public demonstrations of the mosaic effect which help to put the risks in perspective, we need to focus our efforts on being able to create safe ways to advance research efforts, particularly those in the AI community (because they rely heavily on access to large datasets for training purposes).

Some techniques like homomorphic encryption [33] help to address this to a certain degree by leveraging cryptographic and statistical principles to allow computations on encrypted datasets without having to release the underlying data, yet limitations remain around the high cost of computations required to process data in this form. Keeping this in light, there definitely remains dangers of releasing particular snippets of information because it is hard to predict what this can be combined with to unlock other previously existing datasets.

Other challenges remain like the collection of encrypted traffic for future decryption when novel methods might be invented (including the use of quantum computers) further demonstrating the threat from leaked data even if it is in an encrypted format.

8. CONCLUDING REMARKS

The role of the technical community in providing accessible discourse and guiding popular coverage (specifically keeping it within the bounds of practicality) will be crucial as this serves the stakeholders in taking appropriate decisions rather than being swayed by uninformed opinions, especially from a technical feasibility perspective.

It is of paramount importance to educate the public on the existence of such techniques and their effects as highlighted in this paper and the consequences that arise when they are combined together for malevolent purposes. A strong demand from the user base for stricter and more privacy-enhancing regulations will only serve to encourage policy makers and law-making bodies to put together committees to assess the potential impacts and put in place appropriate measures to ensure the wellbeing of their citizens.

ACKNOWLEDGEMENT

The author is grateful to Sydney Swaine-Simon for discussions on the work presented in this paper.

REFERENCES

[1] "'Mosaic Effect' Paints Vivid Pictures of Tech Users' Lives, Felten Tells Privacy Board." Princeton University, The Trustees of Princeton University, www.princeton.edu/news-and-events/news/item/mosaic-effect-paints-vivid-pictures-tech-users-lives-felten-tells-privacy

[2] Williams, Alex. "The Power of Data Exhaust." TechCrunch, TechCrunch, 26 May 2013, https://techcrunch.com/2013/05/26/the-power-of-data-exhaust/

[3] Angela Mitchell, "Artificial Intelligence in Government.", Deloitte UK, 30 June 2017, https://www2.deloitte.com/uk/en/pages/public-sector/articles/artificial-intelligence-in-government.html

[4] Greene, Tristan. "AI Is Analyzing You on Social Media for Market Research." The Next Web, 7 Aug. 2017, https://thenextweb.com/artificial-intelligence/2017/08/03/this-ai-company-leverages-social-media-speech-for-your-marketing-needs/

[5] "How Artificial Intelligence Is Changing Online Shopping." Time, Time, https://time.com/4685420/artificial-intelligence-online-shopping-retail-ai/

[6] Scott Zoldi, Analytics & Optimization, Risk & Compliance. "How to Build Credit Risk Models Using AI and Machine Learning." FICO, 22 May 2017, www.fico.com/en/blogs/analytics-optimization/how-to-build-credit-risk-models-using-ai-and-machine-learning/

[7] Alexander Hellemans, "Why IoT Needs 5G." IEEE Spectrum: Technology, Engineering, and Science News, 20 May 2015, https://spectrum.ieee.org/tech-talk/computing/networks/5g-taking-stock

[8] Unsupervised Feature Learning and Deep Learning Tutorial, https://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/

[9] "Explainable Artificial Intelligence (XAI)." Defense Advanced Research Projects Agency, www.darpa.mil/program/explainable-artificial-intelligence

[10] "Why AI Development Is Going to Get Even Faster. (Yes, Really!)." Import AI, 4 Apr. 2016, https://jack-clark.net/2016/04/03/why-ai-development-is-going-to-get-even-faster-yes-really/

[11] Edith Ramirez et al, "Data Brokers: A Call for Transparency and Accountability", Federal Trade Commission, pp. 97-100, May 2014.

