At the Mercy of Prediction in the Age of Predictive Models and Scoring

Stanley Greenstein

1 Introduction

It is in the nature of technology to progress and develop at the speed of light and this is reflected in the manner in which the digital environment and its associated applications are playing an increasingly important role in the manner we communicate as well as how we choose to document our lives in public. And as we spend more time in this digital environment, more and more of our every-day actions are reduced to data – data that can be recorded, data that can be stored, data that can be analysed and data that can reveal the most intimate of things about us that we did not realize ourselves. This mass of data that is increasing every second is made up of both content data as well as meta-data and is commonly referred to as 'big data'. It is all these individual data points concerning us and forming a part of big data that can be inter-connected to form what is sometimes referred to as our 'data shadows', i.e. a digital representation of who we are.

As our lives are reduced to data, technologies are being developed using mathematics, statistics, machine learning and artificial intelligence to extract the knowledge that is hidden in big data. The human brain does not possess the capacity to identify this hidden knowledge, but it is no contest for the problem-solving algorithms. It is undisputed that the advantages to be gained from the exploitation of the knowledge held within big data are numerous, however, their use also entails the exposure of individuals and society to risks and vulnerabilities, which have been somewhat neglected until now. The reason for these new risks is that this hidden knowledge gained from big data is increasingly being incorporated into decision-making systems, which lack transparency and which are relied upon to make important decisions about us humans in order to increase effectiveness and also facilitate social control. At the core of the problem is the fact that these decisions are not an exact science – rather, they are for the most part based on prediction!

The goal with the use of these technologies is prediction. The more one can predict, the more one can control and the more one can control, the more effective one can be, and the more effective one can be, the greater one's margins of profit and effectiveness. However, the blind striving after effectivity and profitability brings with it certain risks to society and a central question is whether we as a society are prepared to stand at the mercy of prediction in return for the effectiveness and profitability that these new technologies promises?

One would like to argue that the current state of affairs requires a serious discussion and consideration concerning what kind of technologically orientated society we desire. However, considering the extent to which these technologies are already being used in society to judge and make decisions concerning human beings, it would seem that we have already accepted the existence and use of these technologies without actually making a conscious decision regarding their application and without a democratic debate. If it is not too late – and it is argued that it is not yet too late – then what is required is a discussion and public debate concerning the use of predictive technologies and scoring embedded into decision-making systems and how great a degree of human agency people are willing to acquiesce to technology, in return for the benefits that these technologies bring. As society's reliance on decision-making systems increases, a concern is whether, in the near future, humans will even be in a position to have a say in the nature of the fabric of society, human agency having slowly been eroded as digital outputs determine the basis for social steering.

2 The Technology

A technology that is currently being employed in conjunction with big data and which analyses big data using machine learning techniques is that of 'predictive modelling'. Predictive models are increasingly being used by commercial actors as well as public authorities in order to facilitate their interaction with their clients or potential clients and citizens in the digital environment.

At the core of these predictive models is the mathematical algorithm, which has been described as, '[a] process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer'.[1] Predictive modelling can be described as a practical application of machine learning whereby algorithms are used to identify patterns in data, after which these patterns are transposed into rules that are in turn placed in models that form the basis of decision-making systems. Basically, the algorithm examines data and finds correlations or links between those data. The unique aspect of this process is the manner in which a mathematical weight is given to the correlations between the data points. It is this process that is beyond the mental capacity of human beings and what makes this 'knowledge' invisible to people. A consequence of this is the increased trust in these systems – trust being the inability to comprehend the technology behind these systems and a reliance on the assumption that the technology is always accurate. Put another way, algorithms learn from historical data, after which this knowledge can then be applied to novel data or situations i.e. future situations. For example, an algorithm can be used to find patterns in the historical data of a credit institution in relation to the taking and repayment of loans at that credit institution, e.g. a bank. The historical data comprises two types of data, predictor data and behavioural data and the algorithm is able to identify, make correlations and ultimately weigh the predictor data in relation to the behavioural data. In other words, the algorithm is able to identify which characteristics or attributes were present (predictor data) and what the outcome was, based on those attributes (behavioural data). The algorithm can identify the characteristics of those who took a loan as well as those who had difficulty paying back a loan and analyse this in relation to the outcomes. This knowledge can then be applied to new loan applications. The following is a formal definition of the predictive model:

> A predictive model captures the relationships between predictor data and behaviour, and is the output from the predictive analytics process. Once a model

1 English Oxford Living Dictionaries, *Algorithm*, available at "en.oxforddictionaries.com/definition/algorithm" (last accessed on 2017-03-19).

