Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

A vulnerability analysis: Theorising the impact of artificial intelligence decision-making processes on individuals, society and human diversity from a social justice perspective

Tetyana (Tanya) Krupiy

Tilburg University, Montesquieu Building, Room 808, Prof. Cobbenhagenlaan 221, Tilburg, North Brabant, 5037 DE, the Netherlands

ARTICLE INFO

Keywords:
Artificial intelligence
Data science
Decision-making process
Social justice
Human diversity
Vulnerability theory
Feminism
Queer legal theory
Critical disability theory

ABSTRACT

The article examines a number of ways in which the use of artificial intelligence technologies to predict the performance of individuals and to reach decisions concerning the entitlement of individuals to positive decisions impacts individuals and society. It analyses the effects using a social justice lens. Particular attention is paid to the experiences of individuals who have historically experienced disadvantage and discrimination. The article uses the university admissions process where the university utilises a fully automated decision-making process to evaluate the capability or suitability of the candidate as a case study. The article posits that the artificial intelligence decision-making process should be viewed as an institution that reconfigures the relationships between individuals, and between individuals and institutions. Artificial intelligence decision-making processes have institutional elements embedded within them that result in their operation disadvantaging groups who have historically experienced discrimination. Depending on the manner in which an artificial intelligence decision-making process is designed, it can produce solidarity or segregation between groups in society. There is a potential for the operation of artificial intelligence decision-making processes to fail to reflect the lived experiences of individuals and as a result to undermine the protection of human diversity. Some of these effects are linked to the creation of an ableist culture and to the resurrection of eugenics-type discourses. It is concluded that one of the contexts in which human beings should reach decisions is where the decision involves representing and evaluating the capabilities of an individual. The legislature should respond accordingly by identifying contexts in which it is mandatory to employ human decision-makers and by enacting the relevant legislation.

E-mail address: t.krupiy@uvt.nl

https://doi.org/10.1016/j.clsr.2020.105429

Erica Curtis, a former admissions evaluator at Brown University in the United States, has noted that she evaluated each student's application consisting of standardised test scores, the transcript, the personal statement, and multiple supplemental essays within a twelve-minute timeframe.[1] Arguably, this is a very short period of time within which an admissions officer can evaluate the applicant's personality and academic qualities holistically.[2] The time constraints create a possibility that the admissions officer may fail to detect the applicants' capabilities or how societal barriers diminished their ability to realise their potential. Another concern with human decision-making is that the decision-maker officer may act arbitrarily in the course of exercising discretion[3] by putting different weight on comparable attributes that cannot be measured. What is more, an admissions officer could treat applicants on an unequal basis due to being influenced by conscious or unconscious biases.[4] Advances in artificial intelligence (hereinafter AI) technology give rise to a discussion whether organisations should use AI systems to select applicants for admission to university.[5] Technology companies market AI systems with a capability to predict the candidates' performance and to follow a decision-making procedure as possessing the capacity to eliminate bias and to improve decision-making.[6] The computer science community is now working on embedding values, such as fairness, into the AI decision-making procedure.[7] Daniel Greene and colleagues view the focus on achieving fairness by incorporating values into the design of the system as short-sighted.[8] The attention on how to embed fairness into the decision-making procedure of a technical system side-lines the discussion how the employment of AI decision-making processes impacts on achieving social goals, such as social justice and 'equitable human flourishing.'[9] Virginia Eubank's work underscores the importance of investigating how the use of AI decision-making processes impacts individuals and society. Her interviews with affected individuals who applied to access state benefits in the state of Indiana in the United States[10] demonstrate that the employment of AI decision-making processes can lead to the deepening of inequality,[11] to social sorting[12] and to social division.[13] The enquiry is particularly pertinent given the fact that not all sources report adverse outcomes. The British Universities and Colleges Admissions Service asserts that in its pilot project an algorithmic process selected the same pool of applicants to be admitted to universities as admissions officers; the organisation did not reveal the algorithm's design and operation procedure.[14]

