Challenges to society by dig data and algorithms

In recent years, the development of Internet, smart cars, wearable devices and Cloud technology, have led to the rapid growth of data in almost all industry and business areas. Big data has rapidly developed into a hot topic which had attracted extensive attention from academia, industry, and also governments internationally. With this comes major impacts on society- from how individuals shop, interact socially, obtain information to how large companies do business in terms of consumer research, recruitment and policy. Is this just the new convention in society? An algorithm is a series of instructions, like a flowchart or a food recipe, which can be followed to solve a specific problem. Surely technological advancements can only mean progression and with this social justice? There is evidence to suggest otherwise, with some critics such as Cathy O' Neil (2016) stating they provide " a model that contributes to a toxic cycle and helps to sustain it." In terms of data produced via our digital lives – does anyone really collect the private information that we unwittingly disseminate, or is Big Brother really and truly out there? Is there an ethical obligation on digital companies to be more explicit about what information they collect and what they do with it, now, or in the future? Should we change how we behave to protect our privacy? Or has society adapted to their presence so much that there are no alternatives, are we " locked in" to their usage?

Algorithms are routinely employed in the United States criminal justice system to predict the likelihood that a prisoner will reoffend after release. The Level of Service Inventory Revised (LSIR) questionnaire uses a number of socioeconomic factors in its calculations, such as social class, level of education, place of residence. How such factors are used in the algorithm is, according to O’Neil (2016), very opaque. Prisoners are unaware of what their answers will implicate. The algorithm appears to disadvantage those from a poor socio-economic background. For example, one of the questions asks about close contacts and their history with law enforcement: for those in ghetto areas , it is almost inevitable that there is some previous involvement. These responders are flagged as high-risk, not necessarily because of factors that actually contribute to increased recidivism, leading to the “ pernicious feedback loop” of which O’Neil (2016) speaks. Although it is argued by those that support the method that it will decrease the sentencing lengths for the non-threatening criminals, it can pre-condemn the already disadvantaged within society. The risk scores are seen by the sentencing judges and guide them in their decisions. Furthermore, prisoners do not know their risk scores as there is no transparency and prisoner officials do not reveal the purposes of the LSIR questionnaire. The apparent aim of such a process is to “ bring even-handedness and efficiency to the criminal justice system’ (O’Neil 2016). The use of this recidivism model speeds up the sentencing process for the judges and saves the tax payer money, in a society where the criminal justice system costs eighty million per year. It is understandable, but, is it worth it if people are being discriminated against because of their social background? So, since this model saves a substantial amount of time and money and prisoners do not share the same rights as the rest of society will this ever change? Probably not. However, the appeared effectiveness of these “ weapons of maths destruction” within the prison system, as argued by O’Neil (2016) could lead to the spread of these mathematical models throughout the economy and society, leaving the public as collateral damage.

Algorithms have been used in recruitment utilised mainly to speed up the process; reducing initial applicant numbers to more manageable amounts without human intervention. Companies use intelligent and personality tests to decrease the number of applicants in moving to the interview process. However, as O’Neil (2016) puts it, these tests have been known to be discriminatory. In Griggs vs Duke Power Co. (1971), the Duke Power Company used intelligent tests in their recruitment process, proxies which were found to be discriminatory and illegal by the Supreme Court. Subsequently, intelligent tests were replaced with personality tests, which have been identified as poor predictors of job performance (Murphy, 2005). Their aim is not to determine the best employees for the job but to exclude undesirable candidates, thus saving the companies time and money. O’Neil (2016) argues that personality tests could be valuable if used in an appropriate way, for example, with the purpose of educating workers on their behaviours, or to improve communication and teamwork. Supporters of these tests claim that no one answer will disqualify an applicant but the lack of transparency is a problem: applicants do not know what patterns are being measured or why, and do not receive feedback. Kyle Blehm’s story exemplifies the challenges with such algorithms. Blehm, a high-academic achiever elected to break from his college studies due to a new diagnosis of bipolar disease. His attempts to pass multiple entry tests in low-income roles were unsuccessful, leading to his father, a lawyer, filing a class action suit against the companies for discriminating against his son due to his mental health. Most people applying for such roles do not have the resources to do this however; the “ pernicious feedback loop” manifests again. Disadvantaged work-seekers have fewer employment choices so must acquiesce to a company’s requirements. People from lower socio-economic groups have less education and higher health issues, all of which can be identified in these tests in ways which would be difficult to ascertain without the algorithms, thus preventing progression and perpetuating inequality.

