Algorithmic Puppet Masters: The data economy and how our lives are managed, manipulated, and monetized through Algorithms

Dillon Chi and Xinyi Ren

The Dilemma

"It is such a common condition of modern life that roughly six-in-ten U.S. adults say they do not think it is possible to go through daily life without having data collected about them by companies or the government."
-Pew Research

Privacy or Ease of use? It is a dilemma faced by everyone using software, whether that be an app on their phone or service on the internet. People are being traced daily and are not aware of it. The residents of this planet have never before been as polarized and emotionally spent as they are right now. Through the internet, we see incredible acts of kindness, upvoted, and shared across networks. The computer algorithm has learned what keeps our attention by what we clicked and how long we looked at posts.

"So what? Why does it matter if I am being tracked? I've got nothing to hide." This is not true. Algorithms and Artificial Intelligence are profiling individuals by their generated data traces, and utilize predictive analytics to make data-driven decisions about the individual: the daily news feeds, the Google research results, loans, insurance, the employment, even the price on each of the e-commerce platforms differs based on the data traces we have left. People are not aware of how much they are being managed and manipulated in every single choice they make in their daily life.

Moreover, individuals are being monetized by the data they generate on “free” apps such as social media. On average, people spent 144 minutes per day on social media, which has increased by 62.5% since 2012. (Statista and Clement 2020). The attention economy, as the most lucrative part, has the power from algorithms to predict behavior and buying patterns. The so-called free services from social media and other digital platforms are driven by the ad-platform to secure revenue; therefore the algorithm is made to keep users' attention for as long as possible - push notifications and the endlessly scrolling news feeds have created a feedback loop that glued us to our devices. Based on a study conducted by the American Journal of Epidemiology in 2017, higher social media use correlates with self-reported declines in mental and physical health and life satisfaction (Shakya and Christakis 2017).

Cloud Computing, Big Data & Artificial Intelligence

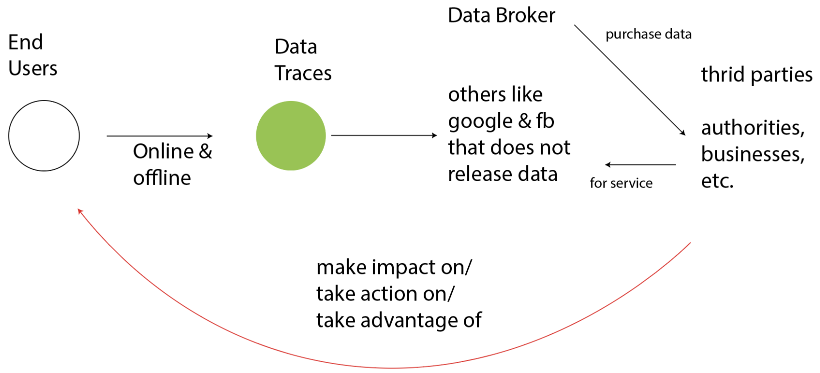
If we look back at the scope of history, we can get a brief overview of how technology evolved into today's data economy and the rise and cause of privacy violations. How did it get to this point that these innovations have become data goldmines? Can we blame Steve Jobs when he turned the simple communication device of a cellular phone into a pocket computer?

Cloud computing has its roots before the internet surprisingly, according to (Foote 2017) MIT and DARPA worked together to "develop a technology allowing a computer to be used by two or more people simultaneously". Several decades later another DARPA project, the internet, would take this concept further and become the platform for the tech companies today to make their millions through cloud computing.

It wasn't until machines started teaching themselves instead of humans telling them how to work that the field exploded. Aided by the massive amounts of data being collected and available on the public web, machine learning has expanded to today where it is being applied to multiple levels of human lives.

The "Predictions" made by Artificial Intelligence

If we look at this evolution through the economic lens, using predictions to advance economically is not new. Advertisers have been using market research for decades to better sell us things. However, due to the availability of data and its traces, these predictions are becoming frighteningly accurate. When technology lowers the prediction cost, we see a rise in more predictions-as-a-business-input applications in multiple ways: traditional businesses that already leverage prediction as input, such as managing inventory can now do so with greater accuracy. We have also seen a rise in problems that we didn't consider as a prediction problem, such as autonomous driving (Agrawal 2018). When the prediction cost keeps dropping, we see the snowball that turned into an avalanche that is the data economy today.


