Instagram, Amazon, and Machine Learning: Ethical Implications of Collecting and Analyzing Commercial User Data

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Tucker Wilson
Spring, 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Instagram, Amazon, and Machine Learning: Ethical Implications of Collecting and Analyzing Commercial User Data

User Data and its Dangers

In the world of smartphones, social media platforms, and smart home technologies all collecting data on their users, those users are left with questions — what happens to that data, what is it being used for, and how will it affect their life? Companies like Facebook, Amazon, Apple, and Google collect user data through social media sites and devices and then use this data to make predictions on a user's preferences and traits. These traits range from the benign, such as food and drink preferences or preferred clothing styles, to the potentially sensitive, such as political leanings, sexuality, or even predisposition to mental illness. Simultaneously, users are generally unaware of the potentially sensitive insights this data generates, as well as what other companies, organizations, or individuals have access to this data, either legally or through potential security breaches. This STS paper uses frameworks of Actor-Network Theory and Technological Momentum to explore the collection and analysis of user-generated data through two case studies: Instagram, and the Amazon Echo, also called Alexa. Using these frameworks, this paper explores the risks posed to users by these machine-learning technologies and potential steps to mitigate those risks.

Research Question and Methods

This paper seeks to answer one main question: does user data collection and analysis pose security, privacy, or safety risks to the users, and are there potential methods to mitigate such damage?

This paper employs documentary research on primary and secondary sources to answer these questions. Sources include studies performed on users of this technology, court cases raised

against companies that employ these data collection techniques (Rediger, 2017), the European Union's recent update to its cybersecurity standards for private companies (EU, 2016), and case studies on instances where these types of data collection methods have been exploited. These case studies include a showcase of how data gathered from Amazon devices at home and at work can be used to track movements of those near the devices, as an example (Do & Choo, 2018). These sources provide a view of the landscape of user data collection for machine learning and illustrate problems with the current systems.

Technological Background

Currently, many technologies that offer support to individual users, that is, the ability to create and manage a unique profile, also collect data on those users' activities, friends lists, etc. Combined with demographic data, this data creates a comprehensive profile on a user (Malthouse et al., 2018). This data is then used to improve a user's experience through the process of machine learning, which analyzes datasets for patterns and then uses those patterns to make predictions. For example, Instagram may use the friends list of a user to predict what other accounts they might be interested in friending. Besides improving user experience, this process is also a large revenue generator for digital companies in the form of advertising. Machine learning algorithms use this data to identify brands a user might be interested in, and then the site advertises those companies to the user. This has become an incredibly profitable strategy — as early as 2006, Google was generating over $9.5 billion in revenue using its targeted advertising services (Moon & Chen, 2006).

Instagram and Amazon Alexa both expand upon this technology in unique ways. On social media sites, a technique known as sentiment analysis, which analyzes user text posts to determine the post's positive, negative, or neutral sentiment, is used to describe a user's feelings

towards a particular topic of interest or product (Katsurai & Satoh, 2016). Instagram offers a unique opportunity for sentiment analysis by also applying the same analysis of sentiment to the makeup of photos posted and the filters used in editing those photos (Reece & Danforth, 2017).

In addition to the order history and viewing history data collected from amazon.com and the Amazon Echo friends lists, Amazon also uses voice recordings from the Amazon Echo to generate data on its users. Amazon uses this data in training the voice recognition software it uses in the Echo as well as classifying user traits as it does with traditional user data (Rediger, 2017).

This data so far is used mostly for advertising and in those cases is mostly benign. However, this data also poses potential risks. First, this data also offers insights about a user on potentially sensitive topics, such as a user's political leanings. Additionally, this data is not always properly cared for. For example, a 2017 study on breaking the Echo's security (Haack et al., 2017) showed the ease at which Amazon user security PINs could be guessed, and stated that there are perceivable cases where "listeners may be able to recover personal details, including payment information" from the information the Echo is constantly sending to Amazon's servers. These types of risks, and the strategies that could be used to mitigate these risks, are the main topic of exploration in this thesis.

