ENCODING INEQUALITY: THE CASE FOR GREATER REGULATION OF ARTIFICIAL INTELLIGENCE AND AUTOMATED DECISION-MAKING IN NEW ZEALAND

Ella Brownlie*

Automated decision-making systems, developed using artificial intelligence and machine learning processes, are being used by companies, organisations and governments with increasing frequency. The purpose of this article is to outline the urgent case for regulating automated decision-making and examine the possible options for regulation. This article will argue that New Zealand's current approach to regulating decision-making is inadequate. It will then analyse art 22 of the European Union's General Data Protection Regulation, concluding that this regime also has significant flaws. Finally, this article will propose an alternative regulatory solution to address the novel challenge posed by automated decision-making. This solution aims to strike a balance between the interests of organisations in capitalising on the benefits of automated decision-making technology and the interests of individuals in ensuring that their right to freedom from discrimination is upheld.

I INTRODUCTION

Artificial intelligence is reshaping the world around us.[1] The world we live in is increasingly customised by thousands of invisible decisions that decide the content we interact with, the decisions we make and even our emotions.[2] Within the next two decades, advancements in this

* Submitted for the LLB (Honours) Degree, Victoria University Faculty of Law, 2019. I would like to express my thanks to my supervisor Marcin Betkier for his invaluable guidance and support.

1 Shoshana Zuboff "The Surveillance Threat is Not What Orwell Imagined" *Time* (online ed, New York, 6 June 2019).

2 Zuboff, above n 1.

technology may change the very fabric of human civilisation.[3] How we regulate artificial intelligence systems is therefore one of the most urgent questions facing humanity.[4] This article will argue that neither New Zealand's current regulatory framework nor the framework proposed by art 22 of the European Union's General Data Protection Regulation (GDPR), is an adequate regulatory solution to address the issue of automated decision-making (ADM).[5] An alternative and innovative approach to law reform in this space is needed for New Zealand to lead the way in effectively regulating ADM and artificial intelligence.

II UNDERSTANDING AUTOMATED DECISION-MAKING (ADM)

A What is ADM?

ADM is a term used in the GDPR to refer to decisions made by technological means.[6] The term covers a broad range of technological processes. At the most simple level, ADM could involve a computer applying a mathematical formula that has been predetermined by a human to new sets of data.[7] At the more complex level, machine learning and narrow artificial intelligence systems can be used to generate an automated decision that reflects the computer's interpretation of past data.[8] Through the use of machine learning technology, computers can be "taught" to identify patterns in large data sets.[9] The machine then develops a predictive formula based on the patterns the machine has observed in the past data. This formula can be applied to new data inputs to model the likelihood of different outcomes occurring in the new scenarios.[10]

The difference between simplistic and more complex ADM can be illustrated with an example. If an employer knows they want to hire a university graduate, they could write a formula

---

3 Tad Friend "How Frightened Should We Be of AI?" *The New Yorker* (online ed, New York, 7 May 2018).

4 Meredith Whittaker and others *AI Now Report 2018* (AI Now Institute, December 2018) at 7.

5 Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1, art 22 [GDPR].

6 Article 29 Data Protection Working Party *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* (17/EN WP 251, 6 February 2018) at 8 [*Working Party 29 Guidelines*].

7 John Zerilli and others "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" (2019) 32 Philosophy & Technology 661 at 663.

8 *Working Party 29 Guidelines*, above n 6, at 4.

9 Russell Brown "Actually Interesting: A machine can make decisions but can it ever understand why?" (1 August 2019) The Spinoff <https://thespinoff.co.nz>.

10 Brown, above n 9.

which filtered out the CVs of prospective candidates without a university degree. In this case, the computer would be applying a set formula that has been predetermined by a human.[11] Under this formula, the human has decided that they will only hire candidates with a university degree, and the computer has efficiently applied that decision to the data it is given by filtering out candidates without a degree.

In contrast, if an employer intended to use a machine learning system to filter the CVs of prospective job candidates, they would give the machine a database of the CVs of past candidates, identified by those who were successful and those who were not. Using this information the machine learning system would determine a formula by which CVs should be screened out.[12] In this case, the human is not fully aware of what patterns the computer might recognise in the past data and apply to new applicants.[13] The computer might recognise, for example, that based on the past data, features of a candidate correlate more strongly to their being offered the job than whether they have a university degree. For example, if there was a higher proportion of Pākehā applicants who had been successful in obtaining the role in the past, when compared to the proportion of applicants of other ethnicities, the computer might assume that being Pākehā is a stronger predictor of a successful candidate, than having a university degree. This could lead the computer to perpetuate systemic racism, allowing Pākehā applicants without a degree to pass a screening process whereas applicants of other ethnicities would have been rejected.[14]

This article will focus on the latter type of automated decisions, produced through machine learning and "big data" processes. It is the use of these intelligent systems to generate automated decisions through pattern recognition which represents a novel shift in human decision-making processes.[15] Such a shift warrants a thorough analysis of the legal implications and problems created by ADM technology.

B Where is ADM Currently Used?

ADM is currently being used across a plethora of different areas in our society, ranging from the seemingly mundane or trivial, to more important decisions. ADM is used to tell us which TV show on Netflix we should watch next and to control the content we see in our Facebook

---

11 Zerilli and others, above n 7, at 3.

12 Cathy O'Neil "How can we stop algorithms telling lies?" *The Guardian* (online ed, London, 16 July 2017).

13 Cathy O'Neil *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing Group, New York, 2016) at 8–10.

14 There are multiple real life examples similar to this one. See for example Jeffrey Dastin "Amazon scraps secret AI recruiting tool that showed bias against women" (10 October 2018) Thompson Reuters <www.reuters.com>.

15 Eliana Herrera-Vega "Artificial Intelligence and law" [2019] NZLJ 64 at 67.

Newsfeeds every day.[16] In the United States, the technology has been used to assist judicial decisions on criminal sentencing, determine police resource allocation and determine performance-based pay for teachers.[17] In all of these cases, the algorithms behind these decisions are highly complex, based on multiple factors which are continuously being adjusted.[18]

New Zealand has been slower in adopting ADM technology compared to the United States, however it is beginning to be used.[19] A report from Statistics New Zealand has documented the use of algorithms by the public sector.[20] ADM is currently being used by the Accident Compensation Corporation (ACC) to process claims for accident cover that are submitted.[21] The algorithm was developed to analyse past claims data and determine a formula for identifying claims that are "complex" and need to be referred to a human and claims which are straightforward and can be approved straight away.[22] Though the system does not have the power to grant a person cover, this use of past data to design a filtering system for claims may result in inequities. It is possible, for example, that claimants with pre-existing health conditions or disabilities may be classified as "complex" even when they have a simple claim for an accident. Such a classification would be based on the fact that in past data similar claimants were less likely to have their cover approved because ACC does not provide cover for issues arising from pre-existing medical conditions. The possible issues arising from reasonably minor uses of ADM such as this demonstrate the need for regulatory attention to this issue.[23] This is particularly so given that public and private actors in New Zealand are likely to adopt ADM technologies at increasing scale over the coming years.

