Spring 5-5-2023

Digital DNA: The Ethical Implications of Big Data as the World's New-Age Commodity

Clark H. Dotson
Coastal Carolina University, chdotson@coastal.edu

Digital DNA:
The Ethical Implications of Big Data as the World's New-Age Commodity

By

Clark H. Dotson

Information Systems

Submitted in Partial Fulfillment of the
Requirements for the Degree of Bachelor of Science
In the HTC Honors College at
Coastal Carolina University

Spring 2023

Louis E. Keiner Director of Honors HTC Honors College	Ross Foultz Doctor of Computer Science Concentration Information Assurance Department of Computing Sciences Gupta College of Science

Abstract

In the emerging digital world that we find ourselves in, it becomes apparent that data collection has become a staple of daily life, whether we like it or not. This research discussion aims to bring light to just how much one's own digital identity is valued in the technologically-infused world of today, with distinct research and local examples to bring awareness to the ethical implications of your online presence. The paper in question examines anecdotal and research evidence of the collection of data, both through true and unjust means, as well as ethical implications of what this information truly represents. Through examining the leaps made in machine learning and digital forensics, the information about you is often just as valued as gold… Even more so, as its use in advanced AI (artificial intelligence) systems has begun to skyrocket in recent years. Such a thorough investigation on Internet identity aims to pull back the curtain over the good, the bad, and the ugly when it comes to ethics and data privacy, all to find out if your so-called "Digital DNA" is often more coveted than the real deal.

Keywords: artificial intelligence, big data, hyper-segmentation, data analysis, ethics

Contents

I.) Introduction: A Supposed Expert Weighs in on the Subject?

> *The ethical implications of Big Data are very powerful in relation to being a commodity. There are a large variety of ways that these implications have an effect on people living their lives, both in an online space and in an offline space as well. Understanding these implications is only the first step of being able to change them, as large amounts of data continue to become very important in all facets of tech work. Due to the use its in many large computing systems, being aware of ethical requirements for protecting data is only the first step, as we must ensure that we protect this digital form as if it was a real thing. You should always try to protect your digital self. After all, what have you got to lose?*
> (ChatGPT, 2023)

It may surprise you to learn that this introductory quote was written by an artificial intelligence. Now, being so brash as to make this statement within the opening of a paper may be seen as bombastic, strange, and potentially damaging to the argument being upheld through the title itself. Shouldn't I know that machine writing is a massive point of modern contention?

And yet, the excerpt included does exactly what it needed to do. The purpose of this inclusion of an artificially-generated statement is to highlight that from a glance, the introduction appeared as if nothing was wrong whatsoever. Looking over the excerpt shows very little indication of the text in question being different from any other piece of the document. It appears normal, and is normal, as up until reading this explanation, the artificial text was very much indistinguishable from the real deal. It is this concept that drives the idea of the discussion at hand, the main factor that forgoes what we understand as being the so-called correct form for the idea of the self. If an author is to truly write something that they wish to speak on, an artificial intelligence cannot speak on their behalf! It must be from the author, by the author, and of the author's intent! That is what makes it their own, right? Indeed, this is the main issue that arose when I decided to include this artificially-made section. Besides, who would I even cite as my source? Could an AI even give permission to be cited in a thesis paper? And yet, what makes the discussion of artificial intelligence in media that much more potent is the fact that the words written by the artificial intelligence, throughout the entire introduction just discussed, are entirely learned through the data given to it from hundreds of scholars, thousands of papers, and millions of sources. With so much digital data flowing through it, the neural network is technically the most versed to speak on the subject!

Of course, I do not truly believe this statement of hyperbole. Instead, this where exactly where the discussion needed to start. The moral implication of AI and the concept of online information is a topic that has come under great scrutiny in the modern age, a constant struggle of where exactly it begins or ends. Within the technological age that we find ourselves, it becomes more and more apparent that the concept of a digital self is all the more implicit to that of a person. Gone are the ages in which separation could easily be made, for now technology has bridged the gap by such a large margin that it is often difficult to distinguish the difference between each. It is through a thorough examination of each individual factor that has begun changing the digital landscape, one to directly understand how this online fingerprint is changing the lives of those offline. This concept has been expressed by Internet historian and researcher Luca Belli (2018) as a "Network Self-Determination," an innate and active part of our lives, a so-called digital DNA. This online presence that constitutes the information regarding a person has only become more and more important as technology continues to evolve, especially in relation to the field of data analysis and information systems. The digital self has also become far more coveted as artificial intelligence systems and data brokers require information more and more, which has turned the data on a person into a commodity that is just as valued as silver or gold. A rare treasure, and one that is held in such high regard that it is often becoming more valued than the real person it is tied to. The artificial response I generated to begin the text should be proof enough, where even the smallest modicum of data can create entire descriptions down to the minute detail of writing style. It is due to this emerging New Age commodity that the concept of self must be protected despite the impressive creations using such information through leaps and bounds, as the world of tomorrow may be far more different than we expect due to the rapid technological feats of today.

II.) The Good: The Origins of Data Analysis and the Evolution of Big Data

> *The pace of technological change has not slowed down… Some of the traditional views that the building of the technology would require the large companies that have taken long-term views, that's been reinforced. But the advances in the connectivity speeds, the advances in the chips that drive the different devices from the PCs to the new mobile phones to the new types of entertainment that we'll have in the home environment; all of these things are proceeding at full speed. So we can expect power and connectivity as much as we desire. The only limiting factor in the whole picture is what kind of software is created, what kind of easy to use applications can be built*. (Gates, 2001)

A few months ago, the Midway Fire Rescue company of Conway, South Carolina came to an agreement with Coastal Carolina University for the graduating seniors of the Information Systems capstone program to develop an online application able to store information in a database. This specific application was designed to replace the rather obtuse and outdated paper forms that had been previously implemented. Up until that point of time, the entire fire department would fill out all inspections, documents, and information on paper forms, which would then be manually typed and entered into a digital spreadsheet, then stored directly on a computer within the firehouse itself. With the creation of this database project, the information would be able to be input through an online form, which then could have a direct implementation into a full SQL database. With full online access and login credentials for administrator access, the project was to be fully secured and a direct upgrade from the previous paper form and unsecured computer storage. That was, however, until the day on which the project itself was to be proposed to the fire department in question. For on that previous day before the students were to meet, a malicious bad actor had gained access to the fire department’s system holding the original data, having made off with all the information collected over the decades that the department had been open (GAB News, 2023). It truly was a terrible robbery of such valuable information, all done in a single clean sweep of a cyber-attack.