Actor-Network Theory and Technological Momentum

In studying a large, widely distributed network of companies, algorithms, databases, governments, and users, this thesis employs Actor-Network Theory to map complex relations and model their change over time. Actor-Network Theory, or ANT, is a method of formally describing a complex network of stakeholders and contributors to a technological system. It consists of defining actors, which can be people, companies, technologies, or any other entity

that affects the system, and that exist within a network, with unspecified and constantly changing relationships connecting the actors. Defining intermediaries, the languages through which actors communicate, is how ANT seeks to explain the complex relationships within the network. A common critique of this theory is that it is purely descriptive and does not seek to explain the impact any actors have or why the network takes its current form. However, in this thesis ANT is used only as a method of mapping the landscape of user-generated data, not as a framework for explaining the actions of any stakeholders. Therefore, this limitation is not an issue. Instead, the theory of Technological Momentum, first developed by technological historian Thomas P. Hughes (1994), is employed to study how these systems of data gathering and executing on that data grew out of small-scale experiments and will be much harder to change in the modern era. Technological Momentum is a theory that technologies, when first created, are done so because of and subsequently shaped by the society they were created in and the stakeholders that created and used them. However, at some point, a technology grows large enough that further shaping is incredibly difficult and the technology begins to shape the society. One critique of this framework comes from David Nye (2007), who stated that "cultures select and shape technologies, not the other way around," and that no technology is ever "taking humanity somewhere in particular." Nye argues that no technology, no matter the size, is ever deterministic and unshaped by humans. In order to recognize this critique, this thesis seeks to use Technological Momentum to explain why changes to the user-generated data landscape will be more difficult because of its size and avoid the claim that change is impossible.

Analysis, Results, and Discussions

There are several key actors in this landscape identified out of the topic background. First are the users who create accounts with Amazon and Instagram and engage with their

technologies. Second are companies, both Amazon and Instagram and outside parties, who harness the data for machine-learning purposes. Third are government agencies and regulatory institutions, who attempt to set rules on how user-generated data is stored and used. Finally, there are bad-actor individuals and organizations that could seek to exploit user-generated data for nefarious purposes. Note that the methods used to generate and store user-generated data and insights on that data are mostly black-boxed, though some discussion of the technology is still included, to focus on the human interactions surrounding these technologies. Note also that many more stakeholders within this network exist, such as the brands seeking to use targeted advertising, but that they are outside the scope of this topic and thus also black-boxed. With these main players in mind, this analysis seeks to define and consider the implications of the relationships between them. Specifically, this paper investigates how company actors use their power to exploit user actors. As discussed later, these users are generally unaware of the implications their data can have or are powerless to keep their traits as determined by the data private, short of not using the technology altogether. Additionally, this paper demonstrates that these company actors hold greater power than even the governmental entities meant to limit their power, as very little regulation exists on the generation and use of this data. Finally, this paper, by incorporating the theory of Technological Momentum, shows that, while it is the nature of a network in Actor-Network Theory to change, it would be difficult to affect significant change on these relationships.

In the case of the Amazon Echo, many of the concerning vulnerabilities present are linked to the Echo's 'skills,' or its ability to interface with other services such as weather services or other smart devices (Amazon, n.d.). In one such case, researchers at the University of South Australia and the University of Texas at San Antonio (Do & Choo, 2018) were able to

design an adversarial bot to track the schedules of two smart devices: a smart light bulb and a smart switch. On the smart light, they were able to use state pings, requests to receive information on the light bulb's current state (color, power, etc.) as well as pings between a smartphone and the device itself to determine when a user was at home by comparing the light's on/off state to location data from the smartphone pings. On the smart plug, they further defined the details of the user's schedule by tracking the scheduled on/off times of the plug, assuming that the plug would only be set to *on* if the user was indeed at home. All of this was done only by listening to the pings sent between the smart devices, the user's smartphone, and the user's home Wi-Fi network. If the adversarial bot was able to, say, send its own pings through the smart devices, it would have been able to determine with certainty whether a user was home by listening for a response ping from the user's phone (Do & Choo, 2018). While this experiment did not include the Echo specifically, its implications for all smart home technologies, including the Echo, are clear. Not only does the Amazon Echo store its user's schedules, both through alarms, timers, and reminders and through third-party devices such as smart light bulbs and switches, but the Echo can send out its own pings through the 'Alexa, how are you?' command, which pings all the devices Alexa is connected to (Wi-Fi networks, phones, and smart devices) and relays the responses to the user (Amazon, n.d.). Additionally, the Echo logs connections to smartphones not only through the Alexa app but also through other apps such as Spotify, a popular music-listening app (Spotify, n.d.). Thus, not only could a bad actor track a user's home schedule through their own Echo's schedules and smart device and smartphone connections, but also a user's movements to their workplace or to the homes of other Echo owners if that user were to connect to another Echo device to, as an example, play music. All of this is not even considering voice-recognition technology: The Echo records and stores audio from a few

seconds before its activation word ("Alexa") up to when it recognizes a pause in the audio (Jackson & Orebaugh, 2018). If a bad actor were able to access this data, they could use voice recognition to track a particular person wherever they go, so long as an Echo device is being used.