---

16 Wall Room Media "Facebook Newsfeed Algorithm History" <https://wallaroomedia.com>; and Libby Plummer "This is how Netflix's top-secret recommendation system works" (22 August 2017) Wired <www.wired.co.uk>.

17 Julia Angwin and others "Machine Bias" (23 May 2016) ProPublica <www.propublica.org>; and O'Neil *Weapons of Math Destruction*, above n 13, at 4, 85 and 139.

18 Stuart Dredge "How does Facebook decide what to show in my news feed?" *The Guardian* (online ed, London, 30 June 2014); and Alex Hern "Why Facebook's news feed is changing and how it will affect you" *The Guardian* (online ed, London, 12 January 2018).

19 AI Forum New Zealand *Artificial Intelligence: Shaping a Future New Zealand* (May 2018).

20 Statistics New Zealand *Algorithm Assessment Report* (October 2018).

21 At 23.

22 At 23.

23 Cathy O'Neil "When Not to Trust the Algorithm" *Harvard Business Review* (online ed, Massachusetts, 6 October 2018).

C The Opportunities and Risks of ADM

ADM offers an unparalleled opportunity to improve efficiency within organisations.[24] By automating decisions, businesses and organisations can reduce the amount of human employee time spent on basic decision-making tasks, enabling businesses to cut costs or direct their employees' time towards more valuable activities.[25] In addition, technologist Hal Varian has noted the potential business opportunities that ADM creates for content, products and services which are more tailored to individual customers' preferences than ever before.[26] PricewaterhouseCoopers (PwC) estimates that such developments in ADM will increase global GDP by 14 per cent by 2030, an increase of USD 15.7 trillion.[27]

However, despite the economic opportunities created by ADM, this technology also presents some risks. As the ACC example illustrates, ADM has the potential to amplify discriminatory trends and erode human dignity. One of the key attractions of ADM initially was its potential to reduce "human bias" in decision-making processes.[28] However, as Cathy O'Neil has exposed, this is a false expectation of ADM technology.[29] Machine learning systems are based on combining *past data* and a *definition of success*, neither of which are objective.[30] Past data frequently reflects pre-existing discriminatory trends and definitions of success are often based upon assumptions held by dominant cultures, ignoring the experiences of marginalised groups.[31] O'Neil draws on examples from education, banking, recruitment and healthcare to argue that far from eliminating "human bias" in decision-making, destructive uses of ADM *amplify* that bias.[32] Operating completely opaquely, unregulated algorithms can automate the same patterns of discrimination

---

24 Anand S Rao and Gerard Verweij *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* (PricewaterhouseCoopers, 2017).

25 At 1.

26 At 4; and Hal R Varian "Beyond Big Data" (paper presented to the NABE Annual Meeting, San Francisco, 10 September 2013) at 4.

27 Rao and Verweij, above n 24, at 1.

28 Maddy Savage "Meet Tengai, the job interview robot who won't judge you" (12 March 2019) BBC News <www.bbc.com>; and Andrea Gallego and others "How AI Could Help—or Hinder—Women in the Workforce" (13 May 2019) Boston Consulting Group <www.bcg.com/publications>.

29 O'Neil "When Not to Trust the Algorithm", above n 23.

30 O'Neil "When Not to Trust the Algorithm", above n 23.

31 WGBHForum "Cathy O'Neil: How It's Unfair to Use Personality Tests in Hiring" (21 July 2017) YouTube <www.youtube.com>; and Luciano Floridi "What the Near Future of Artificial Intelligence Could Be" (2019) 32 Philosophy & Technology 1 at 4.

32 O'Neil *Weapons of Math Destruction*, above n 13.

on grounds such as race, gender and sexuality which have been prevalent throughout history.[33] Widespread use of these algorithms therefore can exacerbate pre-existing social inequities and violates the right of individuals to be free from discrimination, often in the name of making society fairer.[34]

Subjecting individuals to ADM also poses a threat to the human dignity of individuals. Popular literature has frequently explored situations where humans lose control over or become controlled by machines.[35] This reflects a powerful fear in society of our lives being determined by alien forces we do not understand or trust. The use of ADM systems to make significant decisions that affect people's lives, including loans, employment, healthcare and insurance, strikes up against this underlying fear.[36] As individuals, we may feel a sense of powerlessness, that our distinctive and varied experience is being determined by a set of code that cannot even be explained to us.[37] In addition, collectively, our dignity as a society may also be threatened. Where we allow such systems to be used, unregulated, knowing that they may encode inequality, we undermine the fundamental core of dignity we expect all humans to be allowed; and subject some among us to the unfeeling, mercenary logic of discriminatory machines.[38]

ADM systems offer the possibility of making significant decisions about individuals fairer and more efficient. Yet at the same time, the opaque, unchallengeable nature of some ADM systems and their heavy reliance on past data means there is a real threat that these systems will automate discriminatory patterns and, if unregulated, undermine individual and collective human dignity.

III THE CASE FOR REGULATION OF ADM

A Why Regulate?

The proliferation of ADM systems presents opportunities and challenges for both organisations and individuals. Whilst New Zealand might currently be behind the rest of the world in the large-scale adoption of ADM, our organisations and businesses are likely to catch up to these global trends over the next few years, meaning there is a need to ensure that the New Zealand

---

33 O'Neil "When Not to Trust the Algorithm", above n 23; and Kate Crawford "Artificial Intelligence's White Guy Problem" *The New York Times* (online ed, New York, 25 June 2016).

34 O'Neil "When Not to Trust the Algorithm", above n 23; and Kate Crawford "Artificial Intelligence—With Very Real Biases" *Wall Street Journal* (online ed, New York, 17 October 2017).

35 Some example texts include Mary Shelley's *Frankenstein* published in 1818, popular film *The Matrix* released in 1999 and popular television series *Black Mirror* that first premiered in 2011.

36 Friend, above n 3.

37 Friend, above n 3.

38 O'Neil *Weapons of Math Destruction*, above n 13, at 10.

regulatory framework is fit-for-purpose.[39] Effective regulation in the field of ADM should enable us to balance the benefits offered by ADM, with the need to ensure that individuals are protected from discriminatory decision-making and that human dignity is preserved.

Balanced regulation of ADM could involve numerous options, including establishing prohibitions on ADM for certain significant decisions about individuals, requiring organisations to undertake an "audit" of ADM systems, or establishing review processes for individuals to challenge the results of ADM.[40] Improving transparency over ADM systems may also assist with ensuring ADM is conducted correctly and fairly; however transparency should not be seen as the silver bullet solution for regulation of ADM.[41]

B Is New Zealand's Current Regulatory Framework Adequate?