But instead of the criminal making off with information about their monetary systems, banking systems, or any other seeming-valuable information like that, the only thing that was taken was the files of information. Information on the clients, information on where they were located, information on their phone numbers, their businesses, their constituents, their livelihood. To the hacker, this strange collection of information was more valuable than any of the financial details that could have been stolen. Why was that?

Upon hearing this information during the meeting with Midway Fire Rescue, I was rather surprised to discover this was the case for our capstone project. In addition to now being unable to create the system with online specifications, we would be barred from handling any data to make the system in the first place. This, of course, was understandable considering the circumstances, and yet, that was not what was caught in my head. It was the fact that this bad actor had not made off with banking details or monetary finances, all stored in the same system, but simply made off with… information. This was the question that heralded my topic of understanding, and one that I wished to examine further. Why was there such a desire for this general information, when the monetary information would have been far easier to steal? The answer, surprisingly, is far more complex than one may first assume, and one that brought to light the intricate world of data information to mind.

The Midway Fire Rescue company is not the only one to have this happen, and certainly will not be the last. Through surveys conducted by Duke University (2022) for both private, public-owned, and governmental businesses, approximately 80% percent of companies experience some form of a privacy attack on their systems, certainly highlighting the fact that this is no singular example. When a company has data, someone will always want to obtain that data, for the information it provides inherently makes it valued. This is often the reasoning for bad actors to attempt to obtain this information in the first place, and only serves to note that there has always been and still will be a desire to obtain information the collection of information. But if the information is just that, simple information, why is there a need to take it? For this, the very fabric of information systems must be examined. Both in the example of the fire department and business in general is often applied for the purpose of ensuring that future attempts have documented results that can be utilized. Information protection is just a part of working in an online world, now referred to as “data storage compliance” as a concept for keeping it secure as a requirement (Corcetti, 2021). This seemingly innate need to continuously keep track of information is not one of new design due to bad actors, even if the applications of data analysis are rooted in the concepts that make them seemingly negative and detrimental. Despite such doom and gloom, there are a multitude of positive aspects that allow data collection to be utilized in a fashion that assists the user, which is often the reason for its implementation.

However, the idea of data storage compliance, or simply data storage in general, is not a new concept for people. When it comes down to discussing exactly how data is used within the modern age, most of discussions would have assumed to be negative applications, especially with the previously mentioned example. The obvious conclusion would be to immediately jump to the thought of how horrible, terrible, and unsafe every single tech system is. After all, it was this technology that allowed hackers to take down that fire department! Yet, as apt as such a statement may be after having just heard about the hacker in question, that is not to say that there are not good and effective applications for the utility of digital data analysis.

Data itself, as a concept, has always been an innate part of the world around us. Even before the term existed directly, there has always been the analysis of data in human life. Of course, this idea of data being typed into computers and checked into spreadsheets certainly did not exist until they were invented, and yet there was still the need to look over information and draw a conclusion. Even early humans would take in information and gather data, far before civilization as we know it today. Dating even all the way back as 19,000 BC, the use of the Inshango Bone has highlighted a form of information collection even in early humans, carved notches into animal marrow for making astrological marks of mathematical tallies in order to keep a record of information learned (Shatby, 2022). Acting as some of the first forms of data storage, they highlight the fact that information has always been collected and stored as an innate part of human living. These engraved bones would then be kept, far before the creation of books or calculators, and yet serving a very similar purpose through the ages. It should really come as no surprise that even thousands of years in the past, we as a people were collecting data. The long and short of it is that data analysis is simply a part of human interaction, an innate example of our direct desire to analyze and understand. And as technology begins to move forward, there is also a push for the development of better data systems, where the concept of analyzing data is often the driving force in the development of this. Another defining point in data analysis, though perhaps quite a large jump in time between the previous example, was the creation United States Census during the 1800's, and the rather humble and tedious beginnings it was derived from. The census itself was a new-age form of data collection, able to garner demographic information on the various people that lived in the country. However, there was just one slight problem. With such a massive number of people across the entire country that needed to be surveyed at that point, the current system of paper forms and tallying results by hand was far too outdated and sluggish to keep up with the massive data set of the entire country. Due to this, they might as well have been using the carved bones mentioned previously, due to the archaic functionality of the system in question. Humor aside, this was all changed when Herman Hollerith, a German-American statistician, invented a device that utilized punch cards to perform calculations and take in the collected information. This was a major step in the field of data information, where analysis that used to take years to complete could now be finished in only a few months (Rudat, 2018). Information on the country and those who lived within could be obtained in a far shorter period, allowing the information to be utilized far sooner than before. The census was faster, easier, and more efficient all through the machine technology developed by Hollerith, paving the way for the more modern forms of calculators and computers that are seen today.

Now, one might be unsure of why exactly these examples are given when noting the idea of digital media. A punch-card machine and a dusty old bone are not exactly stellar examples of high-end information tech. Simply prattling off points of collecting data throughout history may seem to be a waste, but there is a method to the madness, for these examples of individual forms of data collection evolved past the design of the previous version, which slowly begins to build up a much more modern understanding of how exactly this collection of information has gotten to where it is. The reason that I bring up these for discussion is to highlight that there always was, and always will be, the collection of information. Humans will always want to learn, and as technology grows more powerful, so will the ability to collect data. The evolution of data analysis has brought it to the modern form as it is known today, and it will continue to grow as technology improves as well. This is the direct foundation that creates the core of data analysis itself, and it is the one of the founding factors of the concept of Big Data. This executive-favored buzzword is often thrown about in the field of information systems, yet its importance cannot be overstated. This concept of data sets is based around the fact that large sets of information on people's geography, life, style, and choices can create a better understanding of why those choices are made, with the ability to predict future choices (Rudikowa, 2019). That information can be directly collected and utilized through a massive lens by the observer in question. As silly as it sounds, big data is very much just a simple name for a very large amount of data, where it is far too large for traditional analysis to manually look over. Big data is separate from the previous forms of data collection that had been mentioned for the fact that it pushes past one of the main limiting factors of before, that being of time. Time is a concept that has a great and powerful limitation upon the analysis of information. The amount of time that it takes to analyze any form of information is often what causes the most trouble and ends up being a great cost. If information cannot be analyzed in time, it often is no longer useful, as new information has come along to make it outdated. This was often a limiting factor in decades past, where information collected needed the proper time to analyze it, be it years, months, or days. It is here with the application of big data that that final hurdle is finally leapt over. There is no need for time to be a constraint, for algorithmic design allows a system to look over the information without any need for time. The number of calculations that can be produced in such a short period has made the analysis of big data monumentally easier, which has hastened the data-brokering effects of informatic systems as a whole.