One of the ways that algorithms perpetuate social inequality is by becoming automated versions of human bias. St. George’s Hospital, London used a data process , the algorithm for which was formed from rules applied to previous applicants, which when manually reviewed was subject to unconscious bias. For example, CVs with grammatical errors or names similar to those of unsuccessful applicants were discarded in the manual shortlisting: this rule was also applied in the algorithms. Thus, a major challenge is that an algorithm may have the characteristics of the person who designs and inputs the information. In 1988 the medical school was found guilty of racial and gender inequality in their recruitment process. The school then removed the gender, race, and geographical proxies and only used relevant medical education data. As O’Neil (2016) puts it, the person’s own values and desires influence our choices. Whoever builds or designs these algorithms have an objective and unfortunately those opinions and objectives get rooted into the algorithm. “ Mathwashing” (Woods, 2016) is a term coined by data scientist Fred Beneson who makes the point that just because they use math does not discount any inherent subjectivity. Cathy O’Neil has the same belief; “ Big data doesn’t eliminate bias, we’re just camouflaging it with technology” (O’Neil, 2016). Thus, algorithms are only as good as the information used in their design. Human bias built big data, Du Gay and Pryke (2002) state that “ accounting tools… do not simply aid the measurement of economic activity, the shape the reality they measure”. The challenge here then is not just that these tests perpetuate prejudicial judgments but firstly that they can do this on a much larger scale than when previously done manually; secondly the disingenuousness that a mathematical test must be objective means they are accepted without resistance in most cases.

In the last ten years the introduction of new software (Blackboard, Moodle, and ClassDojo) for individualised learning in schools has produced numerous opportunities for developing the education process while also generating multiple legitimate concerns. The compiling of student profiles and information can later be sold to data brokers, future employers, and universities (Kitchin, 2014). The good news about algorithms and big data in education is that it can be advantageous for personalising teaching, in particular within the e-learning industry and remote learning (Fernández, 2014). On the converse, however, the rise of algorithms and big data bring many ethical issues. Companies like Google have increased the services they offer to include email, document storage, e documents, news, web browsing, social networking and anything else that might interest their users. Such companies gain access to more personal data, which they collect, store, and cross-reference (Boyd and Crawford, 2012). Information that is accessible to the public, is assembled from different sources into a comprehensive profile, thus, creating a revealing portrait of a person. Although everyone in society use these devices and application, students cannot progress through their course without using these technologies, they are obligated to submit assignments via their account, and use ecards in library etc. Student's information may be used against when applying to continue their education or when applying for jobs. Information may be interpreted negatively, for example, if a student has many library fines that have yet to be paid, or spent more time in the campus bar than in the library, these can give negative perceptions about the person's character, when in fact they were engaging in study groups in the campus bar, or that their overdue library fine is being disputed or was an irregularity. All of this information about a person can have a ripple affect going into their adult life, all of which was deciphered without human intervention.

Before the rise of Big Data, consumers were generally exposed to the same advertising and consumer messages, via a standard set of media- posters, newspapers, television, etc. However, Big Data has changed the rules. Individuals are treated differently based on their metadata, for example their browsing history, online shopping habits, or the types of articles they read in electronic magazines and newspapers. The first issue, is that this creates a problem of awareness, salience, and consent. Research undertaken by Madden and Rainie (2015) found that 50% would like to block online advertisers from saving records of their activity. Is it acceptable that users are unaware that they are being specifically targeted, is this a problem, does this lead to unhealthy consumerist behaviours? is it such a bad thing that you will be targeted to buy something you might want or need? secondly, is the threat that users online privacy is compromised? Although consent is given, as Fred Cate (2013) argues, the reality is nobody reads the convoluted policy documents. More importantly, it could be considered that consent is not informed since they cannot predict the future uses of their personal data, Schonberger (2013) proposes. It is this secondary use of information that has become the issue in relation to privacy laws. Companies such as Google are unlikely to embark on a backdated crusade to obtain permission from users to use their previous search history to gather information: it would be both costly and time-consuming.

