VIP3: Big Data

Vandana Anand, Fivos Kavassalis, Daniel Mcdonough

Position: *Big Data will lead to an information divide centralizing wealth, opportunity and control. Big data allows the collection, storage and analysis of datum on a scale never before available. It can be used to identify individuals and groups, and their behavioral patterns, in anonymized data. This places great power in the hands of governments and corporations to suppress and control opposing viewpoints and agendas and is detrimental to the free flow of ideas. Such is detrimental to a well-functioning open society.*

Introduction:

Big Data is a recent buzz-word used widely in the technological age which has derived many definitions over time. For the purposes of this paper we would like to define Big Data as any subset of large volume of data, both structured and unstructured, that inundates any discipline on a day-to-day basis. Big Data has been used as a prescriptive, diagnostic, predictive and descriptive resource for all, if not most, branches of society. It is most used today in a massive amount of sectors such as; Environmental, Social, Economic, Machine, Scientific, Medical, Security and Behavioral. With such a massive amount of data to collect, one must question if the act of collecting such data is even ethical? Do users have a right to privacy when collecting such data? And how does this affect public and private companies in the near future?

Argument One: Privacy & Discrimination

The first argument stems from the topics of privacy and discrimination. Because the government or corporations can collect and sell a large amount of data and information from people, many citizens will have to face the fact of having no privacy. People will have to forgo or compromise their identities. Another point is that the data can provide organizations with bias results and that could affect the decisions they make. Thus, big data can negatively affect our society by placing control and power in the hands of the government and other corporations.

In the age of big data, people will generate numerous amounts of data points based on where they go, what they eat, and what they buy among many other factors that the government can use for example. People are more susceptible to exposure which in turn will

lead to privacy issues, the condition of being free from observation, in this case, by the organizations. Many aspects of privacy rely on informed consent to collect, regulate, and sell a person's private data. big data, though, is a massive resource that can be reused many times usually in ways that were not imagined when it was being collected so informed consent is difficult.[2] In this case, a once well-functioning society can be negatively impacted with big data in terms of freedom. In addition, although there could be an argument that this data is anonymous, the individual's identity could be revealed through relationships between the data. Rebecca Herold, CEO of The Privacy Professor, exclaims, "It will soon become almost impossible to effectively anonymize data in a way that the associated individuals cannot be re-identified".[3] With anonymization out of the picture, there is an even bigger concern for the privacy of individuals and how their vast personal data can be manipulated by corporations.

The decisions organizations make based on detailed information that big data provides can negatively affect people's ability to get employment, loans, or even on a plane if not enforced with care. This introduces the idea of discrimination, the harmful impact on certain individuals and groups. The Federal Trade Commission also highlighted "the need to prevent [big data] from being used to deny low-income communities credit, perpetuate long-standing biases in employment, or exclude underserved communities from other benefits and opportunities".[1] Just because it is date driven, it does not mean Big Data is free of bias. In fact, automated decision-making, a technique used to evaluate the large amounts of data, makes discrimination more difficult to detect no matter how efficient the algorithm is.[1] For example, if data suggests that someone did not get a job, then decision-making techniques will determine and inform the organization that the person should be denied a loan or admission to college. This kind of unfairness against individuals is concerning when it comes to big data. Thus, with the powerful information big data provides, the government and corporations have the power to suppress and control society which is detrimental to the world.

Argument Two: Hinders Competition

One may argue that Big Data allows small businesses to grow faster and keep up with larger companies. This however, has been proven to be a myth time and time again. It is near impossible for new companies access and use data effectively to compete on the global scale our markets are for today.

To obtain good data on must follow the four V's[1] of big data; Volume, Velocity, Veracity, and Variety. Older, larger companies have the advantage of volume and variety in their data. And they are surely able to keep up the speed of updating their data, simply by being in the business longer and having the money to do so[2]. You may ask yourself, "Why can't smaller companies use open-source data?" The sad truth is that open source data collection is usually not of quality, or does not contain the relevant information a young company might need. If a smaller company wants the data it often results in merger or outright purchase of the other company. A great example of this is Facebook and Instagram[3]. It is true however that there are open source tools to analyze and clean large data sets, it is very unlikely however that the users of the tools will be using it correctly as 57%[4] of companies are using big data tools incorrectly. This is tragic in the business world, as many companies are relying on the "Garbage in, Garbage out model"[5] thinking it is completely accurate and failing. It is only those (the large and big companies), that have the money and resources to sustain losses or enhance their collection and analytical strategies. Where the small businesses have little to no safety net to fall back on and end up failing within at least 5 years[6].

Due to the lack of transparency in these analytical strategies, in opens doors for "poorly designed systems and algorithms"[7]. These methods to determine social trees and recommendations that ultimately may give us what we already know what we wanted, but also block us from exploration and experiencing things we do not know we wanted. As mentioned before Big Data is used as a descriptive and predictive[8] tool yet systems are built upon small variety of population data and assume correlation equals causation, which as every statistician knows is faulty basis for action and may lead to discriminatory bias[7]. As previously mentioned these systems have gone wrong before[10], and personally assault their own customer's privacy.