If all of that information sounded like a great load of technical jargon, what it essentially boils down to is the fact that large amounts of data being able to be processed in a short period of time has become quite the fancy in a modern age. Data itself, where it was used specifically to learn about how to obtain commodities, has now become something of a useful commodity all on its own. It is this ability to process so much information that allows larger systems to even function as they are known today, as the data itself is often what allows them to run. Be it a small company like the fire department, or a large website such as Amazon or Google; both are fully dependent on the ability to process immense amounts of data in a short period of time in order to provide the final results to the customers and users that are utilizing their applications. As of April of 2021, Google had an approximate 3.5 billion users accessing from within the United States alone (Walsh, 2021). With such a large number of users continuously accessing the platform, there will always be a need for more information in order to tailor these designs to work exactly for how a customer requires them. And this same concept is seen throughout whether it was this small database system created by senior college students for their capstone, or a massive database system that is used to power Amazon services. The same idea is held between both, in which information must be taken into the system and processed to provide an output that can be read.

All of these big terms, technical notes, and corporate buzzwords apply themselves in such a vague and overarching formats that it feels as if they are concepts far above you or me, invisible hands that push and pull the market through some unseen force. However, the opposite could not be more of the truth. Big data is built from the information collected, which is how the trends generated express itself. It is for this reason that the malicious actor mentioned in the fire department anecdote was so eager to get their hands on such information above all else. For as we see, like all forms of collateral, information itself is power.

III.) The Bad: Information in Marketing and the Emergence of Hyper-Segmentation

> *We are not yet at a point in time where we must seriously consider the ubiquitous consequences of most human bodies as specific data collection sensors. However, given the widespread use of smartphones and the vastly increasing markets for other sensor-based personal devices, it is becoming essential to consider the implications of life in a smart world. Such considerations become even more pressing when the context of the smart world is extended beyond devices that are oriented towards the individual, to the next level.* (Burdon, 2020)

Having examined why exactly there is this desire for data, we then move forward from the need into the want. And it is this want that drives the concept of digital DNA as a fingerprint of one's online presence. If technology has only enabled the ability to collect information, then conversely, there is the reason for such ability to have been pushed in the first place. Yes, as long as there is the ability to collect data, there is the ability to use data. It is here that another important aspect of information systems must be acknowledged, as it is just as important to the concept of digital DNA as big data is to data collection. And yet, it is a concept that does not innately apply specifically to that of information systems. Instead, it is a fundamental form of another defining factor within societal implication, being that of business and marketing.

Let us say a company wishes to sell a product to a customer. They must know that the person even wants to buy it, otherwise there is no product worth selling. This is a gross oversimplification of the process within business, and yet it is this basic factor that gives the simplistic form a highly understandable fashion, especially in relation to that of information collection. Indeed, information on the customer means that a company is more equipped to sell to that customer, because they know what the customer is looking for. If a company knows what someone wants to buy, they can more easily sell the product, which helps them out instead of similar tossing their idea out into the open seas of the mass market. With the example of a company looking to better understand the demographic that they are attempting to sell for, many businesses will implement a concept known as marketing segmentation. With varying groups of customers that exist in the world, trying to appeal to all is a recipe for failure. Instead, the ability to understand which customers want what product allows companies to appeal to a consumer more accurately, which turns a mass market into segmented groups of similar wants or needs (Pride et al., 2018). Although it might sound a bit over the top for what is essentially just making something someone wants to buy, what segmentation essentially boils down to is understanding what people fit into what groups, and then knowing which of those groups to attempt to sell something too. If you are computer mouse salesman attempting to market your brand-new ergonomic mousepad, it is far easier to sell a computer mouse to a person that owns a computer in the first place as a peripheral, rather than someone that does even not own a computer and cannot use the product. The product being provided is specifically tailored to the demographic of those that own a computer and would need a mouse to navigate the desktop, and thus, knowing that a specific group of people own computers means that those people are the most accurate and effective targets to sell to. It applies to the concepts of supply and demand, yet specifically whittling down the demand to people that would demand the product in question. As expressed by an examination on the application of marketing University of Minnesota (2016) for the use of supply, any company can indeed attempt to supply, but supplying everything to everyone is simply bad business. Supplying only what is needed, or what is wanted, creates a more sustainable form for any business. From here, there is a specific portion of marketing segmentation that applies rather well for the data analysis context at hand. This is known as demographics segmentation, which is finding a precise form of audience due to specific data points that could be shared between each of the segments. It is usually the most common method when it comes to marketing, due to the overall efficiency that it provides, and marketers can directly search for specific identifiers that are shared. Knowing the audience of a product allows the creators to figure out who exactly is buying this product, and why they may be buying this product. Due to the upheaval of technological data processing, as mentioned previously, the turn of the century marked when businesses began move towards a much larger and commercialized scale. It was the shift in understanding between a business and a consumer that began marketers' implementation of segmentation to discover what users of a product truly desired, and which users of that product were those to be catered to. And for a long while, this was the way things were.

However, this all came to a grinding halt with the first application of the World Wide Web, the online systems that evolved to be known as the modern Internet. Marketers can no longer rely on traditional forms of demographic segmentation to obtain understanding about audience behavior and habits towards purchasing, as the audience was no longer physically there. As expected when discussing the advent of data systems, this is when another form of segmentation began to emerge, and one that has eclipsed the variants of traditional segmentation that have come before it. Enter hyper-segmentation, in which already segmented markets can be cut down into even smaller portions, which can be catered equally due to technological advancements in data processing. (Tedlow & Jones, 1993). Now, with a name like that, hyper-segmentation sounds like some device straight out of a galaxy far, far away. And yet, its use in marketing today is far from fictitious science fiction. Instead of the requirement for a marketer to see specific economic or lifestyle choices being the main factors for a segment, they merely are guiding posts to the designation of customers, where the true segmentation comes from even more minute factors. Current hobbies, known jobs, or even their favorite animal; all of these are ways to create a marketing base through hyper-segmentation, as the information can be processed by machines in ways that would otherwise be too time-consuming for the output. One no longer needs to market to a person. Instead, the person provides what is needed to be marketed to in the first place. This is where we begin to see the great merging of marketing systems and information systems, despite their seemingly separate identities. The person themselves is what drives the ability to sell, where a company can know how to sell to them based on what they want on even the most minute details. Information on a customer can be gleaned through their habits as a consumer, which then allows one to understand what they, as a consumer, would be more open to buying (Thibodeau, 2000). This is what is known as online profiling, a concept that is defined by collecting specific data points from users utilizing the online behavior of a potential customer, understanding tastes and interests. It is subset of hyper-segmentation and a direct offshoot of big data, a hybrid mixture all stitched into one as a sort of Frankenstein's marketing monster.

