

Lukas Svedberg – Digital and Technological solutions: Discuss how data analytics can be used to influence people.

Throughout history, humans have been shown to be extremely susceptible to having their personality, decisions and beliefs influenced by others. With the ever-more prominent role our digital world plays in our daily lives, corporations can use ‘big data’ to sway populations to serve companies’ needs with more cost-efficiency, accuracy, and success than ever before.

From rudimentary markings by our Palaeolithic ancestors in 19,000 BC to the modern day, the knowledge data insights have provided have proven to be vital for societal progression. In recent years, our world’s total data storage has increased exponentially; ‘There were 5 exabytes of information created between the dawn of civilisation through 2003, but that much information is created every two days,’ remarked Eric Schmidt, former CEO of Google. The sheer influence data has over our modern-day society makes incorporating morality and ethics concerning privacy necessary to ensure users are safe from exploitation. In 2013, Cambridge Analytica harvested the data of many Facebook accounts without user consent, so an ensemble model could analyse these digital footprints to create a psychological profile for over 80 million Americans [1]. This information played an vital role in the 2016 US election, where visualisation tools like heat maps provided by Cambridge Analytica allowed the Republican Party to deliver micro-targeted advertising tailored to sub-groups of the population. Although manipulation of the political landscape has always been the primary focus of all political campaigns, using personalised data from users allows parties to exploit individuals’ personality traits and preferences to influence the course of political history.

Data concerning users has also been exploited by Amazon’s price determining algorithms [2]. This algorithm uses data concerning user spending habits to allow Amazon to alter the price of items to make the most profit, known as price surging or dynamic pricing. Additionally, a similar tactic is

found in Uber's price determining algorithm [3], where Uber incorporates data such as time, traffic density and the customer's phone battery to determine what price would be the most effective for generating profit. These forms of dynamic pricing pose many questions on the ethics of how consumers' own data can be used to discriminate against them based on controllable and non-controllable factors, and whether this economic model is too predatory towards consumers.

Every decade, US legislators of each state modify the boundaries of these electoral districts to reduce geographical malapportionment and eliminate bias in electoral district boundaries. However, these legislators can abuse their point of influence by 'gerrymandering'; unfairly distributing these electoral boundaries through 'packing' or 'cracking' to favour their political standpoint. As the victor is determined based on the greatest number of electoral districts voting for a specific candidate, rather than a pure 'first past the post' system, creating an ideal environment to influence election results. Although gerrymandering has existed throughout US political history since 1811, technology has transformed the current political environment through the analysis of large amounts of data from the American voter file, allowing algorithms to be much more accurate, effective, and subtle in their endeavours. In 2010, the Republican party established the 'redistributing majority project,' which specifically focused on modifying boundaries in politically divided states, leading to a Republican seat majority in Congress despite the Democratic party [4] receiving over a million more votes than the Republican in 2012. Although this political manipulation provides short-term gain for political parties, this misuse of personal data to influence voting results has resulted in gridlock and a more politically divided nation.

In data analytics, laws about data science are the decision of the individual country or of unions like the European Union, whose General Data Protection Regulation incorporate the most stringent privacy and security laws in the world. The lack of an overall international body has resulted in unethical use of data in particular countries due to the lack of protection and privacy for its users. In

China, cybersecurity laws compel all companies operating there to give the Chinese government access to company data, which can be used for military or intelligence purposes. Whilst this wealth of personal data led to a highly effective track and trace system during the COVID-19 pandemic, this data is used by China's authoritarian regime to control its population. The Chinese Government's overreach is seen in mass censorship on news and social media to increase governmental obedience. Moreover, data analytics has been fundamental to the production of a social credit system, which imports data from personal financial and health records, social media usage and your associates with to determine your access to government provisions.

Search engines such as Google and Bing have shaped how our society functions today. Their influence and success are a result of their immense popularity and utility. PageRank algorithms are implemented by all search engines, which utilise data and metadata sent from websites, which are ranked based on relevance to user inputs for personalised searches. The effectiveness of these algorithms has revolutionised knowledge transfer, with over 8.5 billion searches per day on Google alone. Although these systems allow us to gain information at an unprecedented rate, the limitless supply of information stored may reduce memory capacity, with people 50% more likely to forget information if they know it is readily available on the internet [5]. To increase user retention, social media algorithms group users with similar political and social views together to give users more favourable interactions. However, this can result in 'echo chambers', where groups of like-minded people only talk to those with similar beliefs, which can potentially foster dangerous views and actions. This has led to an increasing number of radicalisations, where in 2017 over 83% of those charged with extremist offences had been subject to some degree of internet radicalisation compared to 35% in 2009, correlating with the increase in average internet and social media usage [6].