End Users
Online & offline
Data Traces
Data Broker
others like google & fb that does not release data
purchase data
for service
thrid parties
authorities, businesses, etc.
make impact on/ take action on/ take advantage of

.

In the figure above, the process can be simplified as a single structure. Businesses collect data traces from the end-user (online or offline) and use it to create other technologies, share it with data brokers; third parties, and other authorities like government, and companies. They do this either seek for the service based on the data (see Facebook's ad-platform) or purchase data from multiple data brokers and make a profit from the services they create from the purchased data.

What data traces?

Industries that collect data are widely ranged. Data brokers collect information from online and offline sources that range from social media, web history, online and offline purchase history and warranty information, credit card information, government records, and so on.

By aggregating data on how long we spend on posts, what we talk about with our friends through messaging, and what but also what it hears through your phone microphone the algorithms are able to make better guesses of our behavior. The Verge did an article two years ago on a Times report that had identified 250 games on the Google Play Store that monitored user TV watching habits through the microphone. A company called SilverBullet (Waddell 2015) sells their technology to content producers which allows them to embed sound beacons, ultrasonic frequencies inaudible to our ears to allow precise identification of content tying media consumption directly to the users within eyesight.

Radical Monetization

Data brokerage involves sourcing and aggregating data and reselling the most valuable categories of users to third parties. Most people are not even aware that data brokerage exists, but such companies have become a lucrative industry that generates $200 billion in revenue yearly, and it's still growing (Wolosic n.d.). This describes how profitable this business is and thus how crazy companies are chasing data traces by fair means or foul.

Where is the Problem?

A major problem that we identified emerges from two sections in this process - when the data is collected (source) and when the data is applied in making an impact on, taking action on, or making a profit from the end-users (application) - that intrude human rights and leave people vulnerable.

Problem Part 1: Data Collection

Lack of consent

We are not given full consent on whether we allow our data to be collected. Or rather, we need to give up all or part of the services that are needed in life to protect data privacy.
People choose to minimize the problem to "Americans acknowledge that they are not always diligent about paying attention to the privacy policies and terms of service they regularly encounter." (Auxier et al. n.d.).

The missing puzzle

Do we own our footprints in the sand? Do we own the scan of our body produced by the TSA wave-length scanner? Do we have a claim to the billions generated off of our generated traces?
Because initial ownership of these traces is in a gray area we missed a critical opportunity to define it.

Problem Part 2: Data Application

Our calendars are a collection of data points surrounding where we will be and how long we will be there, the title says what we called it. The writers of algorithms take this data and try to write the story of what happened that day. This is an issue because these teams of coders can be biased in their interpretation of data and then they proceed to have machines make decisions based on their biased interpretations. Because the machines are written by humans and they are trained off of humans, all of which are subject to bias, it is a hotly debated topic if Algorithms can be truly objective.

When data is collected, it is only a generated pool of raw materials that waits to be "translated". It is the algorithm that translates data, and the person or team that constructs that translating system from data to a conclusion is crucial. We can never ensure there is an objective "prediction" in the results, yet people are persuaded that any data-driven conclusion is trustworthy and qualified to draw action.

For example the TSA wavelength scanners we brought up before. When a non-binary presenting person steps in, the TSA agent must select from a menu "Male, or female". These two choices then tell the machine what fat deposits are "Normal" or what it has been fed by coders to be "normal". Inevitably trans and cis-non-confomrative bodies are always asked to step out of the line due to this algorithmic bias encoded by the interface and the system designed.

Who is affected by the problem?

Automated decision making and predictive work affect everyone. The systematic collection of data through our internet-connected devices and use of internet software such as social media, browsers, has made everyone that has ever connected online a target. With the advent of cellphones, adults and children began producing data at an unprecedented rate. The mobile sensor platforms in the pocket collect sound, location, sight, and movement data.

Even the people that did not "accept the terms and conditions" are affected. For example, the Verge reported on the robotic Roomba creating blueprints of peoples' houses and building profiles of the people in the household simply through its touch sensors. By bumping around, the Roomba was able to judge whether children's toys were on the ground and with more data could be trained to recognize specific brands of furniture (Deahl 2017).