The vulnerabilities in the case of Instagram are related specifically to the deceptively strong predictive power of user activity data. In one particularly unsettling experiment, a team of researchers from Harvard University and the University of Vermont (Reece & Danforth, 2017) used Instagram post data — the content of the post (number of people, location, etc.), image pixel data (overall brightness, hue, saturation, etc.), the engagement statistics of the photo (number of comments/likes), and the user's profile activity (number of followers/following, post frequency, etc.) — to prove two hypotheses: first, that clinically depressed individuals can be correctly identified from controls using only Instagram data, and second, that this identification can be done even before the individual's diagnosis. The model these researchers generated could correctly identify depressed individuals at a rate of roughly 70% accuracy, compared to the 42% industry average accuracy of practitioners (Reece & Danforth, 2017). It is important to reiterate that these insights were generated purely by the information publicly available on Instagram's app. Despite the obvious breach of privacy, particularly when studying individuals prior to their diagnosis, this discovery represents troubling implications for these users. This technology could be used to discriminate against potential employees — employers could use this data to rule out employees that would require costly mental health resources or be generally seen as "unproductive" for their mental illness. This data could also be exploited in healthcare. Though discrimination against individuals with pre-existing conditions is generally illegal, a technology

like this could be used to raise the healthcare prices of individuals who are even still unaware that they have clinical depression.

Beyond the scope of these particular case studies, risks for users are present in almost all forms of social media or any website that may generate user data, such as a search engine. Two studies of Facebook engagement showed user data's ability to accurately predict a user's sexuality. In a study of over 58,000 Facebook users, Cambridge researchers (Kosinki et al., 2013) used Facebook likes, both of other users' profiles and of general topics, and were able to predict homosexuality in men with an accuracy of 88% and homosexuality in women with an accuracy of 75%. Another study of roughly 2,300 male MIT undergraduates (Jernigan & Mistree, 2009) used friends lists of the users to classify the students as homosexual, bisexual, or heterosexual with an accuracy of 83%. Note that both of these accuracy numbers are based on the self-reported sexual orientations of the users themselves. It is easy to imagine how more complex models using multiple Facebook metrics or even metrics from several social media sites could be employed to increase accuracy. One other important factor to note is that in both studies, participants cited that this information was gleaned 'accidentally' — that they felt that they were not outwardly portraying their sexuality in either likes or friends lists, but that nonetheless that information could be obtained without anything but publicly available data. The implications for this data are obvious — not only could this lead to users being targets of bullying or harassment, but technologies predicting sexuality could be employed on a systemic level. Countries with anti-LGBTQ policies, including jail-time for homosexuality, could use this technology to target LGBTQ persons who are not even outwardly expressing their sexuality, only participating in normal online behaviors.

Potential bad actors could use the data generated by these technologies to breach the privacy of users and even actively discriminate against them. Many risks also exist in how these companies store user-generated data and who is given access to such data. The Echo is associated with many data security concerns, particularly when third party skills are involved. A comprehensive study of Alexa's security measures (Haack et al., 2017) revealed many issues. First, the use of 4-digit PIN numbers as a second factor of authentication when issuing voice commands is vulnerable to brute force attacks, as the top 20 most frequent user PINs represent 27% of all PINs used. Second, traffic between an Echo and Amazon cloud services can be monitored through man-in-the-middle attacks and successfully replicated, although the network is well-encrypted. Finally, software and firmware updates come in the form of unsecured HTTP data, opening the door for maliciously-designed software updates to substitute for Amazon-issued code. Besides these technological vulnerabilities, more systematic problems are revealed when studying Amazon's numerous relationships with third-party companies, who mainly interact with Alexa by developing custom skills for users to download. In a study of nearly 12,000 skills available for download (Alhadlaq et al., 2017), 76% of them had no privacy policy at all, and only 3% actually reference the Echo in their Alexa Skill privacy policy. These skills have the same access to audio as Amazon's own skills, as well as access to the device's GPS data and the lists and reminders that the user sets up themselves. In this way, Echo users are exposing themselves to risk by downloading third-party skills. Their data is released to companies other than Amazon who are not obliged under privacy policies to adequately encrypt or depersonalize data. Additionally, they may be legally allowed to sell information on users to other companies at their discretion, and even if they choose not to, their relatively minimal security measures compared to Amazon make them easy targets.