The present paper explores some of the hitherto unresolved longstanding societal problems and new issues the employment of AI decision-making processes raises. It contributes to existing literature by proposing that an AI decision-making process should be understood as an institution. The AI decision-making process reconfigures relationships between individuals as well as between individuals and institutions. The paper examines some of the values and types of institutional arrangements the employment of AI decision-making processes embeds into society. This issue is significant. The Council of Europe Committee of Ministers stated that when data-driven technologies operate 'at scale' their operation prioritises certain values over others.[15] The assertion of the Council of Europe Committee of Ministers that data-driven technologies reconfigure the environment in which individuals process information[16] should be extended to encompass

Acknowledgments: I would like to thank Professor Corien Prins for her feedback on the draft version of this article. I am grateful to Atieno Samandari, Stu Marvel, Professor Martha Albertson Fineman, Professor Nicole Morris and Professor Paul Myers for their feedback on a presentation which formed the foundation for this article. Additionally, I wish to thank scholars who asked stimulating questions during the Ethics of Data Science: Addressing the Future Use and Misuse of Our Data Conference, the BIAS in Artificial Intelligence and Neuroscience Transdisciplinary Conference, and the Media & Space: The Regulation of Digital Platforms, New Media & Technologies Symposium where I presented my ongoing work.

[1] Joel Butterly, '7 Admissions Officers Share the Things They Never Tell Applicants' (Insider Inc., 2018) <https://www.businessinsider.com/7-things-college-admissions-officers-wishevery-applicant-knew-2018-2?international=true&r=US&IR=T> accessed 26 June 2019

[2] Ibid

[3] Mark Bovens and Stavros Zouridis, 'From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control' (2002) 62 Public Administration Review 174, 181

[4] Josh Wood, '"The Wolf of Racial Bias": the Admissions Lawsuit Rocking Harvard' *The Guardian* (London 18 October 2018) <https://www.theguardian.com/education/2018/oct/18/harvard-affirmative-action-trial-asian-american-students> accessed 10 March 2019

[5] Moritz Hardt, *How Big Data is Unfair: Understanding Unintended Sources of Unfairness in Data Driven Decision-making* (Medium Corporation 2014)

[6] Ekta Dokania, 'Can AI Help Humans Overcome Bias?' *The Seattle Globalist* (Seattle 22 May 2019) <https://www.seattleglobalist.com/2019/05/22/can-ai-help-humans-overcome-bias/83957> accessed 3 March 2019

[7] Aditya Krishna Menon and Robert C Williamson, 'The Cost of Fairness in Binary Classification' (2018) 81 Proceedings of Machine Learning Research 1, 10

[8] Daniel Greene, Anna Lauren Hoffman and Luke Stark, *Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning* (The Proceedings of the 52nd Hawaii International Conference on System Sciences, Hawaii, 2019) 2122

[9] Ibid

[10] Virginia Eubanks, *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* (St Martin's Press 2018) 10

[11] Ibid 204

[12] Ibid 122

[13] Ibid 196-97

[14] Ben Jordan, *Minimising the Risks of Unconscious Bias in University Admissions: 2017 Update on Progress* (Universities and Colleges Admissions Service 2017) 11

[15] Council of Europe Committee of Ministers, 'Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes Decl(13/02/2019)1' (1337th meeting of the Ministers' Deputies, Council of Europe 2019) <https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b> 15 February 2019

[16] Ibid.

the relationships individuals have with each other and with the institutions. The article examines some of the types of social transformations that the use of AI decision-making processes across domains will accentuate. While the design of AI decision-making processes will shape whether their operation gives rise to solidarity or segregation, there is a potential for these systems to adversely affect individuals who have historically experienced discrimination, disadvantage, disempowerment and marginalisation. The provisions in international human rights treaties prohibiting discrimination provide a non-exhaustive list of individuals who experience discrimination, exclusion, oppression, disempowerment and disadvantage.[17] The characteristics such individuals possess include sex, gender identity, sexual orientation, age, ethnicity, race, colour, descent, language, religion, political or other opinion, national or social origin, property, birth and disability amongst others.[18]

The university admissions process serves as a case study for contextualising the discussion in the present paper. One of the reasons for using a case study for focusing the discussion is that an evaluation of any technology needs to be context specific. Jane Bailey and Valerie Steeves observe that technology is neither good nor bad.[19] Everything depends on how developers design a technology, how the law regulates it and what values the developers embed into the technology.[20] One may add to this observation that how individuals use the technology and for what purpose matters too. Clearly, it is possible to use AI technology to advance societal objectives. Bruce D Haynes and Sebastian Benthall propose that computer scientists should develop AI systems that detect racial segregation in society.[21] This information can then be used to detect similar treatment of individuals.[22] Since individuals have disparate opportunities as a result of living in segregated areas within the same city,[23] the use of AI technologies to remedy segregation would contribute to the attainment of social justice.