Algorithms now control and make decisions on who gets hired in Applicant Tracking Systems. Certain ATS systems that post to job boards now use algorithms to reduce cost and limit where ads are seen, based on their location and "likeliness to click". If people live in an area where nobody has ever gotten hired, they may never see that job ad as well. These same algorithms also determine people's experience level from their resume, use quizzes to see what their personality type is, and can leverage big data to make other judgments about them. Algorithms now make suggestions in all parts of our lives, from the court recommendation system, predictive policing, lending, and credit cards.

A software called Compas (Thadaney 2017), relies on historical data to determine how high a risk of returning to jail people have after they have been discharged. Amazon's internal recommendation system for hiring had been trained on the thousands of white cis-male resumes of past applicants that had been hired and consequently learned to penalize the word “women’s” in resumes (Dastin 2018). Most individuals are probably unaware that their fitness tracker could be selling their activity to advertisers and even insurance companies to determine what rate they should be giving or deny coverage for certain devices like CPAPS (Allen 2018). Based on systemic sexism, Apple’s own credit card gave the husband the higher credit limits even with couples that had been filing jointly for years (Vigdor 2019). Algorithms also make decisions on where police should be sent, based on historical data software such as PredPol (Chammah 2016). This company promises to predict where crime will take place, much like in the Minority Report based on even weather patterns to determine where police should be dispatched.

“If the service is free, you are the product”

Data-driven products and services are often marketed with the potential to save users time and money or even lead to better health and well-being (like the fitness data tracker). Many Americans are willing to share personal information in exchange for tangible benefits (Kerry 2018)

That adage “If it’s free you are the product” rings true in the world of data gathering and brokering. One of the biggest issues with the adage is that it discounts the fact that the selling and monetization of data about the user could and probably subsidizes the cost of the product. The computer industry and the advertising industry conspired together to create the monster that exists today. The feedback loop has produced incredible amounts of data that are sold/given to every sector possible from third parties governments, credit bureaus, and colleges.

Social Discrimination

We know that companies can use data pieced together into big data to create correlations and that these machines create patterns about social category membership and traits. The patterns are later used to make suggestions. Even in cases where users leave out social categories such as Race, Gender, and age, big data can guess. In a paper called “How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications” we learned that machines can still target certain communities based on the patterns of individuals within the same social systems and shared influences as proxy variables. Instead of calling it, “race” Facebook guesses a proxy, a users' “Multicultural Affinity,” based on their site interactions. It then allows advertisers to include or exclude certain groups, which it argues lets companies test different versions of their ads (Williams, Brooks, and Shmargad 2018).

The Losing Battle

Despite that the government is trying hard to save the losing game, tech companies find loopholes to ignore or flout government policy. For example ridesharing companies such as Uber and Lyft have been a point of contention for the past several months have ignored and slightly changed their business model despite legislation saying that their contractors are actually employees.

Unchecked companies continue to develop AI systems without heeding to ethics. The current way of doing business is lucrative. Without government regulation or intervention, it becomes more and more pervasive an issue. There is a misuse of big data at the country government level, for example as China continues to develop its “blood Skynet” as reported by the NY Times (Wee 2020). According to the article, over the past 3 years, the government has systematically traveled the country collecting blood samples to create a vast genetic intelligence network capable of identifying and tracking down relatives.

On top of the normal distrust of governments, there have been numerous data leaks by the Department of Justice and the US Military. for example, data on 4000 US Airforce officers and their families on “unsecured network backup drives" of the US Airforce had “names, addresses, social security numbers, and other contact information” in a passwordless format (TrendMicro 2020).

A research team called VPNMentor found 85,000 customer files on more than 30,000 customers or the retailer THSuite (as reported by TrendMicro) (TrendMicro 2020) were exposed on the internet. These files were found in a “bucket” on Amazon's Web Service platform called S3, the same service trusted by conglomerates, GE healthcare, the Nasdaq, and Unionbank (Amazon n.d.)**.**

The Equifax Data breach in 2017 is a great and terrible example of how a private company through poor data security and IT practices left their internal data on nearly half of the United States open to hackers through a widely-known vulnerability. This data included “ Names, addresses, dob, SSN, and driver licenses. Only 200,000 of those records were direct consumers of Equifax (Fruhlinger n.d.).