Once a profile was created utilizing information from the online understanding, marketers could then search for specific products that tied into the customer's buying habits directly. No longer would you need to make a product to be marketed, where this essentially flipped the traditional forms of marketing segmentation on their head. Instead of having the product come to you after learning about them, you bring the product to them based on what they want to see. In many ways, this is essentially an upgrade from the original version of market segmentation, an enhancement of the system completely given form by the advancing technology found at the time. Certainly, this source of direct and powerful instigation would not have been possible until technology had made it so. Yet, it was this development of technology that provides an immensely powerful service to all manner of business applications. When an online profile can be gathered about a specific user, there is no need for guesswork, because the guesswork is already filled out by the information provided by the data collected. Information on their behavior can give an understanding of interests and tastes, enjoyment and dislike, wants and fears. But most important of all, it can give an understanding of purchasing desires. It is this understanding that makes online profiling just so powerful in the business world, and a tool that no business would want to be without. Gone are the days of potentially missing out on marketing, for the information on what needs to be marketed can be gathered right then and right there.

It is here that we find ourselves in the modern age, where online profiling has continued to develop and move forward, assisting in creating an effective and efficient landscape for businesses in such a digitally-developed age. Take the example of Walmart Inc., a supermarket conglomerate of American origin. Operating a multitude of supercenters across the country that offer shopping for groceries and other goods, having exploded in size and power from its humble beginnings as local general store. Now a major corporation, the supercenter has been at the forefront when it comes to the use of hyper-segmentation and profiling systems. This specific form of data prediction, referred to by the company as “predictive technology” (Venkatesan, 2021) is an algorithmic system that utilizes previous data points from customer purchases in order to predict the needs and wants of future customers in relation to various specific scenarios. For example, within 2004 the raging superstorm Hurricane Frances was whirling its way towards Florida's Atlantic coast, laying devastation in its wake. Now, for a company that wants to help in this time of crisis, executives and analysts would assume that the most natural items to stock up on would be important necessities for a hurricane. Batteries in case of power outage, bottled water in case of lack of drinking sources, and other importantly staple items of a natural disaster. And yet, Walmart's predictive technology came up with a different solution. Sure, those previously mentioned examples were of decently-high regard, but surprising as it would be, there was no distinct change in the purchase of these items in relation to oncoming disasters. However, there was one specific item that would skyrocket in purchase upon the advent of an impending typhoon:

Kellogg's toaster pastry Pop-Tarts. Specifically, that of the Strawberry-Flavor variety.

Now, before you assume that the predictive technology was short-circuited into catastrophic error by the floodwaters, this specific choice was not out of some random assumption. As chief information officer at Walmart, Linda M. Dillman (2004) expressed this thought in a press release regarding the purchases, noting that instead of relying on what analysists and higher-ups had to say on what to keep in stock, the company could instead "start predicting what's going to happen, instead of waiting for it to happen." The system in question was able to come up with predictions that hadn't even been considered previously, although at first. Trusting the system's finalized data statement, massive orders of Strawberry-Flavored Pop-Tarts were sent to all stores within the radius of the disaster, despite such a strange choice for the number one stocking choice seeming to be absurd.

However, upon the arrival of Hurricane Francis, the absurdity of the statement soon changed to that of awe. Just as the algorithm had predicted with almost uncanny accuracy, Walmart sold immense amounts of toaster pastries in high volume and high frequency, even completely selling out of the extra influx purchased by designation of the data system. The predictive algorithm had predicted the requirements for purchases to an exact level, even going so far as to predict that there would be a slight influx of people buying alcoholic beverages such beer beforehand (Hays, 2004). The machine had predicted the buying habits of the consumers affected by Hurricane Frances to nearly perfect degree, one that shocked the marketing world in the wake of the natural disaster. Using the immense amount of data from Walmart’s internal servers, the predictive technology system was able to come up with a result that maximize potential profits in the week of the oncoming disaster, utilizing information from previous examples of similar events across a multitude of other stores nationwide. It was this collection of information from all manner of shoppers that allowed Walmart to create the perfect stock order and stock their stores accordingly. It should be no surprise that the application of such information has unprecedented uses for marketing segmentation, as well as the use of data in information systems. But this information on customers simply did not appear out of the blue. The information collected to develop the predictive technology that Walmart uses was built and designed from the data points of millions of examples, from which the system can scour to create the most effective order possible. The data about each of these shoppers and their order is far more valuable than any item they sell in the store. For one, the information on this specific desire for Pop-Tarts during a natural disaster allowed them to increase their profits twofold. Even so, when a system developed using this information can create 100% accurate predictions for ordering stock, it really isn't surprising to see why.