This article focuses on uncovering a number of adverse impacts the use of an AI decision-making system is likely to have both on individuals and society from the perspective of advancing social justice. It is confined to scrutinising the context where educational institutions automate the process of the selection of students by employing AI decision-making processes. Such criteria could include performance on examinations, extra-curricular activities, personal statements, samples of student work and so on. While the article uses examples from a number of countries, the findings can be extended to all universities that use a variety of criteria to judge the merit of individuals. The analysis does not include within its scope AI systems that allocate students to universities based on the students' preferences for a study programme without reference to the merit criteria. An example of the university admissions processes beyond the scope of this paper is that of the French state universities other than grandes écoles.[24] The algorithm allocates places at French state universities to students according to the student's highest preference for a programme and according to whether a student lives within the district where the university is located; a random procedure is used to break the ties.[25] For reasons of space it is beyond the scope of the present enquiry to consider the beneficial uses to which a variety of AI technologies may be put.

Section 1 maintains that it is more meaningful to talk of an AI decision-making process rather than an AI decision-making system. It defines the elements comprising an AI decision-making process for the purpose of situating the discussion. Section 2 introduces Martha Albertson Fineman's vulnerability theory[26] as a theoretical framework for examining some of the ways in which the use of the AI decision-making processes will affect individuals and society from the perspective of social justice. Section 3 investigates some of the types of values that the operation of AI decision-making processes gives rise to. The discussion draws on the vulnerability theory to illustrate some of the ways in which these processes reconfigure social and institutional relationships in which individuals are embedded.[27] It scrutinises how the employment of AI decision-making processes impacts on how society understands lived human experience and human diversity. It is concluded that one of the contexts in which it is desirable to preserve human decision-making processes is where the decision concerns evaluating the capability of the individuals for the purpose of determining their entitlement to resources. Automated decision-making should be avoided where the decision involves representing individuals in geometric space. The university admissions process is an example where the decision-maker evaluates the capabilities of individuals by assessing their skills and personal qualities. The present work is designed to be a starting point for further scholarly exploration for how the use of AI decision-making processes reconfigures societal arrangements and produces society-wide

[17] Convention for the Protection of Human Rights and Fundamental Freedoms art 14; African Charter on Human and Peoples' Rights art 2; American Convention of Human Rights art 1; International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 art 26; International Convention on Economic, Social and Cultural Rights (adopted 16 December 1966, entered into force 3 January 1976) 993 UNTS 3 art 2.2

[18] Ibid; Convention on the Rights of Persons with Disabilities (adopted 13 December 2006, entry into force 3 May 2008) 2515 UNTS 3 art 5(2); *Identoba and Others v Georgia* App No 73235/12 (ECtHR, 12 May 2015), para 96.

[19] Jane Bailey and Valerie Steeves, 'Introduction: Cyber-Utopia? Getting Beyond the Binary Notion of Technology as Good or Bad for Girls' in Jane Bailey and Valerie Steeves (eds), *eGirls, eCitizens: Putting Technology, Theory and Policy into Dialogue with Girls' and Young Women's Voices* (University of Ottawa Press 2015) 5

[20] Ibid

[21] Sebastian Benthall and Bruce D Haynes, *Racial Categories in Machine Learning* (Association for Computing Machinery 2019) 9

[22] Ibid 8

[23] Ibid 7

[24] Lucien Frys and Christian Staat, 'University Admission Practices-France' (*Matching in Practice*, 2016) <http://www.matching-in-practice.eu/university-admission-practices-france> accessed 1 August 2019

[25] Ibid

[26] Martha Albertson Fineman, 'Equality and Difference–the Restrained State' (2015) 66 Alabama Law Review 609, 614

[27] Martha Albertson Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' in Anna Grear and Martha Albertson Fineman (eds), *Vulnerability: Reflections on a New Ethical Foundation for Law and Politics* (Ashgate Publishing Limited 2013) 22

effects. Greater scholarly attention is needed to address the question in what contexts the legislature should require human decision-making.