With the fast speed in device and algorithm updates, the law system is falling behind. More and more data about each of us is being generated faster and faster from more and more devices without protection (Kerry 2018). The current law system does not provide adequate protection to people because there is a lack of significant punishment on organizations or companies that engage in data breaching and leaking.

The Silver Lining

Legal systems in other countries start to bring awareness and impose punishment on a higher level to address the long-held problem of data leaks and hacks. In October 2020, China, as the country with the world’s most online users, unveiled its draft law that is highly participatory on personal data protection. It states that those who violate the law could face a fine up of up to 50 million yuan ($ 7.4 million) or 5 percent of its past year’s turnover, which will strike a heavy blow to organizations, enterprises, and individuals who have constantly disturbed people’s lives by illegally collecting, using and trading personal information for profit (Global Times 2020).

EU passed “General Data Protection Regulation” (GDPR n.d.) or GDPR Q2 of 2018. This forced a lot of companies that used services provided or operated in the EU to tighten up their data handling practices, as well as changing the way that they ask for consent from the user.

Other corporations emerge to put effort into protecting individuals from data leaks and breaches as well. Websites such as “haveibeenpwned.com” have sprung up over the years to help users keep track and see if their information or account passwords might have been hacked. Corporations like Mozilla have developed their tracking-fighting extensions “Facebook container” which “Stop Facebook from following you across the internet. With the Facebook container add-on for Firefox, people take control and can easily isolate their activity on the net from that on Facebook.

What’s more, citizen-participation kicks in the battle. Research groups have created what are called “contestational AI” programs designed to thwart data collection. For example, there is a project called” Trackmenot” which was developed by Vincent Toubiana. According to its website, it is “ a lightweight browser extension that helps protect web searchers from surveillance and data-profiling by search engines. It does so not utilizing concealment or encryption (i.e. covering one's tracks), but instead by the opposite strategy: noise and obfuscation.” In the background, this extension will run ghost queries to throw off the sent of trackers.

What’s Next?

Many people are only really recently understanding the impacts of algorithms on their lives because of the new Netflix documentary, “The Social Dilemma” (“The Social Dilemma - A Netflix Original Documentary” n.d.). But this documentary just barely scratches the surface. Algorithms aren’t all bad, the fairly innocuous ones can be found in your car, making decisions on how to shift gears and when but their application has stretched way farther than that. Veronica Barassi, a child data science advocate said it best, “Artificial Intelligence and predictive analytics can be great to predict the course of a disease or to fight climate change, but we need to abandon the belief that these technologies can objectively profile humans and that we can rely on them to make data-driven decisions about individual lives.” (Barassi n.d.)

The problem is two-fold - unchecked companies will do what is best to pad their bottom lines, just as the example of the consulting firm who provides service of guiding businesses on how to sell their customers and their data. Even under COVID, Companies like “Enaible” are jumping into the market when they saw surges in the need for AI to remotely monitor employees during quarantine (Edelman 2020). It promises to give the executives the information they need by seeing how often employees use social media, write emails, and even takes into account biometric data such as eye-tracking.

The law needs to keep up with companies because it is clear that companies will not choose ethics over profits. There leaves a question mark as to what steps need to happen in the United States, and how far are we from creating a comprehensive set of guidelines from GDPR. The pandemic has shown that even if the information is available to the people, the masses are unlikely to heed them - the government needs to step in to regulate companies to protect the people from data insecurity.

Work Cited

Agrawal, Ajay. 2018. “The Economics of Artificial Intelligence.” McKinsey & Company. April 28, 2018. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-economics-of-artificial-intelligence#.

Allen, Marshall. 2018. “You Snooze, You Lose: Insurers Make The Old Adage Literally True.” ProPublica. November 21, 2018. https://www.propublica.org/article/you-snooze-you-lose-insurers-make-the-old-adage-literally-true.

Amazon. n.d. “Amazon S3 Customers.” Amazon Web Services, Inc. Accessed October 26, 2020. https://aws.amazon.com/s3/customers/.