With the system of this design built off so many data points, a rather intricate thought begins to develop in relation to this. Of course, there is the understanding for application in profit, but a system that can predict what people will do based on what people have done is nothing that has not been seen in the past. Analysts have spent their entire lives working on understanding the economic changes of the stock market, so why is an algorithm such a big topic? The answer to that might be more apparent than it seems. For the bigger question to be asking, instead of if the algorithm is better, comes from the fact that the algorithm can be better is almost every sense of the word. Large companies will often try to continue and become better, bigger, and more effective for the task at hand, as it is the basis of corporate design. Due to this, Walmart isn't the only one that has their eye on the information of the consumer in order to bolster their business. Take Amazon for example, one of the largest tech firms in the world. A massive conglomerate focusing on the work of e-commerce and product fulfillment, Amazon has built its entire brand around being able to understand the consumer and properly give them what they want. As such, it should be no surprise that Amazon also utilizes forms of predictive technology in their online profiling. However, one of the larger points that makes Amazon stand out in relation to other companies is the often-aggressive forms that it takes in order to obtain this information. For example, Amazon recently purchased the robotic tech company iRobot within 2022, an acquisition that caused quite a stir in many tech-related firms due to the sheer amount of information that could be obtained through such a purchase. The main issue that arose was the statement regarding the use of the device peripherals to enhance Amazon data systems. For those unaware, the iRobot Roomba vacuum itself is able to navigate around the interior of a house, and through this, it can then create a routine based on mapping out the available space. With this data collected, the vacuum can move through the house and clean more efficiently as it was designed to do, becoming better at the task with each scan (Webb, 2021). In a way, the iRobot vacuum is a small microcosm of the process of data collection and data application. The robot collects information about its surroundings, what rooms lead to which areas, and what furniture can be found there. However, in an almost ironic twist, the device designed to collect the data for itself is simply another point of data for an even larger company due to the acquisition previously mentioned. While the information about your house assists the Roomba in completing a task, that very same information is incredibly valuable to marketers, especially Amazon. In a similar vein to Walmart's predictive technological system, there is an innate and powerful understanding that comes with such a collection of information. With an understanding of what a consumer does and where they live, is that much easier to collect information on possible products they could need, as well as their hobbies or jobs, and that much easier for an algorithm to recommend products that they should purchase on Amazon. It is here that we begin to see that there is often a deeper underlying motive to many of these technological advancements. Of course, the assistance that these technological feats provide can certainly improve and marginally bolster the way that tasks are done, but there always seems to be a clause at the end of such a statement. Despite this application of information for use in our everyday lives, the acquisition of information itself is often just as coveted.

Through examining the application of big data and the implication of market segmentation, the curtain has been pulled back to reveal the work behind the scenes that apply within the field of information collection. Like the two halves of yin and yang, each form of segmentation works in tandem with the other. As information is collected, a company can better create segmentation for the customers they are appealing to. And as they continue to segment customers, more information can be collected on the customers that they sell to. It is this continuous cycle that fuels information analysis systems like some gargantuan engine, roaring to life with the intake of information and data as fuel for the fire. But just like any engine that requires fuel, the bigger the engine, the more fuel it needs.

And if you constantly need to keep a massive, churning engine running at full speed, you'll eventually need to start getting fuel from other sources just to keep up.

IV.) The Ugly: Data Security, Artificial Intelligence, and the Future of Digital DNA

> *It is a matter of fact, that AI influences our lives and transforms societies in a variety of ways… For example, AI can be used to control and filter information flows or to exercise automated censorship of content published on social media. Lack of transparency in data collection and its use by algorithms reduces the ability of human users to take fully informed decisions… Therefore, it is of utmost importance to strike the right balance between mitigating the risks and making full use of the advantages that AI can offer, I believe that one of the first steps to build citizens trust in AI and its benefits is to set the rules of the game*. (Bergamini, 2022)

Indeed, these so-called rules of the game are often a point of contention when it comes to the idea of analysis collected, especially with the looming presence of these fuel-hungry data engines just reaches away. Information has become a requirement in order to keep many of these so-called engines up and running. There is only so much data that can be collected from a single source that is willingly provided, such as a Walmart shopper's product choice. With the requirement to constantly obtain new data in order to make bigger and stronger models, it is only natural that this influx of information needs to start coming from somewhere. And this is where the discussion begins to curve towards a more moral obligation towards the information that is obtained. It is here that the application of both big data and market segmentation hit the biggest roadblock of them all. The idea of ensuring that the data collected is only what is needed and allowed is the main focus of protecting the moral obligation of information brokering, which is more commonly known as data security (Franzke et al., 2021). In short, data is valuable in what it represents, so it should be protected from the moral obligation to do so. It should come as no surprise that it is only natural that people wish to keep their information secure. Especially since it was just deliberated how powerful information on people can be for a company to collect for market purposes, there has been a recent surge in that people want to protect this information. This results in a constant struggle, a tug-of-war between two sides, in which one wishes to keep their information to themselves, and the other needs information just to continue functioning.

When it comes to the concept of security, there's a very fine line of which information security teeters on between the brink. On one hand, people in general will always want their information to be secure, and will allow others to ensure that security by any means. However, on the contrary to such a statement, people will also be skeptical of the ability of another to keep their information secure. It is this dichotomy that creates such a powerful and divisive stance on the topic of digital security. There is quite a bit of hubbub regarding how exactly digital security can be assured, when the companies that apply such security measures are just as likely to look into this information as well. Of course, these accusations are not unfounded, as the general populace has quite the reason to be skeptical of digital security. No better is this highlighted than than when it was discovered 2020 that many popular virtual private network services, designed to encrypt a secure connection online, were doing practically the opposite. A large group of secure connection dividers including VPN UFO, Secure VPN, and Rabbit VPN were found to be directly controlled by a singular white-label system, of which the information on over 20 million users was stored and consolidated for collection purposes (Owaida, 2020). This meant that users were being tracked with personally-identifiable data that could then be brokered to other companies despite the purpose of VPN systems as to not track users. When even the options marketing themselves as secure ways to avoid data collection are the ones stealing data on users, there is certainly cause for alarm. In this case, the security of the consumers that placed their information in the hands of noted companies cause these consumers to have their trust broken. It is here that this broken trust creates a rift in what can be truly understood as actual digital security. If the security being provided cannot securely lock down information from being stolen from outsider sources or even the security company itself, then why is there any need for security in the first place? It is this idea that often poisons discussion regarding the concept of a digital self. If this information is so freely available, why is there any need to protect it? Such a laid-back way of looking at the concept is exactly why this concept is just so important. The digital self does not exist in a format where it can be easily obtained. Rather, it is much easier for this information to be given than to be scoured.