In today's high pace economic center, time is money. Every second spent collecting and analyzing data is critical in a company's life span. For instance, as a college student may I want to make a start up and sell goods. In order to compete on a local or global scale, I need to be able to find my base audience, potentially hire employees or interns, and know when, and where to buy or sell my product[6]. It is no doubt that having data would help in this situation, as these kinds of tasks have been done before. I just need to obtain the data. Seems easy enough right? But I would have to collect a significant amount of data, store it, analyze

it, and hire or be a data scientist to analyze the data that may or may not be accurate in the first place. As a small start up business, it would most certainly fail. Even if I bought the data from an external information source, it would be one of the many taxes hindering the growth of my company which the money and time spent could have been spent advertising or analyzing traditional, more effective statistics[9].

Argument Three: Big Government

With big data, governments can accumulate an enormous amount of information that otherwise would be impossible to gather. From matters of everyday life like the precise coordinates of citizen location or their consumer trends to more complex information, like predictive applications to integrate their real-time health data with the corresponding historical data, there is no denying that "Big Data" has the potential to revolutionize the way we are governed. But what would that actually mean for the people? Ideally, governments and corporations can actually use big data for the betterment of people's lives. They can provide smarter and personalized solutions for each and every one. However, it has been shown that this is not usually the case. On the contrary, governments and governmental agencies can become corrupt and thereby, abuse this almost too-powerful-to-handle, tool.

A great number of legal issues emerge from the rise of big data and government usage, for example, in policing. These issues involve direct or indirect government surveillance and may become a serious civil rights issue in our digital age. Without significant community input new technologies employed by governmental agencies may turn against citizens protection and may endanger democratic values. In the state of California for example, efforts to update legislative protections are underway. The arrival of digital and therefore smarter, law enforcement means is imminent however the fear of totalizing surveillance is real and can lead to overreach. Civilians and governments must respond to these concerns before the relevant technology is implemented. Without regulation the Internet of things may become the "Wild West of the Internet." [4]

Even in western traditional democratic societies, the use of big data can tempt governments to use them and abuse their power towards citizens' rights. The situation is much more extreme in totalitarian regimes. In the instance of China, big data can be used to enforce traditional abusive tendencies, such as the "social credit system". This system aims at registering not only financial credit-worthiness of citizens and companies but also their social and possibly political behavior. In other words, it is a system of digital social control. In democracies, laws limit what governments can get their hands on but such protections do not exist in China. The national-security law and the new cyber security law give the government unrestricted access to almost all personal data. In addition,

although big data systems elsewhere are not necessarily designed for social control, in China the contrary is in effect and these instruments of social control are used for political purposes. When all the elements will be ready, China may evolve into the first digital totalitarian state.

Another example that comes to mind especially due to recent events is the one of North Korea. This country discriminates against individuals and their families on political grounds in key areas such as employment, residence, and schooling through a system of socio-political classification groups people into “loyal,” “wavering,” or “hostile” classes. This classification continues to enable the government to privilege or disadvantage people based largely on family background, personal performance, and perceived political loyalty. One can easily imagine how big data can enhance the role of such a totalitarian state that already has the background infrastructure for such discriminatory practices.

Conclusion:

Big Data will lead to an information divide centralizing wealth, opportunity and control. Its ability to be used unethically to identify individuals, groups, and their behavioral patterns places substantial power in the hands of governments and corporations. It allows organizations to suppress and control opposing viewpoints, agendas and is overall detrimental to the free flow of ideas and increases inequality across nations. Thus, a once well-functioning open society would be falling with all private information becoming public, monopolizing data collection/analysis, and mass surveillance throughout the nation.

References:

Argument 1:

1. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf
2. https://epic.org/privacy/big-data/
3. https://www.csoonline.com/article/2855641/privacy/the-5-worst-big-data-privacy-risks-and-how-to-guard-against-them.html

Argument 2:

1. http://acco.gencat.cat/web/.content/80_acco/documents/arxius/actuacions/Eco-Dades-i-Competencia-ACCO-angles.pdf

2. http://www.foxbusiness.com/features/2014/06/30/google-capital-leads-110m-round-in-big-data-startup-mapr.html
3. http://www.the-social-pro.com/big-data-facebook-2/
4. https://www.cio.com/article/3003538/big-data/study-reveals-that-most-companies-are-failing-at-big-data.html
5. http://digitalforward.blogspot.com/2015/09/big-data-shit-in-shit-out.html
6. https://neilpatel.com/blog/things-startup-needs-stop-immediately/
7. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf
8. http://dataconomy.com/2015/06/the-four-essentials-vs-for-a-big-data-analytics-platform/
9. https://www.nbcnews.com/news/other/nsa-program-stopped-no-terror-attacks-says-white-house-panel-f2D11783588
10. https://www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/&refURL=https://www.google.com/&referrer=https://www.google.com/
11. http://metro.co.uk/2017/06/29/how-snapchat-knows-youre-sleeping-driving-flying-and-listening-to-music-6744109/

Argument 3:

1.https://www.economist.com/news/briefing/21711902-worrying-implications-its-social-credit-project-china-invents-digital-totalitarian
2.https://www.hrw.org/world-report/2017/country-chapters/north-korea
3.http://www.oregonlive.com/today/index.ssf/2013/06/government_surveillance_is_big.html
4.https://today.law.harvard.edu/feature/new-age-surveillance/