> has been created, it can be used to make new predictions about people (or other entities) whose behaviour is unknown.[2]

In other words, one can imagine a bag of data. The data points within this bag can be divided into two types of data points, namely data characterized by means of varying attributes (e.g. characteristics associated with human attributes such as age, gender, salary, marital status etc.) and data that represents human behaviour (payed loan on time, defaulted on repayments etc.). By identifying the correlations between these two types of data, human behaviour can be predicted. One way of maximizing a commercial institutions profit is by minimizing expenses and risk, and one way of minimizing expenses and risk is by identifying them at the earliest possible stage in order that they can be avoided. In other words, predictive models are used by commercial actors in order to make strategic commercial decisions, a substantial part of this process requiring the identification of various forms of risk.[3] As indicated by its label, prediction is central to predictive modelling. This is so as it is the function of predictive models to identify future risks and ultimately mitigate them.

Associated with predictive models is the notion of 'scoring'. Scoring can be described as follows:

> The process of using a model to make predictions about behavior that has yet to happen is called "scoring." The output of the model, the prediction, is called a score. Scores can take just about any form, from numbers to strings to entire data structures, but the most common scores are numbers (for example, the probability of responding to a particular promotional offer).[4]

3 The Notion of Prediction

The notion of predicting human behaviour is gaining momentum. This is so as the technologies used to identify future human behaviour as well as make decisions about people based upon this knowledge is becoming more effective and therefore more popular in the commercial sphere. As the effectiveness of the technologies used to make these predictions is viewed as becoming more accurate, the more applications this technology is being put to. And as the benefits of prediction are seen as a triumph within the commercial sector public authorities are jumping onto the bandwagon, also using prediction e.g. in the fight against crime or to detect fraud in advance.

---

2 Finlay, Steven, *Predictive Analytics, Data Mining, and Big Data: Myths, Misconceptions and Methods*, Palgrave Macmillan, 2014, at p. 215.

3 In search of a definition of 'risk' many alternatives exist. For example, The Society for Risk Analysis (SRA) defines it as, 'the potential for realization of unwanted, adverse consequences to human life, health, property, or the environment', in Wahlgren, Peter, *Legal Risk Analysis: A Proactive Legal Method*, Jure, 2013 at p. 20.

4 Thearling, Kurt, *Scoring Your Customers*, available at "www.thearling.com/text/scoring/scoring.htm" (last accessed on 2018-06-05).

There are a number of initiatives that illustrate the increased reliance on prediction. For example, in an attempt to model almost all aspects of the future, the Swiss Federal Institute of Technology has initiated the FutureICT Project, which is building a single model that will be fed almost all available data in the world and which will potentially predict anything. The heart of the project is the Living Earth Simulator, a huge predictive model that will model economies, governments, cultural trends, epidemics, agriculture and technological developments by employing algorithms, and ultimately have insight into the future.[5]

Just as predictive models are being used to predict the weather, earthquakes and volcanic eruptions, so too can they be used to predict human behaviour. This is not particularly difficult considering that human beings are creatures of habit. It is estimated that habit accounts for forty-five percent of the choices humans make every day.[6] The thrust, therefore, of predictive modelling is that the greater the extent to which a human action has been repeated in the past, the greater the probability will be that it will occur in the future. It is on this premise that algorithms, using machine-learning techniques, search data and identify examples of human actions, acquiring the ability to assess the probability of these actions re-occurring in the future. Also central here are the concepts of digitalization and digitization, the former referring to, 'the adoption or increase in use of digital or computer technology by an organization, industry, country, etc.',[7] while the latter refers to, 'the action or process of digitizing; the conversion of analogue data (esp. in later use images, video, and text) into digital form'.[8] Ultimately, the data created by humans as they make their way through the digital sphere speaks volumes about their personalities as well as their behaviour.