Anyoha, Rockwell. 2017. “The History Of Artificial Intelligence - Science In The News.” Science In The News - Harvard University. August 28, 2017. http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

Auxier, Brooke, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. n.d. “Americans And Privacy: Concerned, Confused And Feeling Lack Of Control Over Their Personal Information.” Pew Research Center: Internet, Science & Tech. Accessed October 19, 2020. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

Barassi, Veronica. n.d. “What Tech Companies Know About Your Kids.” TED. Accessed October 27, 2020. https://www.ted.com/talks/veronica_barassi_what_tech_companies_know_about_your_kids/footnotes?trk=organization-update-content_share-video-embed_share-article_title.

Caddy, Becca. 2020. “Are Fitness Trackers The Future Of Healthcare?” TechRadar. September 28, 2020. https://www.techradar.com/news/are-fitness-trackers-the-future-of-healthcare.

Chammah, Maurice. 2016. “Policing The Future.” The Marshall Project. February 3, 2016. https://www.themarshallproject.org/2016/02/03/policing-the-future.

Dastin, Jeffrey. 2018. “Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women.” U.S. October 10, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

Deahl, Dani. 2017. “Roombas Have Been Busy Mapping Our Homes, And Now That Data Could Be Shared.” The Verge. July 24, 2017. https://www.theverge.com/2017/7/24/16021610/irobot-roomba-homa-map-data-sale.

Doffman, Zak. 2019. “Hong Kong Exposes Both Sides Of China’s Relentless Facial Recognition Machine.” Forbes. August 19, 2019.

https://www.forbes.com/sites/zakdoffman/2019/08/26/hong-kong-exposes-both-sides-of-chinas-relentless-facial-recognition-machine/#3bb33dcd42b7.

Edelman, Gary Grossman. 2020. “Work-at-Home AI Surveillance Is A Move In The Wrong Direction.” VentureBeat. July 18, 2020. https://venturebeat.com/2020/07/18/work-at-home-ai-surveillance-is-a-move-in-the-wrong-direction/.

Firefox. 0. “Facebook Container – Holen Sie Sich Diese Erweiterung Für 🦊 Firefox (De).” Mozilla.Org. 0. https://addons.mozilla.org/de/firefox/addon/facebook-container/.

Foote, Keith. 2017. “A Brief History Of Cloud Computing - DATAVERSITY.” DATAVERSITY. June 22, 2017. https://www.dataversity.net/brief-history-cloud-computing/.

Fruhlinger, Josh. n.d. “Equifax Data Breach FAQ: What Happened, Who Was Affected, What Was The Impact?” CSO Online. Accessed October 26, 2020. https://www.csoonline.com/article/3444488/equifax-data-breach-faq-what-happened-who-was-affected-what-was-the-impact.html.

GDPR. n.d. “General Data Protection Regulation (GDPR) – Official Legal Text.” General Data Protection Regulation (GDPR). Accessed October 26, 2020. https://gdpr-info.eu/.

Global Times. 2020. “China Unveils First Law On Personal Data Protection - Global Times.” Copyright 2019 By The Global Times. January 10, 2020. https://www.globaltimes.cn/content/1203363.shtml.

haveibeenpwned. 0. “Have I Been Pwned: Pwned Websites.” Haveibeenpwned. 0. https://haveibeenpwned.com/PwnedWebsites.

Howe, Daniel C. 2019. “TrackMeNot.” March 17, 2019. http://trackmenot.io/.

Jowitt, Tom. 2017. “Classified US Army Data Found Unprotected On AWS Server.” Silicon.Co.UK. November 30, 2017. https://www.silicon.co.uk/security/cyberwar/us-army-leak-225441.

Kastrenakes, Jacob. 2018. “Some Android Games Are Quietly Using Your Microphone To Track Your TV Habits.” The Verge. January 2, 2018. https://www.theverge.com/2018/1/2/16842294/android-apps-microphone-access-listening-tv-habits.

Kerry, Cameron F. 2018. “Why Protecting Privacy Is A Losing Game Today—and How To Change The Game.” Brookings. July 12, 2018. https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/.