[12] Mirani, Leo, and Max Nisen. "The Nine Companies That Know More about You than Google or Facebook." Quartz, Quartz, 27 May 2014, https://qz.com/213900/the-nine-companies-that-know-more-about-you-than-google-or-facebook/

[13] Singel, Ryan. "Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims." Wired, Conde Nast, 17 Dec. 2009, www.wired.com/2009/12/netflix-privacy-lawsuit/

[14] Michael Barbaro And Tom Zeller Jr. "A Face Is Exposed for AOL Searcher No. 4417749." The New York Times, 8 Aug. 2006, www.nytimes.com/2006/08/09/technology/09aol.html

[15] Stempel, Jonathan, and Jim Finkle. "Yahoo Says All Three Billion Accounts Hacked in 2013 Data Theft." Reuters, Thomson Reuters, 4 Oct. 2017, www.reuters.com/article/us-yahoo-cyber/yahoo-says-all-three-billion-accounts-hacked-in-2013-data-theft-idUSKCN1C82O1

[16] Tzink. "How Rainbow Tables Work." Terry Zink: Security Talk, https://blogs.msdn.microsoft.com/tzink/2012/08/29/how-rainbow-tables-work/

[17] Whittaker, Zack. "AdultFriendFinder Network Hack Exposes 412 Million Accounts." ZDNet, ZDNet, 13 Nov. 2016, www.zdnet.com/article/adultfriendfinder-network-hack-exposes-secrets-of-412-million-users/

[18] Zetter, Kim. "Hackers Finally Post Stolen Ashley Madison Data." Wired, Conde Nast, 18 Aug. 2015, www.wired.com/2015/08/happened-hackers-posted-stolen-ashley-madison-data/

[19] "Pastor Outed on Ashley Madison Commits Suicide." CNNMoney, Cable News Network, https://money.cnn.com/2015/09/08/technology/ashley-madison-suicide/

[20] Ragan, Steve "Raising Awareness Quickly: The EBay Data Breach." CSO Online, 21 May 2014, www.csoonline.com/article/2157782/security-awareness/raising-awareness-quickly-the-ebay-database-compromise.html

[21] "Lyrebird – Create a Digital Copy of Voice." Lyrebird – Create a Digital Copy of Voice, https://lyrebird.ai/

[22] Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman, 2017, Synthesizing Obama: Learning Lip Sync from Audio. ACM Trans. Graph. 36, 4, Article 95 (July 2017), 13 pages. DOI: http://dx.doi.org/10.1145/3072959.3073640

[23] Smith, Ms. "Voice Hackers Can Record Your Voice Then Use Morpher to Trick Authentication Systems." CSO Online, CSO, 30 Sept. 2015, www.csoonline.com/article/2988133/security/voice-hackers-can-record-your-voice-then-use-morpher-to-trick-authentication- systems.html

[24] "23andMe Genotypes One Millionth Customer." 23andMe Media Center, https://mediacenter.23andme.com/press-releases/23andme-1million/

[25] "What Can a Hacker Do with Your Genetic Information?" Motherboard, 26 July 2016, https://motherboard.vice.com/en_us/article/gv5w7j/what-can-a-hacker-do-with-your-genetic-information

[26] 1139 – Cloudflare: Cloudflare Reverse Proxies Are Dumping Uninitialized Memory-Project-Zero-Monorail, https://bugs.chromium.org/p/project- zero/issues/detail?id=1139

[27] "Check out All Twilio APIs." Communication APIs for SMS, Voice, Video and Authentication, www.twilio.com/

[28] Ellis, Emma Grey. "Doxing Is a Perilous Form of Justice-Even When It's Outing Nazis." Wired, Conde Nast, 17 Aug. 2017, www.wired.com/story/doxing-charlottesville/

[29] Sanchez, Cassie. "At a Glance: Generative Models & Synthetic Data." Mighty AI, 22 Mar. 2017, https://mty.ai/blog/at-a-glance-generative-models-synthetic-data/

[30] "Cambridge Analytica – Data Drives All That We Do." Cambridge Analytica – Data Drives All That We Do., https://cambridgeanalytica.org/

[31] "How Trump's Campaign Used the New Data-Industrial Complex to Win the Election." USAPP, 29 Nov.2016, https://blogs.lse.ac.uk/usappblog/2016/11/26/how-trumps-campaign-used-the-new-data-industrial-complex-to-win-the-election/

[32] "Home Page of EU GDPR." EU GDPR Portal, www.eugdpr.org/

[33] Gentry, Craig "A Fully Homomorphic Encryption Scheme", Stanford University, 2009.