The power emanating from the access to predictive models increases as the knowledge they identify is combined with the behavioural sciences, which provide even greater insight into human behaviour. It has been stated that, '[w]e live in an age of psychology and behavioural economics – the behavioural sciences'.[9] Consequently, with this insight into human behaviour, comes power, as the ability to identify human behaviour brings with it the ability to influence and alter human behaviour. Subsequently, people have never been more

5 Weinberger, David, *The Machine That Would Predict the Future,* Scientific American, December 2011 at p. 32.

6 Neal, David T., Wood, Wendy and Quinn, Jeffry M., *Habits- A Repeat Performance*, Current Directions in Psychological Science, August 2006**,** Vol. 15 no. 4, pp 198-202, at p. 198, available at "cdp.sagepub.com/content/15/4/198.abstract" (last accessed on 2016-11-25). *See* also Chi, Kelly Rae, *Why are Habits So Hard to Break?*, DukeToday, available at "today.duke.edu/2016/01/habits" (last accessed on 2016-11-25).

7 Oxford English Dictionary, available at "www.oed.com/view/Entry/242061#eid189542747" (last accessed on 2018-05-22).

8 Oxford English Dictionary, available at "www.oed.com/view/Entry/240886?redirected From=digitization#eid" (last accessed on 2018-05-22).

9 Sunstein, Cass R., *The Ethics of Influence: Government in the Age of Behavioural Science*, Cambridge University Press, 2016, at p. 1.

susceptible to being manipulated into making choices that they otherwise may not have made, this manipulation taking place covertly, in turn threatening personal autonomy and the notion of the individual as an autonomous agent.[10]

Many actors within society are utilizing predictive modelling, e.g., governments, public authorities, law enforcement, the health care sector and most notably, private commercial actors. Sunstein remarks:

> For-profit companies are using behavioural research every day. They want to learn how people think and to use that learning to make money. Charitable organizations consult behavioural scientists to find out how they might attract donors and increase donations. For their part, public officials are increasingly turning to the behavioural sciences to promote their goals. They are influencing people in multiple ways in order to reduce poverty, to increase employment, to clean the air, to improve health, to encourage people to vote, and to increase safety on the highways. What are the ethical constraints on their actions?[11]

Witten and Frank highlight the connection between predictive modelling and business:

> A scientist's job … is to make sense of data, to discover the patterns that govern how the physical world works and encapsulate them in theories that can be used for predicting what will happen in new situations. The entrepreneur's job is to identify opportunities, that is patterns in behaviour that can be turned into a profitable business, and exploit them.[12]

For a commercial actor, a research and development project can entail a risk, taking on a new marketing strategy can be a risk or entering into a contract can entail a risk. However, people's behaviour in relation to that commercial actor can also be a risk. For example, an existing customer about to leave for a competitor is a risk or a potential customer can become a risk should he or she result in unnecessary and unexpected costs. The ability to predict and therefore pre-empt these risks places the commercial actor in a powerful position.

However, the process need not stop there. In this regard, reference is made to the concept of 'nudge', which Thaler and Sunstein use to describe the manner in which the actions of individuals can be modified using subliminal techniques which are seemingly insignificant yet which can have a substantial effect on behaviour.[13] In other words, using small increments, people can be pushed or

---

10 For a more detailed examination of the effects of predictive modelling on human autonomy, reference is made to Greenstein, Stanley, *Our Humanity Exposed: Predictive Modelling in a Legal Context*, dissertation Stockholm University, available at "su.diva-portal.org/smash/record.jsf?pid=diva2%3A1088890&dswid=-7446".

11 Sunstein, above n. 9, at p. 1.

12 Witten, Ian H. and Frank, Eibe, *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, 2005, at p. 4.

13 Thaler, Richard and Sunstein, Cass R., *Nudge: Improving Decisions about Health, Wealth and Happiness*, Penguin Books, 2008. It should also be noted that the notion of 'nudging' can have both positive and negative connotations.

nudged into changing their behaviour to the benefit of commercial actors. Not only can human behaviour be predicted, but using the right techniques, it can even be modified or manipulated.

4 The Advantages and Disadvantages with Prediction

It can be said that all technologies are in essence neutral and it is the uses that they are put to that determines the extent to which they are perceived as a risk or threat to society. Predictive modelling is no different. It is a powerful technology that has both positive and negative aspects, and while it is important to consider the immense benefits that it has for society, it remains an important task to identify its dangers in order that steps can be taken to mitigate these.