Unfortunately, this landscape is likely already in the state of high structure as described in the theory of Technological Momentum. In the early days of development for this kind of technology, companies like Instagram and Amazon grew this technology landscape by deploying machine learning. These technology companies recognized the money-making opportunity of programmatic advertising — other companies would be willing to pay a premium for ads that were shown to be more effective at actually selling the product in question and, as shown by Google discussed earlier, this strategy was immensely successful. Additionally, these companies also deployed such technologies in the name of user experience. By this time, ads were a well-established revenue generator, and so these companies have used user-generated data to make an inclusion necessary for company revenue into something that, at least on the surface, benefits the users by showing them products that align with their interests. In turn, these companies were influenced by the technology they deployed. Machine learning requires strict parameters and outputs to function properly — these companies had to strictly log likes, views, follows, etc. and categorize *what* users are liking, viewing, and following. Definitions had to be formed for specific types of accounts, whether it be the type of product a business account was promoting or the occupation of an individual influencer, and then these definitions were used in the algorithm themselves. In this way, all popular Instagram accounts have been organized and definitions assigned to them, even if those definitions do not fit neatly. Additionally, users that follow accounts for myriad reasons are now assumed, by the definitions set in place, to be interested not only in that specific account but also in accounts with the same or similar definitions. As is posited by Technological Momentum (Hughes, 1994), technology both influences and is influenced by the societal actors in play. Also posited by Technological Momentum is the idea that this landscape will be harder to affect now that it exists in a state of increased size and

complexity, also called momentum, wherein it becomes "less shaped by and more the shaper of its environment." These algorithms and advertising strategies are deployed not just by Amazon and Instagram, but by Facebook, Google, Twitter, Tiktok, and many other technology and social media companies that employ advertising. To dismantle or change them would require influencing not one but many private entities located across the globe. Additionally, the data used to power these strategies now exists, not only in private and protected Amazon and Instagram databases but publicly on the internet. The famous adage that nothing can be erased from the internet applies well here. Social media data is accessible to anyone with an internet connection and an Instagram account, and those people can freely replicate and store that data on whatever personal devices they please. Erasing this information would be nearly impossible, and thus these kinds of algorithms can always run so long as these platforms exist.

With the understanding that users face significant exposure to risk in interacting with technologies that log user data and that companies are not adequately protecting that data, and therefore their users, and further that change in this landscape would be extremely difficult, what can be done? For one, programmatic changes that could be made by these companies. Amazon, for example, could require a privacy policy specifically written for the data generated by the Echo for any of its third-party skill providers. There are also governmental actions possible, such as the landmark General Data Protection Regulation passed by the European Union in 2018. This regulation heightened the data security standards for any company that collects data on EU residents, requiring getting expressed consent for data collection, fully anonymizing (removing names and any other identifying factors) the data gathered, notifying users of data breaches, and appointing a full-time data-security team. States like California are also moving towards safer data security regulations, including by prohibiting the sale of audio data generated by voice-

activated technologies as of 2015 (Rediger, 2017). These types of policies create a safer data environment for users and protect them against many of the most pressing risks, including the sale of their personal data to other companies and the ability for bad actors to use data to identify them as individuals. However, these regulations do not solve the issues posed by user-generated data. As seen above, many of these insights come from the normal data generated by these technologies or, even more concerningly, from the information that is publicly available on the social media site. However, ways to mitigate these issues do exist on the user side. For example, an exploration of privacy concerns on the Amazon Echo (Jackson & Orebaugh, 2018) recommends several ways to increase the Echo's security capabilities, including muting the device when not in use, changing the device settings to notify the user of requests it receives, and adding a randomly-generated PIN code to be required to use the device. These types of protections could be extended outwards to companies like Amazon and Instagram as well. Amazon could automatically require a PIN code in using its device, or install a factor of authentication that requires a person to be physically in the room to issue commands to the Echo device. Instagram could disable its API, a function published by a company that allows code written by outside entities to request certain pieces of data from that company's databases, that allows entities to gather statistics on an account, sacrificing convenience in order to make it more difficult for bad actors to gather machine-learning readable data on Instagram users.