Despite these issues the majority of citizens continue to use data-collecting services such as Google or Facebook without appropriately protecting their privacy by means of proxy servers, encryption, or any other technical measures. As Maddie and Rainie (2015) identified, only 7% of adults in the U. S. A. were confident that their records would remain private and secure with online advertisers. This creates another issue, the issue of the ‘ lock in’ effect, relating to the fact that in order to keep up with society you have to have an online presence. Users remain loyal even though their information has been compromised. and they often don’t have the ability to protect their privacy (Strater and Lipford, 2008). Furthermore, the consequences of Big Data collecting private information can have much more far-reaching consequences than simply bombarding one with personalised advertisements. It can prevent people from obtaining mortgages, bank loans, insurance, and housing (O’Neil, 2016). For example, if someone has been overdrawn continuously finance institutions will see this and perceive them as high risk where the fact could be that they are constantly using account for work expenses which take time to be reimbursed. Furthermore, if someone’s online shopping habits corelate to an unhealthy diet and can be seen as a health risk they may not be successful in getting health insurance or their premiums may be too costly. Schronberger (2013) describes this excellently as a “ dictatorship of data”.

The scale of mass surveillance by governments around the world through, for example, the collection of metadata and the monitoring of social media shocked everyone in the post- Snowden society. Data mining and surveillance techniques such as images, videos, and interactive maps as well as associated metadata such as geolocation information and time and date stamps are all used within the policing and national security backgrounds. These raise serious human right rights concerns about the ability of modern states to monitor, dominate and control their citizens (Haraszti et al, 2010). As Schonberger (2013) states, this is reminiscent of the days of the Stazi in East Germany- thousands were employed to monitor the public, collecting multiple miles of documentation. In our modern society an equal amount of data is gathered on each individual. This is collected when we use our phones, devices, debit cards, travel cards, and social security cards, all of which are near impossible to avoid using in the modern world (Bernal, 2016). Unlike the Stazi, companies such as Amazon, Google etc. are not law enforcement agencies; but there is evidence to suggest that government security agencies may be gathering our personal information. The National Security Agency (NSA) are alleged by the Washington Post (Priest and Larkin, 2010) to intercept and store 1. 7 billion communications every day. This indicates that changed immensely since the days of the Stazi; with the rise of technology and big data it has become far easier to compile information on the public. This information is gathered not to investigate in the present but rather to have a head start when or if someone becomes of interest to law enforcement or government agencies, O’Neil (2016) gives the example of PredPol, a predictive policing tool. As put by Schonberger (2013), the danger of big data will shift from ‘ privacy to probability.’ For example, predicting crime in advance of it happening, using algorithms to predict which areas for police to patrol. This is reminiscent of O’Neil’s (2016) idea of the “ pernicious feedback loop” where the already discriminated become more targeted. There are other results that can occur from this too. Does digital surveillance mean we are heading towards a world of constant surveillance and the auto-policing of Orwell’s 1984? Due to the constant surveillance, the public may change their behaviours. This phenomenon has been described by Tijmen Schep (2018) with the following quote-“ like oils leads to global warming, data leads to social cooling”. “ social cooling” He states this is a concept relating to the negative effects of living within a “ reputation economy”. This brings up similarities with Focualt’s panopticism (1979) and can be defined as, “ a type of power that is applied to individuals in the form of continuous individual supervision”. The result of “ social cooling’ is making society better behaved (at least where they can be observed) but the argument is that it is making people less human. Society’s data is compiled and categorised into scores and compared against the information of others to form a pattern and guess the details of a person. Here are some of the categories used, religious views, agreeableness, sexual orientation, and even ‘ had abortion’ or rape victim to name just a few. It is considered that these factors may restrict people in taking risks or expressing their views on political or social issues and hence cool down society.