4.1 The Advantages

There are indisputable advantages with the use of predictive models and there is no doubt that society has much to benefit from their use. First, from the economic perspective, the availability of data and the technical means to exploit these data, is valuable for the economic development of society. Some examples mentioned by the OECD are the effectivization of the manufacturing industry, the more effective use of labour, the ability to tailor services and the more effective use of energy resources by implementing 'smart' solutions.[14] In addition, according to the EU, '[d]ata has become the essential resource for economic growth, job creation and societal progress. Data analysis facilitates better decision-making, innovation and the prediction of future events'.[15]

Within the health industry, the use of technology based on data and modelling practices is helping medical practitioners diagnose and treat illnesses.[16] For example, models assess the risks associated with surgery, models predict influenza trends (Google flu trends), models predict breast cancer, sepsis, HIV progression and the effect of medical drugs and models identify the risks for

14 OECD, *Data-Driven Innovation for Growth and Well-being*, available at "www.oecd.org/sti/ieconomy/data-driven-innovation.htm" (last accessed on 2017-01-18).

15 European Commission, *Public Consultation on Building the European Data Economy*, available at https://ec.europa.eu/digital-single-market/en/news/public-consultation-building-european-data-economy (last accessed on 2017-03-30). *See* also European Commission, *Commission Staff Working Document on the Free Flow of Data and Emerging Issues of the European Data Economy*, {COM(2017) final}, Brussels, 10.1.2017, SWD(2017) 2 final, available at "ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy" (last accessed on 2017-03-30).

16 Memorial Sloan Kettering Cancer Center, *Memorial Sloan Kettering Trains IBM Watson to Help Doctors Make Better Cancer Treatment Choices*, available at "www.mskcc.org/blog/msk-trains-ibm-watson-help-doctors-make-better-treatment-choices" (last accessed on 2018-06-04) and TMF, *Can an Algorithm Diagnose Better than a Doctor,* available at "medicalfuturist.com/can-an-algorithm-diagnose-better-than-a-doctor/" (last accessed on 2018-06-04).

babies born prematurely.[17] Concerning the treatment of prematurely born babies, a venture between IBM, the University of Ontario Institute of Technology and Canadian Hospital uses software developed by IBM to increase the chance of survival of these babies. By monitoring various biomedical information with the use of sensors, doctors can gain access to information to help them make better decisions concerning treatment, where the model can notify of a change in a baby's medical condition up to twenty-four hours in advance by analysing the physiological data.[18] Predictive modelling also improves consumers' experience, where more relevant advertising is received and tailored digital services more accurately reflect the tastes of the consumer. Initiatives at airports using predictive models are also not only solving congestion problems, but also reducing emissions from planes, thereby resulting in gains in relation to the environment. For example, predictive models are being used at Heathrow airport to predict the best order of take-off for the planes. This system then allows air traffic control to determine with more precision, when planes should push back from their gates, thereby reducing the amount of time that they stand idle with their engines running.[19] Finally, the use of predictive models by public authorities benefits society as a whole, e.g. in reducing crime or ensuring that the economy is strong and that taxes are collected, allowing for society to prosper.

Siegel outlines many predictive modelling incentives by private companies and public authorities that are of benefit to society, both from an economic perspective but also from the perspective of the individual.[20]

4.2 The Disadvantages

For all the benefits associated with predictive modelling, there are a number of potential harms where predictive models are used to gauge people's personality and predict what type of person they are. This process can negatively affect the individuals upon whom these models operate, in a number of ways. There is little public awareness of this practice, and when knowledge of its existence is publicised, it results in a feeling of uncertainty, insecurity and a loss of control on the part of individuals, who are the target of this powerful technology. However, the harms resulting from predictive modelling are not only emotional.

---

17 Siegel, Eric, *Predictive Analytics – The Power to Predict Who Will Click, Buy, Lie or Die*, John Wiley & Sons, 2013, at p. 274.

18 IBM, *First-of-a-Kind Technology to Help Doctors Care for Premature Babies*, available at "/www-03.ibm.com/press/us/en/pressrelease/24694.ws"s (last accessed on 2017-01-18) in Siegel, above n. 118, at p. 60.

19 Institute for Aerospace Technology, *IAT Academic discusses airport scheduling at Heathrow on BBC series*, available at "www.nottingham.ac.uk/aerospace/news/iat-academic-discusses-airport-scheduling-at-heathrow-on-bbc-series.aspx" (last accessed on 2017-02-27).