This is where a new contender in the technology data system race begins to come out of the shadows, emerging like some ancient machine creaking to life. This metaphor is perhaps more apt than intended, for the emerging power in question is that of the machine learning of artificial intelligence and learning systems. Examples such as ChatGPT, OpenAI, and other learning language models are able to scour immense amounts of data and generate coherent responses when prompted, often with pinpoint accuracy to the request (Bajarin, 2022). This is only possible through the application of big data and information analysis, in which a computer can respond as if it was any person standing right beside you. Truly, the technological feats are impressive, and yet, it is this impressive nature that is cause for alarm. This emerging technology is the reason why digital security is just so impactful in relation to protecting digital DNA as a concept, as it provides some modicum of protection against one of the most data-hungry systems that has emerged to this day. Information on a person is just so important in order to keep data tech running, where it is starting to become seen as more valuable than the person themselves. It is this value that has caused many companies to overstep the boundaries when it comes to the moral obligation to protect data, yet with just how powerful data can be in the right hands, it is not hard to see why. This idea of the value for digital DNA is especially apt when it comes down to the fact that information about a person can be obtained in ways that do not even require the system searching for it. Sometimes, the system can just be given information all on its own. This was the case with the recent leak having occurred through Samsung and their semiconductor division within 2023. Supervisors at the company had implemented use of the artificial intelligence chat system ChatGPT in order to check over source code and ensure quality assurance. However, other users that were accessing the AI became rather surprised to see that prompts to the system began highlighting rather sensitive information, a large amount belonging to Samsung themselves. It was discovered that employees, looking to streamline their job processes, had been uploading sensitive data directly into the AI model to gain responses. As expressed through the media backlash covering the leak, the information shared was nothing short of sensitive: “In one instance, an employee pasted confidential source code into the chat to check for errors. Another employee shared code with ChatGPT and ‘requested code optimization.’ A third, shared a recording of a meeting to convert into notes for a presentation” (Mauran, 2023). This information is now freely available through the entire ChatGTP model, in which the sensitive trade secrets were taken in by the artificial intelligence and internalized through having been given this information by users. Sharing confidential information with artificial intelligence systems only allows it more understanding of this information, as the intelligence itself has no recollection of what makes the information just so sensitive. It is why this data is so incredibly valued when it comes to obtaining such from users, whether that be about the person themselves or the information they know. With users so open to giving away this information, it is no surprise that such digital data has become such a powerful commodity. However, Samsung is not an outlier in this example. Many companies have been incredibly quick to jump on the bandwagon of implementing AI systems into their everyday interactions. On one hand, this is not exactly surprising, as the amount of utility that a learning model can provide is practically innumerable. On the other hand, this continuous rush to implement artificial intelligence systems and data analysis has resulted in rather hasty and sloppy work. In global survey conducted by artificial intelligence broker McKinsey & Company (2021) on the business implementation of AI, 56% of all companies had at least some form of learning model working with company data to assist business operations, with 64% of those running on a public cloud system. Over half of these businesses that use artificial intelligence systems have them open to online access, which provides no security for the information being fed. With the example highlighted previously of such incredibly sensitive information being uploaded into the system with reckless abandon, this is a rather concerning statistic to consider. How can one truly protect the information they are handling as a company if that information is freely being given away? And yet, despite such moral quandaries regarding the use of sensitive user data in these systems, there seems to be no slowing down. In a keynote release from public engagement analyst Amanda Russo of the World Economic Forum (2020), approximately 84% of employers are on track to digitalize the work process within their companies, from remote access all the way to artificial intelligence. As these forms of AI continues to develop, Russo expresses that there will be a “new normal” for learning machines integrating into company work by the year 2025. With such high statistics and such a looming date, it is no wonder that companies are so eager to train new models as fast as possible, with as much information as possible.

It is this constant fetishization of both big data and market segmentation that has caused such gross negligence for the moral obligations of handling information. As data-analysis technology rushes forward in order to create the new and exciting, the information about a person is beginning to become more important than the person themselves. The vast amount of training that AI systems are being given is a rather disturbing thought to consider even with the technological marvel that it presents, as the information provided is beginning to rival that of trained professionals. Multiple AI systems were able to complete the certification requirements needed to become a certified and fully-licensed doctor, being able to score a passing grade on the United States Medical Licensing Exam. In fact, tests completed with both the ChatGPT and Flan-PaLM learning models were able to pass these exams with a consistent 60% score and above, with even higher marks after being trained on relevant medical data (DePeau-Wilson, 2023). Simply by obtaining the data needed for the project at hand, the artificial intelligence was able to score far better than just being trained on basic information. Now, this might seem like a no-brainer. Of course, when machine learning is given information pertaining to the question, it learns better and can answer more accurately. And yet, that is exactly why such information is just so important to the creation and sustaining of these models. More data means more accurate results, and there is only one place to obtain information: Digital DNA.

Perhaps the most potent, and the most eerie concept to consider in regards to your digital DNA is the fact that it is indeed digital. And as with most digital data, it can be replicated, just as easily as a Copy and Paste. While there is ever only one of a single person, the information on a person is just as it intends to be, which is information. And as such, enough information is enough to be able to replicate the person, in a form that often is indistinguishable from the real deal. And for every great and powerful advancement in the field of technology to help us, there is always a leech on the bottom of the big fish. No better is this shown that the recent surge of scamming attempts that have been on the rise, in which family and friends that reach out may not be who one thinks they are. Bad actors, often using freely-available AI systems, have begun taking advantage of this new age of technology wonder. Using autonomous systems to collect information on people such as banking information, living location, names of family, and even simple hobbies that a person engages in, the bad actors were able to create a so-called replica of the person, right down to mimicking their voice through reconstructed audio samples. This false replica could then call up relatives, friends, or even coworkers, acting as an impersonation of uncanny accuracy (Belanger, 2023). Despite sounding as if it was lifted directly from a movie script of The Matrix, this creation of artificial copies is indeed very real, virtually indistinguishable, and powered by the learning models systems and data collection used to give it form. In one such case, the online AI replica impersonated a man's grandson being wrongfully imprisoned and needing bail, of which it was able to respond to all questions prompted by the family with fully-accurate answers when attempting to verify them as being the real person (Verma, 2023). It could be family members hearing the voice of a loved one crying out for assistance, or an employee hearing their supervisor ask for some important documents, none the wiser that a machine on the other side is simply impersonating them. A rather dystopian concept to consider when it comes to moral application of data collection, especially since this mimicking of another is not limited only to sound, but to the digital DNA of that person as well. The inclusion of these examples highlights that artificial intelligence can respond to questions attempting to validate identity with pinpoint accuracy, all from the information it has gathered on the person is attempting to mimic. And although artificial intelligence is not at fault, the technology itself is still what allows these bad actors to utilize such, all with the application of the data at hand. It is this specified conundrum which causes an immense moral implication to the idea of artificial intelligence, in which the information that it provides is often not of its own, but rather mimicked and regurgitated from the data points collected to create it. The ability for such learning models to be able to replicate another person to nearly identical details is certainly amazing from a technological standpoint, but it is due to this continuous surge of technology systems that such cracks have begun to show. Even with such potential ethical concerns, companies have begun to willingly ignore the possible negative applications of this new technology, simply to be the first to create it. They choose to actively ignore moral implications of artificial intelligence, such as the example of the tech giant Microsoft Corporation completely firing the entire AI-Ethics team after gaining acquisition of the ChatGPT intelligence model. At first, the team was reduced in numbers while working to connect the intelligence system to online databases, citing simple corporate restructuring. However, as soon as the system was up and running, Microsoft fully abolished the ethics overseers with the publishing of the New Bing variation of ChatGPT, resulting in a major setback in the ethical implementation of AI systems. (Smith, 2023). What sounds like some cartoonish caricature of corporate ignorance is very much a real example of negligence in the field of AI, in which companies are so focused on moving forward that moral obligations are starting to turn into roadblocks to them. In such a desperate bid to create the new and the better, the moral implication of these systems is beginning to fall on the wayside. It can be said without pushback that data security, artificial intelligence, and other information systems made to handle information have begun to crumble under the weight of trying to become bigger and better. Truly, the technology that is created is remarkable and utterly stunning in what it can do, but that in and of itself is the issue. This remarkable utility comes at the cost of the information being used to uphold it, and many times, this information is far too focused on what can be done instead of what should be done.

