Tony Janus

NET-100 Computer Ethics

09/01/2024

Data Mining: An Ethical Conundrum, and The Death of Anonymity

I'm going to approach this assignment from a more personal perspective than I was taught is appropriate for an academic paper. However, I can't think of a more appropriate way to approach it, given the subject matter for this week. The question of whether data mining is an ethical practice, I've come to understand, is a deeply personal one, and is a question which I believe has no right or wrong answer.

Depending on the perspective one has on this topic, and the approach they take to considering its meaning and its implications, a person can come to any number of conclusions, which are mostly all valid. I have a conflicting perspective, personally, and I find that to be deeply discomforting when in regard to something so powerful.

On the one hand, there is legitimate value in collecting data to, for example, advance medical research, which ultimately may lead to both improving the quality of life for those who suffer and extending the lifespans of those who would otherwise die prematurely from illness and disease. From this perspective, it could be argued that we have a moral obligation and even duty to mine this data for the betterment of mankind.

On the other hand, this same medical data in the hands of malicious actors could lead to mass death and disaster. As technology continues its tireless march of progress, is it not

inevitable that enemy nations could obtain this data and develop pathogens targeted to specific genetic lines, for example?

This exact scenario has already been discussed, as nations like China sit on an enormous amount of global genetic data. Upon hearing this, a rumor may come to mind regarding ancestry company 23andMe supposedly selling data to the Chinese government, which has been debunked as originating as satire, according to a Yahoo News fact check. However, as is the case with most rumors, there can be a nugget of truth found in this one if we dig deep enough, though unrelated to the original rumor.

In an article from the Washington Post, it was reported that BGI Group, a Shenzhen-based company, produced laboratories called "Fire-Eye labs," a portable laboratory that was developed to detect coronavirus from genetic fragments. As the article states, these labs could sequence the genetic code for not just the virus, but for humans as well. When confronted, BGI Group insisted they had no access to the genetic data from the labs.

However, the article continues "But U.S. officials note that BGI was picked by Beijing to build and operate the China National GeneBank, a vast and growing government-owned repository that now includes genetic data drawn from millions of people around the world…" and "… a 2021 U.S. intelligence assessment linked the company to the Beijing-directed global effort to obtain even more human DNA, including from the United States."

This situation continues to be debated, but as China has public plans to "become the world's leader in biotechnology by 2035," I believe one would be justified in taking a cautious position regarding China's planned dominance in genetic technology.

Circling back to the practice of data mining in general, there are numerous examples of benevolent uses for the practice. As discussed by Bruce Schneier in *Data and Goliath*, companies like Nest are selling smart devices which collect data on energy usage, emissions, and customer habits, which can all be used to improve the efficiency of our power grid and allow for more environmentally friendly methods of consumer climate control.

He also references Fitbit (now owned by Google) and Jawbone (liquidated in 2017) as companies collecting biodata linked to consumer fitness, diet, and sleeping habits. While this data collection is somewhat invasive, it can lead to the development or more effective fitness programs, opening the industry to more jobs for personal trainers as the level of required proficiency is lowered the more that technology takes some of the burden of "knowing."

However, all this data makes it so that we are more identifiable than ever, regardless of privacy measures taken. Schneier discusses a Stanford University experiment where they tracked the metadata produced by participants' phones for a period of many months. He gives 5 examples who, based purely on phone calls made and received, could be identified as "a multiple sclerosis sufferer, a heart attack victim, a semiautomatic weapons owner, a home marijuana grower, and someone who had an abortion."

This level of identification is already concerning to me, but when used in data correlation practices, or as Schneier puts it, "the linking of identities across different data sets to draw inferences from the combined data," I believe that anonymity in society is entirely a thing of the past.

To answer the assigned question “Can anyone be truly anonymous on the internet?” I would suggest that no, we cannot. In fact, I think it’s become impossible, given any length of time.

If a single mistake can lead to full identification, as was the case for hacker w0rmer, who linked a photo of his girlfriend’s breasts on Twitter, the GPS metadata of which led to his identification as Higinio Ochoa of Australia, I think it’s impossible to maintain anonymity. We are only human beings, meaning we are fallible. We will make mistakes, and this means we will eventually be identifiable no matter how careful we are.

This does bother me, but what bothers me more is the fact that I’m starting to believe it’s <u>not</u> worth the effort to try to develop the technologies to preserve privacy. I’m not sure they exist, or even <u>can</u> exist. If I were to walk away from society and live in the woods, the void I’d leave behind would raise red flags on its own.

Then, using ultra-high resolution and thermal imaging from satellites, combined with accelerometer data taken from my previously owned devices about my gait, fidgeting habits, and other body language, as well as my decision-making preferences and methodology put together from correlated data from dozens or hundreds or even thousands of databases containing profiles of my previous life, I’d be found and identified probably within a week or less.

I think we are past the point of no return. Pandora’s Box was opened long ago. While reading chapter two of the textbook, I was most intrigued by the idea of

Bentham's Panopticon. This was a new term for me, but I instantly recognized it as being descriptive of the world we live in today. The concept makes me consider the idea of free will, and of American freedom and liberty. I wonder if these things even exist anymore.

I think of Fyodor Dostoevsky, who once said, "The best way to keep a prisoner from escaping is to make sure he never knows he's in prison." While Schneier only used Bentham's Panopticon as a metaphor for modern society, I wonder if Dostoevsky would see it as more of a literal description. I think he would. I do.

The more I learn about mass and ubiquitous surveillance, the more I question everything that is presented to me. Are we ever shown the truth of reality? Can we even find it if we tried? I think, perhaps I was alive at a time when we still could, but I'm not so sure that's true anymore. At only two weeks into this class and only three chapters into our textbook, I'm already nearing an existential crisis for which I am not sure there's a solution to.

The concept of anonymity is dead, I am sure of that, and I'm not so sure it can ever be resurrected. Technology has granted us unimaginable power of knowing, and to quote Lord Acton, "Power tends to corrupt and absolute power corrupts absolutely."

Sources:

- The Washington Post, “China’s quest for human genetic data spurs fears of a DNA arms race” - https://www.washingtonpost.com/world/interactive/2023/china-dna-sequencing-bgi-covid/
- Yahoo News, “Fact Check: About That Claim 23andMe Sold Genetic Data to Chinese Government” - https://uk.news.yahoo.com/fact-check-claim-23andme-sold-234200069.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAANa2RT0dvT_xK_r6rr2xSyuSSSzZCH1j4umFhYO3Hikuzi5VszToJymymhOwOlWyXj1XrcJYZstkbJ3o1FcpHVUZ0oUFvF9cCtWKeEFPjGYmpW_ztUA7jfPVrODP-yVeXsGBOnG3asTEhIfZzDRe3kmYpuCXkaJlJM4Yerk5knln
- *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*, Bruce Schneier