Kim, Whizy. 1970. “Uber Says Gig Workers Are Their Own Bosses — But Courts Disagree.” Refinery29. January 1, 1970. https://www.refinery29.com/en-us/2020/08/9958817/what-is-gig-worker-uber-employees-ceo.

Lapowsky, Issie. n.d. “Mark Zuckerberg Answers To Congress For Facebook’s Troubles.” Wired. Accessed October 27, 2020. https://www.wired.com/story/mark-zuckerberg-congress-facebook-troubles/.

Leetaru, Kalev. n.d. “The Data Brokers So Powerful Even Facebook Bought Their Data - But They Got Me Wildly Wrong.” FORBES. Accessed October 27, 2020. https://www.forbes.com/sites/kalevleetaru/2018/04/05/the-data-brokers-so-powerful-even-facebook-bought-their-data-but-they-got-me-wildly-wrong/#44f3de803107.

Pudwell, Sam. 2017. “US Military Suffers Massive Leak Of Confidential Air Force Data.” Silicon.Co.UK. March 14, 2017. https://www.silicon.co.uk/security/us-military-data-leak-207093.

Ruokonen, Mika. 2020. “Six Inspirational Ways To Make Money With Data — Futurice.” Futurice. August 5, 2020. https://futurice.com/blog/six-inspirational-ways-to-make-money-with-data.

Shakya, Holly B., and Nicholas A. Christakis. 2017. “Association of Facebook Use With Compromised Well-Being: A Longitudinal Study.” *Am. J. Epidemiol.*, February. https://doi.org/10.1093/aje/kww189.

Statista, and J Clement. 2020. “Daily Social Media Usage Worldwide | Statista.” Statista. February 26, 2020. https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/.

Stern, Joanna. 2018. “Facebook Really Is Spying On You, Just Not Through Your Phone’s Mic.” WSJ. March 7, 2018. https://www.wsj.com/articles/facebook-really-is-spying-on-you-just-not-through-your-phones-mic-1520448644.

Thadaney, Ellora. 2017. “DOpinion | When An Algorithm Helps Send You To Prison (Published 2017).” The New York Times. October 26, 2017. https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html.

“The Social Dilemma - A Netflix Original Documentary.” n.d. The Social Dilemma. Accessed October 26, 2020. https://www.thesocialdilemma.com/.

Thompson, Stuart A., and Charlie Warzel. 2019. “Opinion | Twelve Million Phones, One Dataset, Zero Privacy.” The New York Times. December 19, 2019. https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html.

TrendMicro. 2020. “Unsecured AWS S3 Bucket Found Leaking Data Of Over 30K Cannabis Dispensary Customers - Security News - Trend Micro USA.” TrendMicro. January 27, 2020. https://www.trendmicro.com/vinfo/us/security/news/virtualization-and-cloud/unsecured-aws-s3-bucket-found-leaking-data-of-over-30k-cannabis-dispensary-customers.

Vigdor, Neil. 2019. “Apple Card Investigated After Gender Discrimination Complaints.” The New York Times. November 10, 2019. https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

Waddell, Kaveh. 2015. “Your Phone Is Listening—Literally Listening—to Your TV.” The Atlantic. November 19, 2015. https://www.theatlantic.com/technology/archive/2015/11/your-phone-is-literally-listening-to-your-tv/416712/.

Wee, Sui-Lee. 2020. “China Is Collecting DNA From Tens Of Millions Of Men And Boys, Using U.S. Equipment.” The New York Times. June 17, 2020. https://www.nytimes.com/2020/06/17/world/asia/China-DNA-surveillance.html.

Williams, Betsy, Catherine Brooks, and Yotam Shmargad. 2018. “How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications.” *Journal of Information Policy* 8: 78. https://doi.org/10.5325/jinfopoli.8.2018.0078.

Wolosic, Michal. n.d. “What Is A Data Broker And How Does It Work? - Clearcode Blog.” Clearcode | Custom AdTech And MarTech Development. Accessed October 26, 2020. https://clearcode.cc/blog/what-is-data-broker/.

Zhao, Jingcong. 2020. “Why Is Data Privacy Important And How To Stay On The Right Side Of Data Privacy Laws.” Hyperproof. February 5, 2020. https://hyperproof.io/resource/understanding-data-privacy/.