New Zealand does not have a single statute that directly addresses ADM. The Privacy Bill 2018 offered one opportunity for Parliament to explore the adequacy of New Zealand's regulatory coverage of ADM, however this opportunity was passed over.[42] At the Select Committee stage, several submitters, including the Privacy Commissioner, John Edwards, submitted that ADM regulation should be included in the Bill in order to improve transparency in automated decisions and increase individuals' control over the use of their personal data in these processes.[43] These submissions were rejected by the Justice Select Committee.[44] The Committee argued that New Zealand's current legislation adequately covered ADM and, in addition, humans, not computers, are making almost all of the most significant decisions affecting individuals in New Zealand.[45] The following two sections will challenge both of these claims, arguing that New Zealand's current regulatory environment is *not* fit to regulate ADM and that *even if* humans *are* involved

---

39 PricewaterhouseCoopers *Storm clouds and silver linings: what's the outlook for New Zealand CEOs?* (February 2019) at 3.

40 GDPR, art 22; Rumman Chowdhury and Narendra Mulani "Auditing Algorithms for Bias" *Harvard Business Review* (online ed, Massachusetts, 24 October 2018); and O'Neil "When Not to Trust the Algorithm", above n 23.

41 Statistics New Zealand and the Privacy Commissioner *Principles for the safe and effective use of data and analytics* (May 2018); Cade Metz "Seeking Ground Rules for AI" *The New York Times* (online ed, New York, 1 March 2019); and Colin Gavaghan and others *Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and the Law in New Zealand Project* (New Zealand Law Foundation, 2019) at 57.

42 Privacy Bill 2018 (34–2), cl 3.

43 John Edwards "Submission to the Justice Select Committee on the Privacy Bill 2018" at 29–30; and Gehan Gunasekara "Submission to the Justice Select Committee on the Privacy Bill 2018" at 3.

44 Privacy Bill 2018 (34–2) (select committee report) at 39–40.

45 At 39–40.

in all uses of ADM in New Zealand, the research on automation bias demonstrates there is still pressing need for regulation of ADM processes.

1 The limited efficacy of New Zealand regulations in the context of ADM

Contrary to the Justice Committee's assertion, regulation of ADM in New Zealand is limited. Whilst the Official Information Act 1982 (OIA) requires public parties to exercise a minimum degree of transparency when using ADM, there is no legislative requirement that private parties be similarly transparent. This makes it unlikely that the Human Rights Act 1993 (HRA) could be effectively utilised by individuals to challenge discriminatory ADM by private parties. In addition, even where the OIA does apply, such as for decisions made by public parties, it is possible that the opaque nature of ADM technology may mean that a claimant has insufficient evidence of discriminatory decision-making to be able to meet the standard of proof under the HRA. A brief review of the available legislative avenues will discuss these issues in more detail.

Individuals can use the OIA to demand that a Crown agency provide reasons for a decision that is made using an ADM function.[46] Section 22 provides that individuals have a right to access the "internal rules" of government departments or agencies which are used to inform decisions made by that agency.[47] These rules or policies can be withheld only if there is good reason for doing so.[48] Individuals or organisations could accordingly use 22 to require a government agency to disclose the rules that inform the decision the agency makes automatically, such as a rule that student applicants earning over a certain income level will be denied a government-funded student allowance. In addition, under s 23 of the Act, individuals have a right to be provided with the specific reasoning behind decisions which affect them.[49] This could include decisions made by social welfare agencies, education providers or other Crown entities.[50] Under this section individuals could use s 23 to require an agency to disclose the reasoning behind a particular use of automated decision that affected them. For example, if they were the student applying for a student allowance, they would be told the particular reasons it was denied in their case, such as that their income was over a certain level.

The transparency requirements under the OIA may help individuals to bring a claim under the HRA against public parties for the use of discriminatory decision-making processes. It is illegal to discriminate against citizens on any of the prohibited grounds established in s 21(1) of the Act, which includes race, gender, sexual orientation and religion. A claimant could use the OIA to

---

46 Official Information Act 1982, ss 22–23 [OIA].

47 Section 22.

48 Section 22(4).

49 Section 23(1).

50 Schedule 1.

force a public party to divulge information about the reasons for an automated decision and, if these reasons were discriminatory, the claimant could bring a claim under the HRA.

Nevertheless, because the OIA does not apply to private parties, there is no legislative requirement that private parties provide any rationale for decisions made automatically.[51] Without any information about the reasons for a decision, a claimant would lack the evidence to argue under the HRA that the decision was made on one of the prohibited grounds in s 21(1). Therefore, in order for statutes such as the HRA to provide a remedy for discriminatory ADM, some disclosure requirements would need to be enforced against private operators of ADM systems.

One example of where disclosure is currently being enforced is the Credit Reporting Privacy Code 2004 which places disclosure requirements on private parties in the limited context of generating credit scores. This Code could be used to require private parties, who are using ADM to generate credit scores, to disclose how these scores are produced.[52] Under r 6(2A), when an individual is given access to a credit score, they must also be provided with a statement outlining the methodology used to create the score, the information used and the range their score is placed within.[53] This information could enable an individual to bring a claim for breach of s 21 of the HRA if their credit score was automatically created on a prohibited ground. However, this Code clearly has a limited application, and will be of no assistance to claimants seeking to challenge the use of ADM to make other significant decisions affecting them.

Yet even if claimants have enough evidence, obtained through the OIA or the Credit Reporting Privacy Code, to bring a claim under s 21, they may face another challenge in proving that the ADM produced results based on a prohibited ground. Discrimination in ADM systems frequently emerges, not through inclusion of the protected ground itself, but through another data source which acts as a proxy for that characteristic.[54] For example, a person may not have disclosed their ethnicity on Facebook, but may have entered their postcode when purchasing something on a site where Facebook is embedded. Postcodes are proven to sometimes operate as a proxy for race.[55] Such discrimination by proxy may make it harder for claimants to prove discrimination has occurred on the basis of a prohibited ground in s 21. It therefore remains to be seen whether

---

51 OIA, sch 1.

52 Credit Reporting Privacy Code 2004, r 6(2A).

53 Rule 6(2A).

54 Anupam Datta and others "Proxy Discrimination in Data-Driven Systems: Theory and Experiments with Machine Learnt Programs" (25 July 2017) arXiv <https://arXiv.org> at 1–2.

55 At 3; and Louise Matsakis "Facebook's Ad System Might Be Hard–Coded for Discrimination" (4 June 2019) Wired <www.wired.com>.

sufficient information or explanation regarding how an opaque ADM system is functioning could be obtained to enable a claimant to satisfy the discrimination standard required by the HRA.

New Zealand's current regulatory oversight of ADM is clearly insufficient. The next section will examine whether this is counterbalanced by the inclusion of human involvement in ADM processes in New Zealand.

