Mattia Danese
CS 183: Privacy in the Digital Age
Professor O'Brien
Assignment 1

Company Practices Memo

For this assignment, I chose to download and explore the data that Google has collected on me. I ended up choosing Google because it is an internet juggernaut, so it is nearly impossible for me to avoid interacting with Google, or a service/app provided by Google, the vast majority of the time I use the internet. Being presented with my data and being able to sift through it was a very interesting and honestly surprising experience. For quite some time now, it has become colloquially accepted that companies, especially big-tech and social media companies, track and record your usage and activity, so I thought I was pretty well prepared in regards to the type of data that I would see. That being said, the more data that I saw, especially data from many years ago or about very specific things, the more dystopian and shocking this whole experience became.

The main thing that surprised me was the sheer amount of data that Google is tracking and recording. Of course this is to be expected given how many services (i.e. the Google search engine, Google Drive, Gmail, Google Maps, YouTube, etc.) Google offers, but there is a significant difference between simply imagining the data from these services being recorded and physically seeing just how much Google knows about you. Moreover, being able to see all this data gave merit to the otherwise unsupported yet frequent comments made by many of always being surveillanced, and this was a bit scary. Most, if not all, Google services record exact queries or information (i.e. what was searched on the Google search engine or which video was watched on YouTube); this was expected but still somewhat surprising considering the great amount of queries one person can carry out on only one Google service.

Out of all the Google services, though, the most shocking in terms of data collection were Google Search, YouTube, Google Translate, and Google Maps. The data from the first two services contain every single thing I have looked up on Google and every single video I have watched on YouTube. Every ad I have encountered on either service was also included. While sifting through this data, it was fairly easy to pinpoint (almost exactly) when I started taking a certain course or when I bought a new video game. More surprisingly, however, was how Google Translate recorded every single time I used the service. This type of data was certainly not something I expected to see and it was honestly eerie to be staring at pretty much everything I have ever Google Translated. Similar to the prior two services, it was easy to pinpoint when I started taking an Arabic course or when I went to Spain a couple of years ago solely based on what I was Google Translating and when I was doing so. Lastly, Google Maps, on occasion, recorded specific addresses that have been queried, but mostly recorded when the service was used and a *general area* that was viewed while using the service. Even though physical addresses were mostly absent, my home address still appeared several times and it was extremely unsettling to see my house, Tufts campus, the restaurants I have gone to, or any other place that I have visited be at the exact center of these rather small *general areas*.

Even with all of this information, I cannot say that I learned much about myself because, frankly, this data is mine. I am aware of my interests and internet habits, so seeing soccer videos being watched on YouTube at three in the morning or my endless Google searches about the syntax of a programming language come to no surprise. From the perspective of Google or someone other than myself, however, there is much to learn about me; though, this learning can be done rather efficiently as aggregating all this data makes building a profile on me almost a trivial task. It would not be hard for someone to learn my day to day schedule if presented with

my Google Maps data; it would not be hard for someone to learn my interests or what I am studying if presented with my Google and YouTube search history; my Gmail queries may also give away what kind of email I receive and with who, or what stores, I interact with over email. Needless to say, the task of learning who I am becomes almost insignificant when this amount of detailed data is recorded at all times.

With the amount of data about their users that Google records, to the point where crafting a profile on each user becomes a mere simple task, it comes to no surprise that Google uses this data in many ways and ultimately monetizes it. Namely, I think Google uses the data they record to tailor the ads users see to topics they have a high(er) likelihood of enjoying, and therefore engaging with. An example of when I experienced this was when I was preparing for coding interviews and had done a lot of Google and YouTube searches on best coding interview practices. Consequently, I started receiving numerous ads, primarily on YouTube, of coding interview help services I could buy or subscribe to, which I did not receive prior to my initial searches. I also noticed how the ads Google presented to me depend on my current location; I receive vastly different ads when I am in Boston, New York City, or Italy. Aside from ads, the data Google collects also affects the actual content presented by certain Google services. For example, it is no secret that YouTube has an *algorithm* which recommends personalized videos based on the user. It would come to no surprise if Google uses Google searches, frequent Google Map locations, or Google Shopping activity to help personalize YouTube video recommendations.

Being able to download all of your own data from Google, or any other company for that matter, is great if you are interested in analyzing your activity within Google or familiarizing yourself with what data of yours is actually getting recorded, but nothing more. Requesting to

download your own data from Google is very straightforward and the data comes pre-organized into folders pertaining to a specific Google service. The data itself is also easy to analyze as it is displayed in table format and solely showcases the query, a timestamp, and sometimes other metadata.

That being said, I do not think being able to download your data from Google provides any more value for two main reasons. First, there is simply too much data for someone to realistically sift through all of it. With every Google search, YouTube search, Gmail search, etc. being recorded, the amount of recorded data quickly ramps up, much of which may be repeat queries. In turn, the amount of time needed to go through all of this data becomes much greater than what the average person would most likely want to spend. This is analogous to Terms or Service or Privacy Policy documents being too lengthy and verbose for the average person to actually take the time and read. Secondly, I question the accuracy of the data being presented upon submitting a data download request. Even though a user produces way too much data, across the various Google services and across time, for them to precisely pinpoint if the data they see is wrong, I am more specifically concerned with the possible omission of certain data, whether that be certain instances in known categories (i.e. Google search engine) or publicly unknown categories. This may stem from a legal misunderstanding, but I find it hard to truly trust Google and believe that this is really *all* of my recorded data. Given that companies in the internet and ad-deployment industries have a colorful history of exploiting loopholes in otherwise ethical and legal gray-areas, it may very well be that Google is not being as open and transparent as they are making it out to seem.