One of the major difficulties of big data and algorithms is the obvious inequalities they foster; inequalities that may have already existed but have been made easier and occur on a much wider scale when employed via technological advances. It appears that those who are disadvantaged in society- prisoners, low-income workers- are particularly vulnerable to the formulas which seek to effect cost savings for large institutions. While such biases may always have existed, technology has facilitated a situation whereby they are considered more acceptable under the mistaken premise that science is non-discriminatory. In some cases, the fact that such algorithms are being used for particular outcomes is unclear, bringing up issues around the ethics of such methods. One of the bigger challenges of Big Data is how to manage digital footprint – everywhere the public go they leave behind a distinct picture of their personality and characteristics. It is considered that societal shifts may occur where people will seek to manage their privacy by changing their publicly recordable behaviours- a mammoth task considering all the different aspects of our lives which can now be monitored. There are few legal protections in place- do governments need to legislate for this or are they complicit in the data-gathering to monitor their citizens? It would seem that the technological progress which has occurred evolved faster than society’s ability to adapt accordingly and there is significant danger that the “ information superhighway” will lead us on a journey to a world where privacy is a public commodity.

References –

- Bernal, P., 2016. Data gathering, surveillance and human rights: recasting the debate. *Journal of Cyber Policy* , *1* (2), pp. 243-264.

- Boyd, D. and Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* , *15* (5), pp. 662-679.
- Cate, F. H., Cullen, P. and Mayer-Schonberger, V., 2013. Data protection principles for the 21st century.
- Fernández, A., Peralta, D., Benítez, J. M. and Herrera, F., 2014. E-learning and educational data mining in cloud computing: an overview. *International Journal of Learning Technology* , *9* (1), pp. 25-52.
- Foucault, M., 1979. *Panopticism* . na.
- *Griggs v. Duke Power Co.* , 401 U. S. 424, 91 S. Ct. 849, 28 L. Ed. 2d 158 (1971).
- Haraszti, M., Roberts, H., Villeneuve, N., Zuckerman, E. and Maclay, C., 2010. *Access controlled: The shaping of power, rights, and rule in cyberspace* . Mit Press.
- Joel Reidenberg et al., *Privacy and Cloud Computing in Public Schools* , CTR. L. & INFO. POL’Y
- John Walker, S., 2014. Big data: A revolution that will transform how we live, work, and think.
- Kitchin, R., 2014. *The data revolution: Big data, open data, data infrastructures and their consequences* . Sage.
- Labrinidis, A. and Jagadish, H. V., 2012. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment* , *5* (12), pp. 2032-2033.
- Mary Madden & Lee Rainie, *Americans’ Attitudes About Privacy, Security and Surveillance* , PEW RES. CTR. 7 (May 20, 2015)

- Mayer-Schonberger, V. and Cukier, K., 2013. *Big data: the essential guide to work, life and learning in the age of insight* . Hachette UK.
- Murphy, K. R., 2005. Why don’t measures of broad dimensions of personality perform better as predictors of job performance?. *Human Performance* , *18* (4), pp. 343-357.
- Privacy-
- Michael, K. and Miller, K. W., 2013. Big data: New opportunities and new challenges [guest editors’ introduction]. *Computer* , *46* (6), pp. 22-24.
- O’Neil, C., 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy* . Broadway Books.
- Penney, J. W., 2016. Chilling effects: Online surveillance and Wikipedia use. *Berkeley Tech. LJ* , *31* , p. 117.
- Priest, D. and Arkin, W. M., 2010. Top Secret America. *Washington Post* , *19* .
- Strater, K. and Lipford, H. R., 2008, September. Strategies and struggles with privacy in an online social networking community. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1* (pp. 111-119). British Computer Society.
- Reno, J., 2012. Big Data, Little Privacy. *CA Technology Exchange* , p. 24.
- Woods, T. (2016). ‘ Mathwashing,’ Facebook and the zeitgeist of data worship. *Technically Brooklyn* . Accessed online https://technical. ly/brooklyn/2016/06/08/fred-benenson- mathwashing-facebook-data-worship/ December 5th, 2018.