2 Human involvement in ADM: An antidote to the challenge of regulation?

The Select Committee's second argument against regulating ADM was that humans, not computers remain involved in all major decisions.[56] This section argues that even when humans *are* involved in making decisions with the use of ADM tools, oversight of these tools is needed.

ADM can be used in combination with human decision-making to varying extents.[57] The table below illustrates the spectrum of human involvement that might be included in ADM processes.[58] In all of the situations ranging from 2–10, ADM will have some effect on the decision outcome:[59]

Automation level	Automation Description
1	The computer offers no assistance: human must take all decision and actions.
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans and

56 Privacy Bill (select committee report), above n 44, at 40.

57 ML Cummings "Automation Bias in Intelligent Time Critical Decision Support Systems" (paper presented to American Institute of Aeronautics and Astronautics 1st Intelligent Systems Technical Conference, Chicago, Illinois, 20–22 September 2004) at 1.

58 At 2.

59 At 2.

8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

The centre of the human involvement spectrum – or the middle section of this table – is particularly interesting because in these situations, the ADM system is heavily influential even though a human is involved. In scenarios 3–6 the human plays no role in designing the proposed solution, they merely select from the list available or approve or veto the computer's selection.[60] This is a passive role. The human is relying heavily on the accuracy of the ADM system to select the correct list of options and propose the best one.[61]

This kind of human–computer decision-making requires regulation for two reasons. First, if the ADM system is biased in any way, this bias is likely to translate straight into the human's decision.[62] In these examples, the human decision-maker is not privy to the list of rejected options and does not have sufficient information to consider possible options the computer might have missed. Therefore, even with the best of intentions, humans do not have sufficient oversight to spot instances of bias or inaccuracy in the ADM system and to mitigate these to produce a fair result.[63]

In addition, humans have been shown to engage in automation bias when presented with a decision generated by a machine.[64] Automation bias refers to the tendency of humans to ignore or fail to seek out contrary evidence when presented with an automated decision. Humans are likely to assume the automated decision presented is correct.[65] Extensive studies have been done on automation bias in the aviation sector, where ADM processes are heavily relied upon in complex environments such as route configuration or air traffic control.[66] In one study on

60 At 2.

61 At 2.

62 At 5.

63 At 5.

64 Eugenio Alberdi and others "Why are People's Decisions Sometimes Worse with Computer Support?" (paper presented to Computer Safety, Reliability, and Security 28th International Conference, Hamburg, Germany, 15–18 September 2009) at 2.

65 At 1.

66 Cummings, above n 57, at 3–4.

automation bias, pilots were given an automated solution to a flight route configuration.[67] Over 40 per cent demonstrated over-reliance on the solution, completing none of their own decision-making and accepting flight plans that were significantly substandard.[68]

Therefore, in both fully automated and partially automated decision-making systems, regulation is needed to ensure that the systems are accurate and free from bias. The design of human–computer decision-making systems and the phenomenon of automation bias means that we cannot trust that the mere presence of a human, often with limited oversight into the ADM process, will mitigate the risk of a biased or inaccurate result.[69] Given the inadequacy of New Zealand's regulatory environment, urgent attention should be paid to developing regulations to provide much needed oversight of ADM systems, even where a human is involved.

IV THE GENERAL DATA PROTECTION REGULATION (GDPR) AS AN EXEMPLAR REGULATORY INSTRUMENT

This Part will analyse the regulatory solutions proposed in the GDPR to address ADM and what this regime might offer to regulators looking to legislate ADM in a New Zealand context. It will conclude that art 22 of the GDPR, which covers ADM, fails to strike an appropriate balance between the interests of organisations utilising ADM and the interests of individuals. Instead of creating a holistic regulatory framework designed to manage this balance, art 22 merely creates a strict prohibition on ADM that applies only to an extremely narrow range of cases.

A Introduction to the GDPR and Art 22

The GDPR entered into force in the European Union in April 2016, applying from May 2018.[70] The Regulation applies to all businesses doing business in the European Union, and in some circumstances, to businesses operating outside of the Union.[71] The Regulation covers the relationship between data controllers (parties with control over the means and purposes for which personal data is processed) and data subjects (individuals whose personal data is being used in these processes).[72] For consistency this article will use the word "organisation" to describe a data controller, and "individual" or "person" when referring to a "data subject".

---

67 At 4.

68 At 4.

69 Alberdi and others, above n 64, at 1.

70 Russell McVeagh *Information Sheet for GDPR* (May 2018).

71 Article 3.

72 Articles 4(1) and (4).

Article 22 of the GDPR explicitly addresses ADM. It provides that individuals shall have the:[73]

> … right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

Article 22(1) operates as a general prohibition on "solely … automated processing" for decisions that "significantly affect" individuals. There are three exceptions to the prohibition, covered in art 22(2). These are:

(a) where the decision is necessary for entering into or performing a contract;
(b) where the decision is authorised by Member state law;
(c) and where the decision is based on the individual's explicit consent.

Article 22 was initially misinterpreted as a "right to an explanation provision" for automated decisions.[74] This interpretation has since been proven incorrect by legal experts who affirmed that art 22 is a prohibition clause on the use of ADM in certain contexts.[75]

1 Effect of arts 13 and 15 in regards to ADM

Article 22 is the best option for a remedy in the case of automated decision making. Articles 13 and 15 of the GDPR do offer a possible avenue through which an explanation of ADM might be obtained, however this is likely to be limited. Article 13 requires that an organisation undertaking ADM must notify the affected individual or individuals that ADM will take place and provide "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject".[76] Article 15 provides that individuals have a right of access to the same information.

However, as Wachter and others and Edwards and Veale have pointed out, the explanation that is required by arts 13 and 15 is merely an ex ante explanation.[77] It requires individuals to be told general information about how the automated decision will be made. There is no requirement

---

73 Article 22(1).

74 House of Lords Select Committee on Artificial Intelligence *AI in the UK: ready, willing and able?* (HL Paper 100, Report of Session 2017–2019, 26 April 2018) at [101].

75 Sandra Wachter, Brent Mittelstadt and Luciano Floridi "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation" (2017) 7 IDPL 76 at 77–78.

76 GDPR, art 13(2)(f).

77 Lilian Edwards and Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You Are Looking For" (2017) 16 Duke Law & Technology Review 18 at 52.

that individuals be provided with an explanation of the logic of the decision as it relates to their case, such as the specific reasons they were denied a loan, for example.[78]

The scope of arts 13 and 15 and their application to ADM cases will likely be elucidated further as case law emerges. Current interpretations support the view that these two articles only allow individuals to access a general explanation of decision-making systems and their foreseen consequences, rather than an in-depth explanation of the decision process as it relates to their case, which could enable an individual to contest a decision. We therefore must return our attention to art 22 because it is the provision that shows the most promise in offering a remedy to individuals in respect of ADM under the GDPR.