1. A definition of an artificial intelligence decision-making process

An evaluation of AI-based decision-making processes necessitates understanding what AI is, how it functions and what elements comprise the decision-making process. The decision to frame the discussion in terms of an AI decision-making process as opposed to an AI decision-making system is intentional. One of the reasons for this choice is that AI technology is evolving. For this reason, it is more meaningful to focus on the types of procedures and processes that underlie present AI technologies rather than on how computer scientists design such systems. The evolving nature of AI systems is illustrated by the fact that multiple definitions of artificial intelligence exist and the definitions have been evolving over time.[28] One of the reasons why it is difficult to define the term AI stems from the fact that it is unclear what society means by the term intelligent.[29] According to John McCarthy, 'the problem is that we cannot characterise in general what kind of computational procedures we want to call intelligent.'[30] Given that AI as a discipline is a social phenomenon shaped by individuals, Bao Sheng Loe and colleagues recommend that the definition of AI be continuously updated.[31] A present common understanding of an AI system is that it autonomously learns from being exposed to its environment and makes changes to its model of the external environment based on the sensed changes in the environment.[32]

It is more fruitful to understand the term AI in terms of how a particular system is designed and operates rather than by reference to the term intelligence. Ig Snellen argues that intelligence is a metaphor in the context of technical systems because human beings do the thinking in the course of creating the system's architecture.[33] Similarly, the Dutch Raad van State (the Council of State)[34] maintains that it is misleading to call AI decision-making systems self-learning because they do not understand reality.[35] The processes underlying the construction and operation of AI systems will be examined to show why the term intelligence should be understood as having a specialist meaning in the context of an AI system. The discussion will demonstrate that it is more fruitful to talk of an AI decision-making process rather than an AI decision-making system.

Computer scientists draw on data science techniques when creating AI systems.[36] When one understands the design and operation of AI systems it emerges that these systems are not intelligent in the sense in which societies attribute the term intelligence to human beings. Computer scientists begin the development of an AI system by formulating a problem for which they aim to generate useful knowledge.[37] Computer scientists then prepare data by converting it into a format an AI system can process.[38] The end result the computer scientists strive to achieve will influence how they manipulate and label the data.[39] The next step is to use the data to create a model of the external environment that captures the object of interest,[40] such as a student's predicted examination grades. The model locates patterns in the data by detecting correlations between pieces of data.[41] The model identifies what pieces of information are related to each other.[42] In the unsupervised learning process computer scientists let the system search for patterns; the system allocates individuals into groups based on shared characteristics.[43] In the supervised learning process the computer scientists formulate a criterion and the AI system sorts individuals into groups based on their likelihood of fulfilling that criterion.[44] The model the system generates allows the user to predict that an individual belongs to a particular group of people with shared characteristics.[45] The AI system predicts an individual's performance based on the performance of individuals whom it treats as having similar characteristics to the individual in question.[46] It applies a

[28] Bao Sheng Loe and others, *The Facets of Artificial Intelligence: A Framework to Track the Evolution of AI* (International Joint Conferences on Artificial Intelligence Organization, Stockholm, 2018) 5180

[29] Max Vetzo, Janneke Gerards and Remco Nehmelman, *Algoritmes en Grondrechten* (Boom Juridisch 2018) 41

[30] John McCarthy, 'What is Artificial Iintelligence? Basic Questions' (*Stanford University,* 2007) <http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html> accessed 13 May 2019

[31] Loe and others, *The Facets of Artificial Intelligence: A Framework to Track the Evolution of AI* 5186