Machine learning algorithm's ability to analyse data almost instantaneously allows these algorithms to have a profound impact on many fields. However, data input into these deep learning algorithms need

to be free from bias to ensure that the results are ethical, due to implicit bias affecting the fairness of machine learning. This has been seen in Watson's IBM [7], a tool developed to diagnose cancer, which used hypothetical data containing doctors' innate biases, making the program far less accurate compared to an individual doctor's diagnosis. Algorithmic data also reduces individuality in samples, which means limited information given to algorithms may lead way to unfair judgement and discrimination based on certain factors. This problem was encountered in the 2019 A-level Ofqual algorithm for predicted student grades [8], where students from disadvantaged backgrounds were 25% more likely to be downgraded than those from higher socioeconomic backgrounds. A further example of algorithmic discrimination is found in Amazon's Artificial Intelligence (AI) Recruitment Program [9], where innate bias gave male candidates an advantage compared to female candidates upon application. Since algorithms make decisions using previous tests, slight deviations in fairness perpetuate rapidly, creating systems which oppose the ideals upon which the algorithms were founded upon. Consequently, this makes data validation a priority among data scientists and programmers for effective AI.

Data analytics are a fundamental of business intelligence because data is a requirement to make informed and safe business decisions. This allows companies to successfully incorporate strategic change, to appeal to the greatest number of consumers. Data analytics is used by the McDonalds corporation [10], who employ data algorithms which take in parameters such as weather predictions, local traffic, and the time of day, personalising the drive-thru menu for the user. Similarly, algorithms also play a crucial role in the Starbucks business model [11]. The company uses the programme Atlas, a geographical information system which provides Starbucks with data on consumer demographics, population density and competing businesses to determine the economic viability of opening stores in different locations. Although machine learning algorithms can make rough predictions about the future, algorithms can only use parameters input by the user to determine decisions they should make. A lack of foresight was seen in the release of Microsoft's AI chatbot *Tay* in 2016 [12], which interacted and mimicked other users on Twitter, where online trolls manipulated the chatbots machine

learning technology to influence the bot to cite hate speech. The failure of *Tay* highlights the potential danger AI algorithms can cause if they are not properly maintained. With an increase in popularity and demand for AI, with 55% of companies accelerated their AI strategy in 2020, we need to maintain these algorithms so similar instances to *Tay* don't arise in more essential projects, both in business and in government institutions.

Data analytics have been fundamental for allowing people to adapt, businesses to thrive and society to develop. However, the ethical and moral implications of data analytics can be questionable, particularly where data analytics can manipulate consumers and control populations. Peter Sondergaard, ex-vice president of Gartner research stated that 'Information is the oil of the 21<sup>st</sup> century, and analytics is the combustion engine.' I believe that the sheer influence data analytics has over society does allow for a powerful resource, but the possibility of data analytics misuse is a problem not to be underestimated. Consequently, it is imperative that ethical frameworks are implemented by companies and governments to avoid potential destruction.

Word Count: 1533

## References

[1] = Alex Hern 2018 - Cambridge Analytica: How did it turn clicks into votes? (The Guardian)

[2] = What is Amazons dynamic pricing Strategy? (reactev)

[3] = Anubhav Pattnaik – How does uber do Surge Pricing using Location Data (medium)

[4] = Azavea Redistricting 2010 - A national study

[5] = Betsy Sparrow, Jenny Liu, Daniel M Wegner- Google Effects on memory: Cognitive Consequences of having information at our fingertips

Lukas Svedberg

[6] = HM Prison and Probation service: Exploring the role of the Internet in radicalisation and offending of convicted extremists

[7] = Sandeep Konam - Where did IBM go wrong with Watson Health (Quartz)

[8] = Pahlmela Duncan, Niamh McIntyre, Rhi Storer, Cath Levett 2020 -Who won and who lost when a-levels meet the algorithm? (The Guardian.)

[9] = Dastins J 2018 - Amazon scraps secret AI recruiting tool that showed bias against women (Reuters.)

[10] = Ryan Owen – Artificial Intelligence at McDonald’s – Two Current use cases (Emerj)

[11] = Max Pakapol – The Perfect blend: Starbucks and data analytics (d3.harvard)

[12] = Gina Neff, Peter Nagy - Talking to Bots: Symbiotic Agency and the Case of Tay