V.) Conclusion: Privacy Nihilism and the Future of Data Ethics

> *Crucially, while we often think of innovations as the product of reflective design, technological evolution can be understood in terms of a stochastic process of natural selection coupled with artificial selection by human stakeholders. Consequently, once technological evolution is considered, it becomes clear that social and ethical issues can be incidentally ignored as users adopt an innovation for one reason while neglecting associated trade-offs.* (Schoenherr, 2022)

Now, with all this talk of how important your data is, and how much companies will be willing to sacrifice to obtain it, many thoughts revolving around the concept will be going through your head. Some good, some bad, and most probably being wary of the advent of this powerful new technology. And one of the greatest, and perhaps most telling, is the thought that since all of this information is so easily obtainable and desired by companies, why is there any reason to worry about it? If the idea of privacy and digital DNA is so out there, why is there any need to try and protect it, since it is apparently no longer possible to do so? This concept often comes up in discussion relating to how exactly the digital self applies itself, in the sense that it is often not worth the effort to try and protect your information if the information is just going to be used anyways. Many Americans talking about their habits on data protection would agree with you, some even succinctly expressing that they “haven't changed it much simply because [they] don't feel it's worth the effort” (Kravets, 2015). When it comes down to the thick of it, it can very much seem as if there is no way to avoid such powerful technological advancements. If such large companies do not seem to care, then it almost appears as if there is no way out from such an incessant march of technology. Regardless of how you think it, in a function that is both conscious and unconscious, the result is the same.

You cannot, and should not, fall victim to this thought. This concept of accepting data collection due to a perceived futility has come to be referred to as known as “privacy nihilism,” in which the infrastructure created to collect data has made it seem as if this is impossible to stop (Hartzog, 2018). If there is nothing that can be done to stop the relentless hunger for data, then why even try? Nevertheless, the outcome is still the same, in which the person in question feels that there is an inability to push back against this kind of data collection. It is this line of thought that is rather problematic for the idea of ensuring the ethical upholding of sensitive data and the protection of digital DNA. In many senses, those that are looking to obtain your information will want you to believe that your information is easy to obtain. After all, it is far easier to take something from your grasp if your grasp has been loosened. Privacy nihilism is a very easy thought to subscribe to and is the most damaging form of thought you could take away from this. It begins to directly impact the concept of the digital self as a concept in general, for if people believe that the information about them is going to be collected regardless of what they do, this information is already as good as gone.

But why even bring up this concept if there is such a risk? Although discussing such a concept may have made you wary, confused, unsure, and potentially betrayed when it comes to the information about you in a digital age, that is exactly the sort of mindset that needs to be had when it comes to the concept of digital DNA. A sense of vigilance in knowing about the data-driven world of today, even if it is gritty and unlikable. As information on a person becomes more and more valuable in the modern age, it becomes more and more important to protect this information as if it were the person themselves. Data brokers and information systems grow larger and more complex with their ability to obtain information, which means it becomes imperative that this information is protected just as much as the person it is tied to. Each of these examples highlighted and discussions made are not to give off a sense of hopelessness for the situation, but rather to inspire hope in the sense that although these things are occurring, there are ways that they can be changed. But that change can only come from those that make this technology, and those that use the technology in their everyday lives.

There is a light at the end of the tunnel, even if the current system now may seem bleak and dark. But it is this push to protect the integrity of the user that needs to be taken care of sooner rather than later. With the massive advancements in artificial intelligence and data analysis systems, the information on a person is often just as valuable as any other commodity, and perhaps, it is worth even more as we move forward into a highly technologically-focused age. One must simply be aware that your information as digital DNA is valuable, keen in knowing that there will always be attempts for your data to be collected. It was collected with ancient mathematical tools created thousands of years in the past and will be collected with the advanced AI models of the future, as the desire to collect and understand is an innate part of the human spirit. It is why today, we see such leaps and bounds in technology, where the systems designed to take information in are made so that they can constantly move forward. It is for this reason, the innate desire of mankind to learn, that we must ensure that the moral obligation of data collection is upheld in this modern age. As technology advances in leaps and bounds with truly amazing feats, it is difficult to stop and pull back the reigns. Should we not consider the implications of our value as people in this digital age, there will be a time when we let the reigns slip away, and it will be too late to grab them. Even with every new learning model or data segmentation system that enhances our everyday lives as we move into the world of tomorrow, your digital DNA should never be more valued than the real deal. Simply being aware of your value as a person and of the data you carry is vastly important with the advances in technology made each day, and taking steps to protect it is imperative in the modern digital age that we find ourselves in. If I can teach you anything today, it is not to be afraid of big data, hyper-segmentation, or anything in between. It is that we must be diligent to morally and ethically protect your digital DNA as a concept.

…Because it’s the real you.

The irony is not lost on me, but as my ChatGPT (2023) impersonation succinctly stated at the beginning of this analysis, “you should always try to protect your digital self. After all, what have you got to lose?”