B Does Art 22 Effectively Regulate ADM?

This sub-Part analyses several areas of art 22, the prohibition clause. This will include the two limbs of the legal test: whether the decision was made "based solely on automated processing"; and whether the decision produced legal effects or "similarly significantly" affected an individual, as well as two of the exceptions to the general rule: the exception where the individual consents; and the contract exception. Ultimately, this sub-Part will conclude that art 22 fails to balance the interests of organisations with the rights of individuals and does little to regulate ADM as used in practice.

1 Limb one of the art 22 test: "Solely Automated" decisions

For art 22 to apply, a decision must be "based solely on automated processing".[79] The word "solely" requires interpretation. The Article 29 Data Protection Working Party, a European Union advisory body formed to provide guidelines on the interpretation of art 22, defined "based solely" to mean that the decision was made "without human involvement".[80] This is supported by the synonyms of "solely": "entirely" or "exclusively",[81] and by the ordinary usage interpretation, as a lay person would likely assume "based solely on automated processes" to mean that the decision needs to be made *only* by a machine.

The Working Party guidelines also specify that superficial human involvement in decision-making does not suffice and that humans must have "meaningful" oversight with sufficient authority and competence to change the result in order for a decision to not be made "solely" by a machine.[82] The underlying assumption in this provision is the same assumption made by the

78 At 52; and Wachter, Mittelstadt and Floridi, above n 75, at 83.

79 Article 22(1).

80 *Working Party 29 Guidelines*, above n 6, at 8.

81 "Definition of 'solely'" Collins Dictionary <www.collinsdictionary.com/dictionary>.

82 *Working Party 29 Guidelines*, above n 6, at 21.

Justice Select Committee on the New Zealand Privacy Bill.[83] The assumption is that through human involvement, the risks associated with ADM can be avoided. The case of *State v Loomis* will be used to build further on the arguments mentioned in Part III; that restricting regulation to solely automated decisions ignores the obvious risks of involving humans in ADM processes, prohibiting the most extreme uses of ADM while providing no oversight over other, similarly impactful, uses of the technology.

State v Loomis concerned the use of the automated tool COMPAS to inform judicial decisions on sentencing.[84] COMPAS was designed for the United States Department of Corrections by a private company, Northpointe Inc.[85] The algorithm was developed using past data on criminal recidivism among individuals.[86] It used information from the defendant's criminal file and an interview with the defendant to produce three "risk scores".[87] In Mr Loomis' case the trial judge used his COMPAS score, along with other factors including the seriousness of Mr Loomis' crime and his criminal history to deliver a relatively lengthy sentence that reflected an extremely high risk of reoffending.[88] Mr Loomis argued that COMPAS had increased his risk profile on the basis of his male gender and therefore that the Court was engaging in unconstitutional discrimination on the basis of sex.[89]

The primary issue when applying art 22(1) to the facts of *State v Loomis* is whether the trial Judge's decision in this case was based "solely" on the automated processing completed by COMPAS. The Judge's decision on sentencing would likely not be considered to be "solely automated", under the definition identified earlier. The Judge took into account other factors in addition to the COMPAS score, including the defendant's criminal history and the seriousness of the crime.[90] The Judge was not bound to apply the COMPAS rating strictly; they had the power and competence to change the decision outcome.[91]

This case clearly highlights the two problems with relying on human involvement to alleviate the need for regulation of ADM systems. First, human decision-makers using ADM tools

---

83 Privacy Bill (select committee report), above n 44, at 39–40.

84 *State v Loomis* 881 NW 2d 749 (Wis 2016) at [4].

85 At [6].

86 At [7].

87 At [6]–[7].

88 At [7].

89 At [13]–[14].

90 At [44].

91 At [5]; and *Working Party 29 Guidelines*, above n 6, at 21.

frequently do not have oversight over the ADM process itself. The trial Judge in *State v Loomis* was not aware of how COMPAS generated its risk scores nor what weighting was assigned to the different characteristics and attributes of the defendant's profile.[92] The power of the Judge only included the ability to take the COMPAS score, consider other factors and make the final sentencing decision. The underlying assumption of the "solely automated" limb is therefore undermined because the Judge had no oversight over the ADM which would have enabled them to identify and mitigate potential biases within it.[93] Here, the human involvement provided no additional assurance that the COMPAS algorithm was fair and accurate. Secondly, human–computer decision-making systems such as COMPAS cause automation bias in human decision-makers. Even if the Judge in *State v Loomis was* presented with evidence that the COMPAS result was unfair, evidence shows that it is likely that the Judge would defer to the automated decision, rather than challenge it.[94]

The narrow scope of art 22 therefore has the effect of restricting the provision to apply to a very limited number of cases where there is no human involvement.[95] The result is that significant uses of ADM processes, such as the COMPAS algorithm, are subject to no regulatory oversight.

2 Limb two of the art 22 test: Legal effects or "similarly significant" effects

Under the second limb of the art 22 test, a decision must be one with legal effects or "similarly significant" effects for the individual. However, this definition of "similarly significant" effects imposes too big a restriction on the range of situations where art 22 can apply.

The Article 29 Working Party defined "legal effects" as decisions that affect a person's legal rights, such as the freedom to vote, entitlement to social benefits, or the granting of citizenship.[96] They found that to be "similarly significant" to these legal effects, a decision must have the potential to significantly affect the choices, behaviour or circumstances of the individual, have a prolonged impact on the individual or lead to the discrimination of the individual.[97] The Working Party stated that decisions affecting access to healthcare or employment would fall into this category.[98] On the other hand, the Working Party used online advertising as an example of ADM

92 *State v Loomis*, above n 84, at [21].

93 At [21].

94 Cummings, above n 57, at 3.

95 *Working Party 29 Guidelines*, above n 6, at 21.

96 At 21.

97 At 21.

98 At 22.

that would not typically "similarly significantly" affect individuals.[99] As an exception to this general rule, the Working Party said that an automated advertisement for a high interest loan that targeted an individual in a precarious financial situation could meet the threshold of producing a "similarly significant" effect.[100]

This definition of "similarly significantly affects" is based on a binary model of one decision directly causing a significant effect on a person's life. In the loan example, this "tipping point" where the effect is created is the moment where the individual takes out the loan after seeing the advertisement. However, this model fails to reflect the "immersion effect" that individuals experience within heavily personalised, deliberately constructed online environments.[101] On social media platforms like Facebook, hundreds of decisions are made almost instantly about the content that should be presented to individuals in order to optimise their engagement on the platform, thereby increasing "clicks" or site purchases.[102]

The effect of these environments on individuals is insidious and gradual.[103] It will not always be possible to identify the exact "tipping point" where one decision had a significant effect on an individual. For example on Facebook, advertisements on dieting and methods to reduce body weight may be targeted at a person based on the fact that their past activity on the site reveals that they are someone who exhibits symptoms of anxiety and negative body image.[104] Individually, the decision to show each of these ads is unlikely to be found to produce a "significant" effect for the individual.[105] The individual might ignore the ads sometimes, whilst other times the ads may lead them to restrict their food intake. However, continued exposure to this advertising could "snowball" and lead the individual to develop an eating disorder.[106] The contraction of a mental

---

99 At 22.

100 At 22.

101 Eric Johnson "Google and Facebook have become 'antithetical to democracy,' says The Age of Surveillance Capitalism author Shoshana Zuboff" (20 February 2019) Vox <www.vox.com>.