20 Siegel, Eric, *Predictive Analytics – The Power to Predict Who Will Click, Buy, Lie or Die*, John Wiley & Sons, 2013, at pp. 265-289.

Predictive modelling can result in harms that are concrete in nature and with a varying degree of recognition in the law.[21]

4.2.1 The 'Washington, D.C.' example

The following scenario illustrates the risks associated with predictive modelling. It is the real-life 'Washington, D.C. example'.[22] In 2007 the mayor of Washington, D.C. wanted to increase the performance of school children. Acting upon the theory that the students' bad results were the result of bad teachers, the newly hired school chancellor for Washington, D.C. embarked on a plan to rid the district of the bad teachers by implementing a teacher assessment tool called IMPACT. The system utilized a 'value-added' method that used mathematical formulas to determine how much value a teacher had added to what a student had learnt.[23] As a result of the system, 206 teachers (the bottom 2 percent identified by the system) were fired. One of the victims of IMPACT was a 5th grade teacher SW. She had a good reputation and had received excellent reviews, both from the school principal and parents. However, she received a poor IMPACT assessment for teaching maths and language, as generated by the IMPACT algorithm. Based on SW's performance as calculated by the model, she was dismissed. Now the company behind the development of the IMPACT model had been given the task of modelling the progress of the students in the district and then calculate how much of their performance (both development or decline) could be attributed to their teachers. This task was complex as many factors can shape the development of a student, both at school but also private.

---

21 For a detailed investigation concerning the potential harms connected to predictive modelling and their establishment in law, reference is made to Greenstein, Stanley, *Our Humanity Exposed: Predictive Modelling in a Legal Context*, dissertation Stockholm University, available at "su.diva-portal.org/smash/record.jsf?pid=diva2%3A1088890&dswid=-7446".

22 This example is referred to in O'Neil Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown, 2016, at pp. 3-11. This real-life example is portrayed by O'Neil who uses the term 'Weapons of Math Destruction' (WMD) to describe models based on mathematics and that are predictive in nature. The above scenario was also investigated in Turque, Bill, *'Creative … motivating and fired'*, The Washington Post, March 6, 2012, available at "www.washingtonpost.com/local/education/creative--motivating-andfired/2012/02/04/gIQAwzZpvR_story.html?utmterm=.c00363531ed5" (last accessed on 2017-02-27), Strauss, Valerie, *Firing of D.C. teacher reveals flaws in value-added evaluation*, March 7, 2012, available at "www.washingtonpost.com/blogs/answer-sheet/post/firing-of-dc-teacher-reveals-flaws-in-value-added-evaluation/2012/03/07/gIQAtmlGxR_blog.html?utm_term=.6e40150898f2" (last accessed on 2017-02-27) and Gillum, Jack and Bello, Marisol, *When standardized test scores soared in D. C., were the gains real?* USA Today, March 30, 2011, available at "usatoday30.usatoday.com/news/education/2011-03-28-1Aschooltesting28_CV_N.htm" (last accessed on 2017-02-27). The term 'Washington, D.C. example' is provided by the author of this article.

23 Strauss, Valerie, *Firing of D.C. teacher reveals flaws in value-added evaluation*, March 7, 2012, available at "www.washingtonpost.com/blogs/answer-sheet/post/firing-of-dc-teacher-reveals-flaws-in-value-added evaluation/2012/03/07/gIQAtmlGxR_blog.html?utm_term=.6e40150898f2" (last accessed on 2017-02-27).

Two anomalies also presented themselves in the investigation of this event. First, SW noticed that the level of proficiency of the students leaving 4th grade was stated to be high but that they had dropped considerably by the time they started SW's 5th grade class. Second, the media that started to investigate this event found that the actual physical tests of the students showed a large degree of erasure, up to seventy percent, a sign that there may have been cheating.[24] Considering that teachers received a monetary incentive for their students outperforming other students coupled with the fact that their own jobs were on the line, being under the constant scrutiny of the IMPACT model, a plausible finding suggested that 4th grade teachers were in fact altering the tests of their students in order to portray themselves in a better light. This resulted in the appearance that SW's students had become worse during the 5th grade, while in actual fact they had started the school year with an inflated level of proficiency, making SW look bad.