[32] Vetzo, Gerards and Nehmelman, *Algoritmes en Grondrechten* 41

[33] Ignatius Theodorus Maria Snellen, 'Het Automatiseren van Beschikkingen Bestuurskundig Beschouwd' in Hans Franken and others (eds), *Beschikken en Automatiseren, Preadviezen Voor de Vereniging voor Administratief Recht* (Samsom HD Tjeenk Willink 1993) 55, quoted in Beppie Margreet Alize van Eck, 'Geautomatiseerde Ketenbesluiten & Rechtsbescherming: Een Onderzoek Naar de Praktijk van Geautomatiseerde Ketenbesluiten Over een Financieel Belang in Relatie Tot Rechtsbescherming' (PhD thesis, Tilburg University 2018) 193

[34] The Council of State Advises the Government and Parliament on Legislation and Governance. Raad van State, 'The Council of State' (*Raad van State* 2019) <https://www.raadvanstate.nl/talen/artikel> accessed 25 July 2019

[35] Raad van State, *Advies W04.18.0230/I: Ongevraagd Advies Over de Effecten van de Digitalisering Voor de Rechtsstatelijke Verhoudingen* (Raad van State 2018) 9

[36] Rosaria Silipo, 'What's in a Name? Artificial Intelligence or Data Science?' (*BetaNews Inc,* 2019) <https://betanews.com/2019/02/05/artificial-intelligence-or-data-science> accessed 14 May 2019

[37] Foster Provost and Tom Fawcett, *Data Science for Business* (O'Reilly Media Inc 2013) 19

[38] Ibid 30

[39] Ibid

[40] Ibid 39

[41] Ibid 25

[42] Ibid 23

[43] Ibid 24

[44] Ibid

[45] Ibid 107

[46] Ibid 146

decision-making procedure for determining whether an individual is entitled to a positive decision.[47]

Presently, AI systems lack human intelligence. They do not have the capacity to understand what the correlation between the pieces of data means, whether the correlation has significance and how this correlation corresponds to phenomena in the world. When an AI system finds a correlation between two pieces of data, this does not signify that A causes B.[48] In fact, the correlations the AI system detects can be spurious or accidental.[49] For instance, there is a high correlation but not causation between ice-cream consumption and shark attacks; both tend to occur during warmer seasons.[50] Another reason why AI systems are not intelligent stems from the fact that they cannot independently reflect on what their predictions signify and whether the decision-making procedure produces societally desirable outcomes.

David Preiss and Robert Sternberg view human beings and technology as having a reciprocal influence.[51] Technology transforms human cognitive skills[52] as well as the understanding of what is human intelligence.[53] Meanwhile, cultural context influences technologies.[54] This observation is corroborated by the fact that as AI gains new capabilities to solve problems, society redefines the term human intelligence in order to differentiate between AI and human beings.[55] The better approach is to acknowledge that any definition of human and machine intelligence is tentative. There needs to be an awareness of what definition of intelligence one chooses, why and with what consequences. Given that some computer scientists seek to replicate human intelligence in AI,[56] there may come a time when the dividing line between 'artificial' and 'human' intelligence is less clear. Society should reflect on the social role the term intelligence has as it continues to refine the meaning of this term.

Currently, different definitions of algorithmic or automated decision systems exist. Definitions framing the subject matter broadly and by reference to an artificial intelligence decision-making process are preferable. The Australian Human Rights Commission defines 'AI-informed decision-making' as 'decision-making which relies wholly or in part on artificial intelligence.'[57] By hingeing the definition on the term artificial intelligence and the notion of decision-making, the Australian Human Rights Commission conceives of AI-based decision-making in terms of the computer science techniques underpinning the decision-making process. Provided one gives the term artificial intelligence a holistic interpretation, the definition of the AI-informed decision-making can be interpreted as covering all the stages involved in the decision-making processes. What is significant is that the Australian Human Rights Commission uses the term 'AI-informed decision-making' rather than the term decision-making system.