102 James Bridle "The Age of Surveillance Capitalism by Shoshana Zuboff review – we are the pawns" *The Guardian* (online ed, London, 2 February 2019).

103 Johnson, above n 101.

104 SC Matz and others "Psychological targeting as an effective approach to digital mass persuasion" (2017) 1144 PNAS 12714 at 12717.

105 Michael Veale and Lilian Edwards "Clarity, surprises and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling" (2018) 34 CLSR 398 at 401.

106 At 402. Numerous studies have demonstrated that online advertising, tailored to individuals' personality traits can be effective in changing individuals' behaviour. See for example Matz, above n 104, at 12717; and Michael Kosinki, David Stillwell and Thore Graepel "Private traits and attributes are predictable from digital records of human behavior" (2013) 110 PNAS 5802 at 5805.

illness is clearly a significant effect on an individual. However, because the effect is developed gradually through continuous exposure to the ads it would be impossible to prove that one automated decision had produced that effect and thus that the ADM use should be prohibited under art 22.[107]

In summary, art 22 creates an arbitrary distinction between effects immediately produced by a single automated decision and effects which manifest gradually through long term exposure to a series of automated decisions. However individuals in both scenarios are likely to experience the same exploitation of their particular vulnerabilities through ADM that in turn produces similarly significant effects on their lives. This makes the distinction unjustifiable.

3 Exception where an individual consents

Even where limbs one and two of the legal test under art 22 are satisfied, three broad exceptions to the general rule limit the article's application. The first exception is where the decision-making is completed with the individual's "explicit consent".[108] This exception is consistent with the GDPR principle of enabling individuals to exercise control over their own data.[109] However the exception requires closer analysis. Across all areas of privacy law, increasing scepticism has been raised about the effectiveness of consent as a mechanism for regulating privacy intrusive data management systems.[110]

"Consent fatigue" is a term that has been coined to describe the reduced ability of individuals to offer genuine consent in new online environments.[111] When individuals are constantly asked to give consent on online platforms through click-wrap contracts which are impossibly long and stand in the way of a social or entertainment reward, people are more likely to blindly offer their consent, often without reading the terms.[112] If users do not understand what they are consenting to, have no option to opt-out of processing without being excluded from a digital platform or are

---

107 *Working Party 29 Guidelines*, above n 6, at 22.

108 Article 22(2)(c).

109 Article 5.

110 Luis Alberto Montezuma and Tara Taubman-Bassirian "How to avoid consent fatigue" (29 January 2019) International Association of Privacy Professionals <https://iapp.org>; and Brian Chen "Getting a Flood of GDPR-Related Privacy Policy Updates? Read Them" *The New York Times* (online ed, New York, 23 May 2018).

111 The Workshop *Digital Threats to Democracy: Report on Qualitative Interviews* (2019) [*Digital Threats to Democracy*] at 12.

112 At 47 (Privacy Commissioner John Edwards submission).

unable to make "rational trade-offs" between the privacy risk and associated benefits,[113] consent cannot operate as an effective "shield" to protect users from the negative effects of privacy intrusive systems.[114]

Article 7 sets out the conditions for consenting to processing in the GDPR. Consent for processing must be presented distinctively in an "intelligible and easily accessible form, using clear and plain language".[115] However, as discussed, it remains to be seen whether placing these kind of requirements around how consent should be obtained will actually overcome the core issue of consent fatigue and reinstate consent as a meaningful check operating to prevent individuals from exposing themselves to exploitative and harmful uses of their personal data.

Article 7(4) of the GDPR offers one area which could help to ensure consent is meaningful. This article states that when determining whether consent has been "freely given", consideration should be given to whether "the performance of a contract … is conditional on consent to the processing of personal data that is not necessary for the performance of that contract". This technically makes it more difficult for organisations to engage in ADM using personal data where the use of that data is not necessary for the contract with the customer to be performed. However the strength of this provision has not yet been tested.

Plaintiff Maximilian Schrems is contending before the Vienna Regional Council that Facebook operates in breach of art 7(4) of the GDPR because the processing of individuals' personal data for the purposes of supplying advertising on Facebook is not "necessary" for the performance of Facebook's contracts with its users.[116] Facebook has argued that Facebook users "order" advertising when they sign up for Facebook and therefore that the processing of their personal data for the purposes of supplying advertising is within the scope of the contract. Based on the outcome of this case, it remains to be seen whether art 7(4) will provide adequate protection for users against the exploitation of their personal data for purposes tangential to their contracts with data controllers such as Facebook. Given that ADM may be used in the future to provide individuals with critical services such as bank loans and health insurance, there is a need for regulations to ensure that individuals are able to consent to the use of ADM without being subject

113 Marcin Betkier "Moving Beyond Consent in Data Privacy Law: An Effective Privacy Management System for Internet Services" (PhD thesis, Victoria University of Wellington, 2018) at 37–38.

114 *Digital Threats to Democracy*, above n 111, at 67.

115 Article 7(2).

116 For the case history and a summary see Global Freedom of Expression Columbia University "Maximilian Schrems v. Facebook Ireland Limited" <https://globalfreedomofexpression.columbia.edu/cases>.

to additional processing of their data where that data may not be necessary for the provision of the service.[117]

4 Exception for entering into a contract

Another exception to art 22(1) is where the use of ADM "is necessary for entering into, or performance of, a contract between the data subject and a data controller".[118] This exception gives rise to two problems: first, it can be interpreted very broadly; and secondly it creates inconsistent standards on ADM use for organisations.