References

Belli, L. (2018). Network self-determination: When building the Internet becomes a right. *IETF Journal*.

Bergamini, D. (2022, November 23). *Speech on artificial intelligence and public functions* [Speech transcript]. La Sapienza University Parliamentary Assembly. https://pace.coe.int/en/pages/daems-ai-public-functions

Burdon, M. (2020). *Digital data collection and information privacy law*. Cambridge University Press.

Campbell. (2019). Synthetic Data: How AI Is Transitioning From Data Consumer to Data Producer... and Why That’s Important. *Computer (Long Beach, Calif.),* 52(10), 89–91. https://doi.org/10.1109/MC.2019.2930097

Chui, M., Hall, B., Singla, A., & Sukharevsky, A. (2021, December 8). *The state of AI in 2021*. McKinsey & Company. Retrieved April 3, 2023, from https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021

DePeau-Wilson, M. (2023, January 19). *AI passes U.S. Medical Licensing Exam*. Medpage Today. Retrieved January 12, 2023, from https://www.medpagetoday.com/special-reports/exclusives/102705

Duke University. (2022, June 5). *Global Business Outlook survey polls*. Duke CFO Business Outlook. Retrieved April 2, 2023, from https://cfosurvey.fuqua.duke.edu/press-release/more-than-80-percent-of-firms-say-they-have-been-hacked/

Franzke, Muis, I. ., Schaefer, M. ., Afd Digital Humanities IT, ICON – Media and Performance Studies, & LS Film televisiegeschiedenis. (2021). Data Ethics Decision Aid (DEDA): a dialogical framework for ethical inquiry of AI and data projects in the Netherlands. *Ethics and Information Technology*, 23(3), 551–567. https://doi.org/10.1007/s10676-020-09577-5

GAB News. (2023, March 8). *Georgetown County computers hacked*. Georgetown And Beyond News. Retrieved April 23, 2023, from https://gabnewsonline.com/georgetown-county-computers-hacked-not-as-bad-as-incident-p6965-90.htm

Gates, B. (2001). *Developer Agility in the Digital Decade* [Speech transcript]. TechEd 2001. https://web.archive.org/web/20120213015238/http://www.microsoft.com/presspass/exec/bill/speeches/2001/06-19teched.aspx (Original work archived February 13th, 2012)

Hartzog, W., & Selinger, E. (2018, October 11). *Stop saying privacy is dead*. Berkman Klein Center for Internet & Society. Retrieved April 2, 2023, from https://cyber.harvard.edu/story/2018-10/stop-saying-privacy-dead

Hays, C. L. (2004, November 14). *What Wal-Mart knows about customers' habits*. The New York Times. Retrieved January 11, 2023, from https://www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html

Kravets, D. (2015, March 16). *Online privacy nihilism runs rampant in US, survey says*. Ars Technica. Retrieved April 1, 2023, from https://arstechnica.com/tech-policy/2015/03/online-privacy-nihilism-runs-rampant-in-us-survey-says/

Mauran, C. (2023, April 6). *Whoops, Samsung Workers accidentally leaked trade secrets via ChatGPT*. Mashable. Retrieved April 11, 2023, from https://mashable.com/article/samsung-chatgpt-leak-details

ChatGPT (2023). *OpenAI*. [Artificial intelligence].

Owaida, A. (2020, July 21). *7 VPN services leaked data of over 20 million users, says report*. WeLiveSecurity. Retrieved April 2, 2023, from https://www.welivesecurity.com/2020/07/20/seven-vpn-services-leaked-data-20million-users-report/

Pride, W. M., Ferrell, O. C., Lukas, B. A., Schembri, S., Niininen, O., & Casidy, R. (2018). *Marketing principles* (3rd ed.). Cengage.

Roh, Heo, G., & Whang, S. E. (2021). A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. *IEEE Transactions on Knowledge and Data Engineering*, *33*(4), 1328–1347. https://doi.org/10.1109/TKDE.2019.2946162

Rudat, M. (2018, May 18). *A history of data collection, storage, and analysis*. GutCheck. Retrieved February 6, 2023, from https://gutcheckit.com/blog/a-history-of-data/

Rudikowa, L., Myslivec, O., Sobolevsky, S., Nenko, A., & Savenkov, I. (2019). The development of a data collection and analysis system based on social network users' data. *Procedia Computer Science*, *156*, 194–203. https://doi.org/10.1016/j.procs.2019.08.195

Russo, A. (2020, October 20). *Recession and automation changes our future of work, but there are jobs coming, report says*. World Economic Forum. Retrieved January 22, 2023, from https://www.weforum.org/press/2020/10/recession-and-automation-changes-our-future-of-work-but-there-are-jobs-coming-report-says-52c5162fce/

Shatby, S. L. (2022, June 1). *The history of data: From Ancient Times to modern day*. 365 Data Science. Retrieved April 17, 2023, from https://365datascience.com/trending/history-of-data/

Schoenherr, J. (2022). *Ethical artificial intelligence from popular to cognitive science : trust in the age of entanglement*. Routledge.

Tedlow, R. S., & Jones, G. (1993). *The rise and fall of mass marketing*. Routledge.

Thibodeau, P. (2000, September 18). *Online profiling*. Computerworld. Retrieved April 8, 2023, from https://www.computerworld.com/article/2597220/online-profiling.html

University of Minnesota Libraries. (2016). *Exploring Business* (P. Edition).

Venkatesan, S. (2021, June 24). *How walmart is using A.I. to make smarter substitutions in online grocery orders*. Walmart Corporate. Retrieved January 11, 2023, from https://corporate.walmart.com/newsroom/2021/06/24/headline-how-walmart-is-using-a-i-to-make-smarter-substitutions-in-online-grocery-orders

Verma, P. (2023, March 10). *They thought loved ones were calling for help. it was an AI scam.* The Washington Post. Retrieved April 19, 2023, from https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/

Walsh, S. (2021, May 25). *50 Google Search Statistics & Facts*. Semrush Blog. Retrieved April 10, 2023, from https://www.semrush.com/blog/google-search-statistics/

Webb, A. (2022, August 5). *Amazon's iRobot deal is really about Roomba mapping your home*. Bloomberg. Retrieved April 11, 2023, from https://www.bloomberg.com/news/articles/2022-08-05/amazon-s-irobot-deal-is-about-roomba-s-data-collection#xj4y7vzkg