4.2.2 Creditworthiness in China

Another example concerning prediction regards an initiative in China, where creditworthiness is becoming the measure of the model citizen.[25] A number of companies in conjunction with the state have produced a system of prediction whereby a citizen's reliability is reduced to a score produced by predictive models. The task of determining the model citizen has been left to technology and the criteria being used to determine this is creditworthiness. The system uses big data, against which an algorithm is run, in order to monitor every action that a person takes in the digital environment. A digital image of the person is compiled based on this data. It is possible to acquire a maximum of 950 points. If you have a score of 650, that is considered good and you are entitled to a number of benefits and some of the obstacles that would normally need to be overcome in order to gain certain societal benefits, are removed. A low score, e.g. under 300 points, makes life more difficult. You are not eligible for the better jobs in society and many obstacles appear that make everyday life more difficult.

The Chinese authorities are open about the system and about how the model citizen is determined. They even provide advice on how one can improve one's score, e.g., by purchasing the right goods such as electrical appliances, by refraining from buying frowned upon products such as certain computer games, by refraining from using those social media considered unacceptable and by paying bills on time. Also, having friends with a low score can bring down a person's score and the only way to remedy this is to cut all contact with these 'undesirable' friends. There is also an app that can be downloaded to view what

24 Gillum, Jack and Bello, Marisol, *When standardized test scores soared in D. C., were the gains real?* USA Today, March 30, 2011, available at "usatoday30.usatoday.com/news/education/2011-03-28-1Aschooltesting28_CV_N.htm" (last accessed on 2017-02-27).

25 Stanley, Jay, ACLU, *China's Nightmarish Citizen Scores Are a Warning for Americans*, available at "www.aclu.org/blog/privacy-technology/consumer-privacy/chinas-nightmarish-citizen-scores-are-warning-americans?redirect=blog/free-future/chinas-nightmarish-citizen-scores-are-warning-americans" (last accessed on 2018-06-05).

your score is and the trend is to publish the result for everyone to see.[26] It is the predictive model and scoring system that is at the heart of this programme, where an algorithm compiles a digital identity of an individual "which is reflected as" a score. This is a typical instance where reputation is determined by a computer system and the data points associated with a particular individual.

4.2.3 The prediction of crime

While much of the focus up until now has been the use of predictive models and scoring within the private commercial sphere, some disconcerting examples associated with the use of predictive models and scoring comes from within crime prevention and law enforcement activities. Here, models are being used to determine where spikes in crime will occur so that resources can be deployed more effectively.[27]

In addition, algorithms are being used to develop so-called 'threat scores', that is the degree of threat that a specific person poses, potentially to the surroundings, but also to law enforcement that may be required to enter a premises. One such initiative from Fresno California, USA, uses the software called 'Beware' in order to give specific individuals a score, the data fed to the predictive algorithm originating from official public data sources and combined with searches of the social media and other internet websites.[28] Should an individual call law enforcement in connection with a disturbance at a provided address, the software automatically creates a profile of that person in connection with that specific address (the same is applicable to vehicles).[29] In addition, Beware can make a profile of all the inhabitants in the surrounding area from which a call was made, which means that if a person calls the police in connection with a disturbance, that person's neighbours too will inevitably be associated with a score, even if they had absolutely no involvement with the disturbance.[30]

---

26 Åkerblom, Tobias Andersson, *Så blir du en kinesisk mönstermedborgare*, available at "kit.se/2015/09/25/11823/sa-blir-du-en-kinesisk-monstermedborgare/" (last accessed on 2015-12-01). *See* also Chin, Josh, *China's New Tool for Social Control: A Credit Rating for Everything*, The Wall Street Journal, available at www.wsj.com/articles/chinas-new-tool-for-social-control-a-credit-rating-for-everything-1480351590 (last accessed on 2017-02-07).

27 Data Impacts Case Studies, available at "dataimpacts.org/project/predictive-modeling-fights-crime/" (last accessed on 2018-06-05).

28 Stanley, Jay, ACLU, *Eight Problems with Police 'Threat Scores'*, available at "www.aclu.org/blog/privacy-technology/surveillance-technologies/eight-problems-police-threat-scores" (last accessed on 2018-06-05).

29 Stanley, Jay, ACLU, *Eight Problems with Police 'Threat Scores'*, available at "www.aclu.org/blog/privacy-technology/surveillance-technologies/eight-problems-police-threat-scores" (last accessed on 2018-06-05).