According to the Article 29 Working Party, to prove that ADM was "necessary" under this exception, the organisation or data controller must demonstrate a lack of an equally effective and less privacy-intrusive means to achieve the same commercial objective.[119] The Working Party used the example of a business that receives thousands of applications for a job. Finding it impractical to sift through all the applications manually, the business automates the process of CV screening. According to the Working Party, this example would meet the definition of "necessary"; there is no equally effective, less privacy means to sort the applications. [120]

The broad interpretation of "necessary" in article 22 as *commercially* necessary means that this exception has the potential to apply to a large range of commercial interactions between organisations and individuals. For example, if an employer has thousands of employees and their employment contracts require a performance review, the employer could argue that ADM is "necessary" for performance of the contract because it would be impractical to complete performance reviews without it.[121] Therefore the exception for contract creation makes substantial allowance for organisations and their need for efficiency, without adequately protecting the rights of individuals. Many of the situations where individuals most need protection from the potential harms of ADM are where they are attempting to enter into a contract or enforce performance of a contract against a dominant party.[122]

Not only does the contract exception inadequately protect individuals within interactions where they typically are the more vulnerable party, the exception also creates a hierarchy of interactions within the regime. This exception privileges contract creation and performance over

---

117 Article 21; Betkier, above n 113, at 38–39; and O'Neil *Weapons of Math Destruction*, above n 13.

118 Article 22(1)(a).

119 *Working Party 29 Guidelines*, above n 6, at 23.

120 At 23.

121 At 23.

122 O'Neil *Weapons of Math Destruction*, above n 13.

other equally important interactions.[123] ADM, when used fairly, may be used to further many worthy objectives such as encouraging the provision of credit from lenders, or ensuring adequate resourcing and accurate diagnosis at public hospitals.[124] Therefore it is unclear why the efficiency of contracting has been privileged in art 22 over other legitimate commercial or public interest objectives.

V ALTERNATIVE LEGISLATIVE SOLUTIONS

Given that the GDPR does not offer a clear solution to the issue of ADM in the New Zealand context, this Part will explore two alternative legal solutions which could be adopted in to strike a balance between the benefits and risks of ADM for organisations and individuals. These include a "right to an explanation" and a compulsory consent process which could be used to vet organisations' proposed uses of ADM.

A Is a "Right to an Explanation" a Solution?

A legislative requirement of transparency or a "right to an explanation" is frequently cited as a desirable regulatory solution to the challenge of ADM.[125] Legislation could force organisations to disclose the reasoning behind automated decisions to individuals, which individuals could then rely on to contest the decision if it was unfair.[126] Greater transparency not only appears to reduce the likelihood that ADM processes will be unfair, but it also preserves human dignity by enhancing individuals' understanding of ADM processes and thus trust in these systems.

However, a "right to an explanation" also poses several significant problems. First, private companies who develop their own ADM systems, or commission them, are the owners of these systems.[127] The commercial sensitivity of ADM systems means that companies are typically extremely reluctant to disclose the internal workings of their algorithms for fear of losing the advantage over their competition.[128] The proprietary and confidential nature of ADM systems therefore makes it very difficult for a regulator to require the disclosure of such systems to the

123 GDPR, art 22(2)(b).

124 O'Neil *Weapons of Math Destruction*, above n 13, at 141; and Charles Towers-Clark "The Cutting-Edge of AI Cancer Detection" *Forbes* (online ed, New Jersey, 30 April 2019).

125 Julia Angwin "Make Algorithms Accountable" *The New York Times* (online ed, New York, 1 August 2016).

126 Vanessa Blackwood "Algorithmic transparency: What happens when a computer says 'no'?" (29 November 2017) Privacy Commissioner Blog <www.privacy.org.nz/blog>; and Privacy Commissioner "Submission to the Justice and Electoral Select Committee on the Privacy Bill 2017", above n 43, at 29–30.

127 Gavaghan and others, above n 41, at 41; and Copyright Act 1994, s 21(1) and (3)(a).

128 *State v Loomis*, above n 84, at 5.

public for review.[129] Secondly, even if private algorithms are disclosed, complete transparency over machine learning systems may provide no meaningful information about ADM processes.[130] Many ADM systems are so complex that they cannot identify the factors or reasons for their decisions in a way that an expert could understand, much less a lay person.[131] Developments in explainable artificial intelligence (AI) and counterfactual systems may offer a solution to this issue of "explainability" in the future.

Explainable AI or "XAI" could help to resolve the issue of opaque ADM systems.[132] Explainable AI systems are developed to ensure that the system is capable of explaining its reasoning in a way that is understandable to humans.[133] One example of a partial use of explainable AI is the "why am I seeing this?" feature on the Facebook Newsfeed.[134] This feature tells Facebook users why a certain advertisement or post is being shown to them.[135] Reasons provided include things like "you are 18–24 years old" or "you have interacted with posts from x friend more than posts from others".[136] However developing explainable AI typically involves a trade off as producing fully explainable systems often requires reducing the complexity of the system, and thus its accuracy.[137] For example if the Facebook advertising system were to be made fully explainable, the multidimensional explanations that would need to be given for every ad would be incredibly long and complex. The only solution to make this more accessible would be reducing the complexity of the algorithm that decides what ads are shown, deciding what ads to show based on only two or three factors that can easily be explained, rather than using a highly sensitive, personalised algorithm that could recommend you ads tailored directly to your interests.

Counterfactual explanations represent another possibility for increasing the transparency of ADM systems, without some of the drawbacks of fully explainable systems. Experts such as Sandra Wachter have proposed that counterfactual explanations can be used to avoid disclosing

---

129 O'Neil *Weapons of Math Destruction*, above n 13, at 10.

130 Edwards submission, above n 80, at 23.

131 Gavaghan and others, above n 41, at 42.

132 Louise Matsakis "What Does A Fair Algorithm Look Like?" (10 November 2018) Wired <www.wired.com>.

133 Ron Schmelzer "Understanding Explainable AI" *Forbes* (online ed, New Jersey, 23 July 2019).

134 Alex Hern "Why am I seeing this? New Facebook tool to demystify Newsfeed" *The Guardian* (online ed, London, 1 April 2019).

135 Hern, above n 134.

136 Author's own Facebook newsfeed.

137 Zerilli and others, above n 7, at 664.

the content of proprietary algorithms whilst providing individuals with an explanation for ADM decisions that they can understand.[138]

Counterfactual explanations provide an alternative scenario, the "closest possible world" in which the decision outcome would have been different.[139] For example, if a person was denied a loan because their income was under NZD 50,000 per year, then the counterfactual explanation for the decision would be: if your income had been greater than NZD 50,000, this loan request would have been accepted.[140] Typically there will be multiple factors which weigh into a decision made by an ADM system and separate counterfactual explanations can be provided for all of these factors.[141] Counterfactuals can provide individuals with an explanation that they understand and can act on, for instance individuals may contest the decision if it is based on incorrect or illegitimate information, or may alter their behaviour or circumstances to obtain a more favourable outcome in the future.[142]

However, as Wachter and others point out, counterfactuals provide only some pieces of the puzzle. Counterfactuals do not reveal *how* an ADM system is working and therefore provide no means of ensuring that rules or patterns are being correctly identified and applied.[143] In addition counterfactuals do not produce sufficient statistical analysis of the different data combinations to reveal bias or discrimination against individuals with certain attributes.[144]

Neither counterfactuals nor explainable AI are currently sufficiently utilised by organisations to the point where "an explanation" of ADM could be feasibly mandated by law.[145] However, even if they were, the "right to an explanation" framework still faces some significant philosophical issues. In most rights frameworks, the onus to enforce a right falls on the individual.[146] It would therefore likely be up to the individual to demand that their right to an

---

138 Sandra Wachter, Brent Mittelstadt and Chris Russell "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR" (2018) 31 Harv J Law and Tec 841 at 847.