In contrast, the Directive on Automated Decision-Making of the Government of Canada contains a narrower definition becauses it employs the term an automated decision-making system. It defines an automated decision system as

> [A]ny technology that either assists or replaces the judgement of human decision-makers. These systems draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets.[58]

What is common to the definitions of the Australian Human Rights Commission and the Directive on Automated Decision-Making is that they discuss the respective roles of artificial intelligence technology and human beings in the decision-making process. What is more, the definitions centre on the types of computer science techniques involved. One of the reasons why the term automated decision-making system is narrower than the term decision-making process is because it excludes stages that bear on the outcome of the decision-making process but that take place prior to the actual construction of the system. In particular, the decision-making process begins when the computer scientist formulates a problem to be solved using the AI-driven procedure because this stage bears on the outcome of the decision-making process. Solon Barocas and Andrew D Selbst posit that computer scientists exercise subjectivity when they formulate a problem the machine should solve.[59] They do this by defining the variable for which the machine makes a prediction.[60] How computer scientists define this variable, such as a good employee, shapes what relationships between data the machine finds and therefore its predictions about the suitability of the applicant for the position.[61] According to Reuben Binns, if the computer scientist uses a biased variable as a benchmark for the basis on which the AI decision-making process predicts

[47] Sorelle A Friedler, Carlos Scheidegger and Suresh Venkatasubramanian, 'On the (Im)possibility of Fairness' (2016) 1609.07236v1 arXiv 1, 10

[48] Eubanks, *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* 144

[49] Ibid 144-45

[50] Ibid 145

[51] David Preiss and Robert Sternberg, 'Technologies for Working Intelligence' in David Preiss and Robert Sternberg (eds), *Intelligence and Technology: the Impact of Tools on the Nature and Development of Human Abilities* (Routledge 2005) 199

[52] Ibid

[53] Ibid 184-85

[54] Ibid 199

[55] Chris Smith, 'Introduction', *The History of Artificial Intelligence* (University of Washington 2006) 4

[56] Ben Goertzel and Pei Wang, 'Introduction: Aspects of Artificial General Intelligence' in Ben Goertzel and Pei Wang (eds), *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms Proceedings of the AGI Workshop 2006* (IOS Press 2007) 1

[57] Australian Human Rights Commission, 'Decision Making and Artificial Intelligence' (*Australian Human Rights Commission*, 2019) < https://tech.humanrights.gov.au/decision-making-and-artificial-intelligence> accessed 22 January 2020

[58] Government of Canada, 'Directive on Automated Decision-Making' (Government of Canada 1 April 2019) <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> accessed 23 January 2019

[59] Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671

[60] Ibid

[61] Ibid 679-80

future performance then the outcome will be biased.[62] The term decision-making process is preferable because it can be defined to include the stage where the computer scientist formalises the problem to be solved using an AI-driven procedure.

The Council of Europe Committee of Experts uses the term an AI decision-making system and in this respect mirrors the approach of the Directive on Automated Decision-Making. The Council of Europe Committee of Experts lists the types of tasks the AI decision-making system carries out and the processes that comprise the operation of such systems.[63] In particular, it defines algorithmic systems as applications that 'perform one or more tasks such as gathering, combining, cleaning, sorting and classifying data, as well as selection, prioritisation, recommendation and decision-making.'[64] The definition of the Council of Europe Committee of Experts is preferable to the definition the Canadian Directive on Automated Decision-Making offers. It focuses on the steps involved in constructing a model that the AI decision-making system uses for making predictions about future performance and the process of producing a decision in respect of an individual. Thus, there is an emphasis on the process leading to the decision rather than on the type of techniques the computer scientists use in order to program AI systems. This aspect makes the nature of AI decision-making systems explicit by listing the steps entailed in producing a decision. This definition makes it easier for the law makers and end users to debate the social consequences of using AI decision-making systems.

Another advantage of defining the term in terms of what elements comprise the AI decision-making process is that it provides an understanding of what the system does and how it achieves its objective. On the other hand, the term AI decision-making system is opaque. Little understanding may be gleaned from this term. Society uses the term system to refer to interdependent and interacting elements.[65] The fact that AI technology utilises a combination of different elements reveals little about the nature of the tasks it performs and how it performs them. This stems from the fact that the term system draws attention to the physical architecture of the system and what components or elements comprise the system. What is core for understanding decision-making is the process through which one arrives at a decision rather than the fact that various interdependent stages are involved in the decision-making process.