30 Stanley, Jay, ACLU, *Eight Problems with Police 'Threat Scores'*, available at "www.aclu.org/blog/privacy-technology/surveillance-technologies/eight-problems-police-threat-scores" (last accessed on 2018-06-05).

Another example comes from Chicago, where police have embarked on a violence reduction strategy and consequently use predictive software including algorithms in order to create a ‘Strategic Subject List’ (SSL), which is also referred to as the ‘heat list’ and which contains the names of 400 individuals considered most dangerous in the city. The list is created by algorithms that use historical crime data, disturbance calls, suspicious person reports but also social media activity in order to compile the list.[31] The people on the list are not necessarily violent criminals. Rather, they have made it onto the list because they have been arrested with a person that has previously been arrested for a violent act. For example, a person may be arrested for a minor offence such as speeding together with a person previously arrested for a violent crime. This connection will automatically ensure that the first person lands on the list of potential violent people even though he or she has never committed a violent act in his or her life.[32] What is interesting here is the technology behind the ‘heat list’, in that it is not necessarily only the actual names of a person’s acquaintances in the social media that result in a person landing on the ‘heat list’, but the the underlying structure of the social media networks is also relevant in this regard.[33]

4.3 Consequences

It is clear from the above scenarios that we are moving to a juncture where predictive models are being used to score people, both in a commercial but also law enforcement context. There are a number of issues that require highlighting.

O’Neil, coining the term ‘weapons of math destruction’ (WMD), depicts one of the main dangers with predictive models:

> But you cannot appeal to a WMD. That’s part of their fearsome power. They do not listen. They do not bend. They’re deaf not only to charm, threats and cajoling but also to logic – even when there is good reason to question the data that feeds their conclusions. Yes, if it becomes clear that automated systems are screwing up on an embarrassing and systematic basis, programmers will go back and tweak the algorithm. But for the most part, the programs deliver unflinching verdicts,

31 Stanley, Jay, ACLU, *Chicago Police ”Heat List” Renews Old Fears About Government Flagging and Tagging*, available at “www.aclu.org/blog/privacy-technology/chicago-police-heat-list-renews-old-fears-about-government-flagging-and?redirect=blog/chicago-police-heat-list-renews-old-fears-about-government-flagging-and-tagging” (last accessed on 2018-06-05).

32 Stanley, Jay, ACLU, *Chicago Police ”Heat List” Renews Old Fears About Government Flagging and Tagging*, available at “www.aclu.org/blog/privacy-technology/chicago-police-heat-list-renews-old-fears-about-government-flagging-and?redirect=blog/chicago-police-heat-list-renews-old-fears-about-government-flagging-and-tagging” (last accessed on 2018-06-05).

33 Stanley, Jay, ACLU, *Chicago Police ”Heat List” Renews Old Fears About Government Flagging and Tagging*, available at “www.aclu.org/blog/privacy-technology/chicago-police-heat-list-renews-old-fears-about-government-flagging-and?redirect=blog/chicago-police-heat-list-renews-old-fears-about-government-flagging-and-tagging” (last accessed on 2018-06-05).

> and the human beings employing them can only shrug, as if to say, "Hey, what can you do?".[34]

The above citation illustrates one of the many harms associated with these black box decision-making systems based on prediction, namely, 'what to do when the computer says "no!"'. In other words, having been on the receiving end of a negative and potentially incorrect decision taken by a predictive model, what can a person do to rectify the situation? The problem lies in the fact that the complexity of predictive models prevents insight into how they work and those relying on their decisions also lack knowledge of how they operate, the model having been developed by a computer scientist or mathematician far away in the commercial chain. A second problem, and something that is partially highlighted in the above examples, is a steadfast faith in technology and the attitude that technology is based on science, which never gets it wrong. However, in actual fact, predictive models can get it wrong if the data that they are trained on is incorrect or biased in some manner. A third issue is that in training predictive models, not all future scenarios can be anticipated, which can lead to an incorrect result where the predictive model is required to make a decision on a scenario it has not encountered previously, i.e. been trained to adjudicate. This is referred to as 'overfitting'. Overfitting occurs when a predictive model, once operating live, is confronted with an example or scenario not encountered during training and therefore cannot deal with it in an anticipated manner. Put another way, the goal when constructing a predictive model is that it should apply to all possible examples that it is confronted with and not only to the examples in the training data set, which is why it is important for the model to generalize beyond the training set. Overfitting is defined as, '… the tendency … to tailor models to the training data, at the expense of generalization to previously unseen data points'.[35] Consequently, overfitting can be associated with model complexity and by decreasing complexity, the extent to which overfitting occurs can be controlled.[36] It is argued that less supervised predictive models are more prone to overfitting mainly due to the fact that humans have less control over the functioning of the model. For example, neural networks are more prone to overfitting.[37] A final harm associated with predictive models pertains to a social consequence of their use. Once again O'Neil puts this eloquently:

> They tend to punish the poor. This is, in part, because they are engineered to evaluate large numbers of people. They specialize in bulk, and they're cheap. That's part of their appeal. The wealthy, by contrast, often benefit from personal input. A white-shoe law firm or an exclusive prep school will lean far more on

34 O'Neil Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown, 2016, at p. 10.

35 Provost, Foster and Fawcett, Tom, *Data Science for Business: What you Need to Know About Data Mining and Data-Analytic Thinking*, O'Reilly Media, 2013, at p. 113.

36 Provost, Foster and Fawcett, Tom, *Data Science for Business: What you Need to Know About Data Mining and Data-Analytic Thinking*, O'Reilly Media, 2013, at p. 140.

37 Finlay, Steven, *Predictive Analytics, Data Mining, and Big Data: Myths, Misconceptions and Methods*, Palgrave Macmillan, 2014, at p. 102.

> recommendations and face-to-face interviews than will a fast-food chain or a cash-stripped urban school district.[38]

Gandy echoes this line of thought, by referring to what he terms the 'panoptic sort', which is the explicit use of technology in order to exclude undesirable people for various reasons:

> The panoptic sort is a screen that excludes a filter that blocks, a magnet that ignores fine wood in preference for base metals … the sorting process works primarily by eliminating those who are too much, too little, too late … too bad!'.[39]

By eliminating the costly clients or potential clients, commercial entities can concentrate on those people who are less risky from the commercial perspective. It is in this manner that the class divisions that characterize the 'real world' are being re-created in the digital environment. Lessig refers to 'hierarchies of social rank', which require information in order to be able to make distinctions based on rank, or what he refers to as 'subtle distinctions of rank'.[40]

Some final considerations on the use of predictive models and scoring conjure up the following thoughts: concerning crime prediction based on predictive models, are we not judging people according to their propensity to be violent or to commit a crime, rather than according to the actual commission of a crime? How does one formally question the decision of a complex algorithm (or multiple complex algorithms) when the technology is complex and the actual algorithms are protected by intellectual property rights? How does one go about getting oneself removed from a list compiled by an algorithm, be it in the commercial or law enforcement setting? What are the implications where the accuracy of the predictive algorithm is called into question? Here one can contemplate the potential consequences where law enforcement has mistakenly been notified by the software that a caller from a residence is a highly violent criminal. Also, do the algorithms take into account that people do change? In other words, once a person has been assigned to a certain category or list, how can he or she remove himself or herself from that stereotype? The questions are endless!

5 Conclusions

The modern society can be characterised by the striving to become more effective. This is relevant both in relation to commercial actors, that are required to be more competitive as well as law enforcement, which is required to demonstrate its effectiveness, potentially under political duress. This has

38 O'Neil Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown, 2016, at p. 8.

39 Gandy, Oscar H., Jr., *The Panoptic Sort: A Political Economy of Personal Information*, Westview Press, 1993, at p. 18.

40 Lessig, L., *Code 2.0*, Basic Books, 2006, at p. 221.

resulted in the move towards the predictive society where the increased use of digital technologies by ordinary people has facilitated the ability to record their every move as well as assign them a score based on algorithms incorporated into predictive models. The availability of these technologies has allowed for a shift in society – a shift from the state of being reactive to the state of being proactive. No longer does the commercial actor need to wait until the client has left for a competitor before attempting to entice him or her back with a special offer. The technology allows the commercial actor to predict the clients intent to leave even before that client knows this him or herself. And no longer is law enforcement required to wait until the crime is committed before approaching the criminal. The technology allows for the potential criminal to be identified in advance. The ethical and moral questions that arise in connection with the above scenarios are boundless. There is no ‘silver bullet’ solution to resolve all the issues in one foul swoop. However, what is required is that we start to question the desirability of these technologies in a democratic debate before it is too late.