Limitations and Future Work

It is important to recognize some limitations of this research and its implications. This project has taken place over a relatively short time frame of roughly 6 months. This research has also mostly focused on extrapolation from exposed dangers on relatively new technologies. Luckily, there are few use cases of these vulnerabilities being exploited to cause harm to real

users, but for research purposes that makes it difficult to measure the impact these vulnerabilities could have. Additionally, this type of research is a classic case of the limitations of technological momentum. Amazon, Instagram, Facebook, Google, and similar companies are monoliths of the digital industry. Though governments like the European Union can incrementally change the data landscape, terabytes upon terabytes of user data already exist, and what's more the publicly available information on users cannot be removed unless the user themselves deletes their account. The landscape is unlikely to transform, but it can improve. In the future, this paper could be extended by performing analytical and classification experiments using Instagram data to avoid extrapolation from studies on Facebook. This research could also include more consideration on how users perceive these risks and a discussion of how filling knowledge gaps in user understanding could affect how users interface with technologies and possibly how these companies could respond.

Conclusions

Users of technologies that generate their data on their demographics and user activity face many risks in interacting with these technologies, including the risk of breach of privacy by bad actors, harassment based on their characteristics determined through machine learning algorithms, and exposure to discriminatory policies from private companies or governments. There are solutions, regulatory, by companies, and by users, but they are piecemeal, and drastic change would need to happen for users to be truly protected from these risks. Though pessimistic, this view is a necessary conclusion to accurately represent the risks of user data collection, analysis, and distribution. In order to improve the landscape of data collection, stakeholders must adequately recognized the risks imposed on users, the responsibilities of the

companies and persons gathering this data, the power of governments to regulate such an industry, and the ever-present danger of user discrimination and exploitation.

Works Cited

Alhadlaq, A., Tang, J., Almaymoni, M., & Korolova, A. (2017). Privacy in the Amazon Alexa Skills Ecosystem. *Privacy Enhancing Technologies Symposium*.

Amazon. (n.d.). Understand How Users Interact with Skills. Retrieved February 21, 2020, From https://developer.amazon.com/en-US/docs/alexa/ask-overviews/understanding-how-users-interact-with-skills.html.

Amazon. (n.d.). Use Alexa: Things to Ask Alexa. Retrieved February 21, 2020, from https://www.amazon.com/b?ie=UTF8&node=17934693011.

Do, Q., Martini, B., & Choo, K.-K. R. (2018). Cyber-physical systems information gathering: A smart home case study. *Computer Networks*, *138*, 1–12. doi: 10.1016/j.comnet.2018.03.024.

General Data Protection Regulation (2016) *Official Journal* L119, 4 May 2016, p. 1-88

Haack, W., Severance, M., Wallace, M., & Wohlwend, J. (2017). Security Analysis of the Amazon Echo. Retrieved from https://pdfs.semanticscholar.org/35c8/47d63db1dd2c8cf36a3a8c3444cdeee605e4.pdf.

Hughes, T. P. (1994). Technological Momentum. *The MIT Press*, 101–113.

Jackson, C., & Orebaugh, A. (2018). A study of security and privacy issues associated with the Amazon Echo. *International Journal of Internet of Things and Cyber-Assurance*, *1*(1), 91. doi: 10.1504/ijitca.2018.10011257.

Jernigan, C., & Mistree, B. F. (2009). Gaydar: Facebook friendships expose sexual orientation. *First Monday*, *14*(10). doi: 10.5210/fm.v14i10.2611.

Katsurai, M., & Satoh, S. (2016). Image sentiment analysis using latent correlations among

visual, textual, and sentiment views. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. doi: 10.1109/icassp.2016.7472195.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. doi: 10.1073/pnas.1218772110.

Malthouse, E. C., Maslowska, E., & Franks, J. U. (2018). Understanding programmatic TV advertising. *International Journal of Advertising*, *37*(5), 769–784. doi: 10.1080/02650487.2018.1461733.

Moon, Y. E., & Chen, D. (2006). Google Advertising. *Harvard Business School Case 507-038*.

Nye, D. E. (2007). Not Just One Future. In *Technology matters: Questions to Live With*. Cambridge, MA: MIT Press.

Rediger, A. M. (2017). Always-Listening Technologies: Who Is Listening and What Can Be Done About It. *Loyola Consumer Law Review*, (2), 229–252.

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1). doi: 10.1140/epjds/s13688-017-0118-4.

Spotify. (n.d.). Spotify on Alexa. Retrieved February 21, 2020, from https://www.spotify.com/us/amazonalexa/.