139 At 847.

140 At 847.

141 At 842.

142 At 881–883.

143 At 884.

144 At 884.

145 At 884; Jakko Kemper and Daan Kolkman "Transparent to whom? No algorithmic accountability without a critical audience" (2018) 22 Information, Communication & Society 2081 at 2086; and Zerilli and others, above n 7, at 4.

146 Mike Ananny and Kate Crawford "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability" (2016) 20 New Media & Society 973 at 977–978.

explanation be fulfilled and to contest an automated decision if that explanation demonstrates the decision is unfair.[147] Given the information overload that individuals are already subject to, it is possible that increased transparency about ADM processes will result in no meaningful improvement in individuals' *understanding* of these processes.[148] Veale and Edwards warn that through seeking greater explainability of algorithms we run the risk of creating a "'meaningless transparency' paradigm", similar to the meaningless consent phenomena discussed earlier.[149]

"Explainability" as a regulatory solution also assumes that effective punishment mechanisms are available to hold organisations accountable for any unfair or corrupt ADMs that are uncovered.[150] As Ananny and Crawford have argued, if transparency produces no meaningful effects, public cynicism and distrust may be increased, undermining the initial purpose of implementing transparency.[151] Online monopolies such as Facebook and Alphabet have increasingly demonstrated an ability to withstand extremely severe ethical and legal scandals and emerge relatively unscathed.[152] It is therefore possible that a regulatory solution based on transparency could wind up being ineffective against these dominant companies.

B ADM "Consents": A New Proposal

In this sub-Part, I will outline an alternative regulatory solution which could be used in the New Zealand context to address the issue of ADM. My solution proposes a broader prohibition on ADM than art 22; it would prohibit all uses of ADM that affect individuals in any way. To ensure that ADM can still be used where it will be useful, and has been assessed for fairness and accuracy, I propose that the prohibition accompany a "consent" process, similar to the resource consent process used under the Resource Management Act 1991 (RMA).[153] This proposal is similar to the Swedish Data Act 1973 which prohibited personal data processing,[154] but permitted

147 At 8.

148 At 8–9.

149 Edwards submission, above n 80, at 23.

150 Ananny and Crawford, above n 146, at 976.

151 At 6.

152 Edwards submission, above n 80, at 41; James Vincent "Google hit with €1.5 billion antitrust fine by EU" (20 March 2019) The Verge <www.theverge.com>; Jeff Desjardins "How Google retains more than 90% of Market Share" (24 April 2018) Business Insider <www.businessinsider.com>; and Julia Carrie Wong "The Cambridge Analytica scandal changed the world – but it didn't change facebook" *The Guardian* (online ed, London. 18 March 2019).

153 Resource Management Act 1991, s 9 [RMA].

154 Sören Öman "Implementing Data Protection in Law" (2004) 47 Scandinavian Studies in Law 389 at 390.

licences for processing to be issued by a Data Inspection Board if organisations applied for them.[155]

Under this alternative regulatory solution, a centralised tribunal could be established to grant "AI consents" to organisations, permitting the use of ADM to produce decisions which affect individuals. The Tribunal responsible for issuing consents could use a broad balancing test to decide whether ADM should be used on the basis of each individual case. The test could be:

> … whether an equal balance between the interests of the organisation and the individuals subject to the decision making is achieved by granting a consent.

The Tribunal could consider a range of factors when assessing whether the right balance is achieved. Relevant considerations could include: the significance of the decision for the individual; the measures taken by the organisation to ensure transparency and fairness in the system; whether the individual has consented to the processing; and the benefit to the organisation if consent was granted. Once issued, decisions would be binding on the organisation, either prohibiting them or permitting them to undertake the processing. The result of the consent application could still be appealed to the High Court by either the organisation or an affected individual.[156]

This solution offers numerous benefits. First, it enables organisations to reap the benefits of ADM, but places the onus on them to ensure that their ADM processes do not subject individuals to discrimination. The benefits that organisations may reap from ADM are substantial, including increased efficiency and increased sales due to improved service offerings.[157] On the other hand, ADM creates substantial risks for individuals, without any corresponding gain. Evidence suggests that ADM systems frequently exhibit biases which reflect and exacerbate historic patterns of inequality and marginalisation.[158] It is therefore justifiable to place the cost of reducing the risk of potentially discriminatory results on the party who will reap the benefits of ADM.

In addition, by using a market modality, this solution incentivises companies to implement good AI practices in order to successfully gain ADM consents and maintain an edge over their competition.[159] Companies are incentivised to use counterfactuals and explainable AI because using these practices would increase their chances of demonstrating to the Tribunal that their

---

155 At 390.

156 This process would be similar to how decisions of the Employment Relations Authority may be appealed to the Employment Court: Employment Relations Act 2000, s 179(1).

157 Varian, above n 26.

158 O'Neil *Weapons of Math Destruction*, above n 13.

159 Lawrence Lessig "The Law of the Horse: What Cyberlaw Might Teach" (1999) 113 Harv L Rev 501 at 507.

systems are fair and offer meaningful transparency to individuals. Longer term, these incentives could also drive the creation of independent "AI Auditing" firms which would use accepted professional processes to assess the validity and fairness of ADM systems for their clients.[160]

However, one drawback of this solution could be that the process of obtaining a consent becomes slow. Indeed, the analogous resource consent process is often lengthy.[161] This issue could be avoided through implementing similar timeframe requirements for decision-making by the Tribunal as are used in the RMA.[162] Consultation with affected organisations could also be used to communicate and convince organisations of the benefits that they are likely to derive from participation in the consent process. These benefits include: potential increases in the accuracy of their ADM systems and a reduction of the risk that an organisation is subject to negative public attention as a result of unwittingly implementing discriminatory ADM systems.

VI CONCLUSION

Commercial values such as enabling efficiency, cost cutting and commercial competitiveness are frequently cited by both governments and corporations as neutral goods that regulation ought to support. However regulation should also uphold the rights of individuals, especially the right to be free from discrimination. As new technologies such as ADM penetrate our society at increasing rates, altering the decision-making processes which govern our lives, it is more important than ever before that regulation is used to balance commercial interests with the interests of individuals. An AI consent process is one possible way this might be achieved.

---

160 Chowdhury and Mulani, above n 40.

161 Nicole Pryor "Christchurch council consents too slow - Key" (17 June 2013) Stuff <www.stuff.co.nz>.

162 RMA, s 21.