A relevant consideration is that since societal understanding of AI is evolving[66] it could be that in fifty years time the definition of what AI is and how it operates will be very different. For instance, the architecture of AI could have a highly distributed form where it is unclear exactly what its elements are and how its different elements interplay. Given that computer scientists use the knowledge about the human brain as inspiration to create new AI techniques,[67] AI could resemble the functioning of a human body very closely. Such developments could make it difficult to speak about the machine as a system.

Of significance is that there is a parallel between the elements comprising human and the AI decision-making processes. The term AI decision-making system fails to capture this important element. This is because society does not conceive of human beings and their deliberation as a system. However, one can talk about the similarities in the decision-making process human beings engage in and the AI systems carry out because human beings develop the AI decision-making process. Given that human beings exercise their judgement in developing AI decision-making processes, it is not surprising that there can be a degree of similarity between human and AI decision-making processes.

The human decision-making process begins with the framing of the goal that the decision-making procedure is designed to achieve and with the identification of the criteria corresponding to entitlement to a positive decision. For instance, when a university sets up an admissions process, it formulates a set of goals. The goals could be to attract students who possess a particular skillset, who have particular estimated academic performance or who are representative of the population as a whole. The university could aim to mitigate the existence of societally embedded inequalities by placing emphasis on attracting candidates who experience social barriers. The goal the university sets will determine what criteria it chooses as a basis on which the officials should select the students. The admissions criteria will determine whether the human decision-maker considers only the student's grades or additional criteria, such as extracurricular activities, work experience and the applicant's personal circumstances. The decision-making criteria determine what qualities the decision-maker takes into account or ignores.

When computer scientists decide how to formulate the problem and what the AI process predicts, they select the goal for the decision-making process and the criteria that form the basis of the decision-making process.[68] When formulating the problem to be solved the computer scientsits are in a similar position to human decision-makers who are tasked with developing and applying a decision-making procedure. In both cases the goal to be achieved determines what criteria the decision-makers adopt for selecting a pool of candidates. The difference is that human decision-makers can choose criteria that can be expressed in both quantitative and qualitative terms. On the other hand, computer scientists can select only those criteria as indicators of a good candidate that can be expressed in quantitative terms.[69] Examples of quantita-

[62] Reuben Binns, 'Imagining Data, Between Laplace's Demon and the Rule of Succession' in Irina Baraliuc and others (eds), *Being Profiled: Cogitas Ergo Sum* (Amsterdam University Press 2018)

[63] Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence, *Draft Recommendation of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems* (Council of Europe 2019) par 3

[64] Ibid

[65] Merriam-Webster Incorporated, 'System' (*Merriam-Webster Incorporated*, 2019) < https://www.merriam-webster.com/dictionary/system> accessed 22 January 2020

[66] Loe and others, *The Facets of Artificial Intelligence: A Framework to Track the Evolution of AI* 5180

[67] Zhongzhi Shi, *Advanced Artificial Intelligence* (World Scientific Publishing 2011) 10 par 1

[68] Barocas and Selbst, 'Big Data's Disparate Impact' 678-80

[69] Friedler, Scheidegger and Venkatasubramanian, 'On the (Im)possibility of Fairness' 3

tive proxies are grades, rankings and number of outputs. Computer scientists need to find quantitative proxies if they want to capture qualitative characteristics.[70] Qualitative characteristics refer to multidimensional and multitextured qualities, such as creativity and leadership skills.

When decision-makers use a grade to capture a qualitative characteristic, such as intelligence, they employ a proxy.[71] For instance, the standardised admission test for graduate programmes General Record Examination measures analytical, quantitative and verbal skills.[72] The scores of this test did not have accurate predictive capacity for how lecturers at Yale University rated the students' analytical, creative and practical skills on a graduate psychology programme.[73] Thus, AI decision-making processes should be viewed as mapping proxies onto the model alongside actual characteristics. It is difficult to express qualitative characteristics, such as interpersonal and teamwork skills, numerically because they relate to how individuals interact with each other. While classmates could be asked to rank each other on the metric of a teamworking skill, such responses would be unreliable. Individuals can be influenced by their personal attitudes, by a desire to compete or by unconscious biases. They may lack the distance necessary to reflect on how all team members interacted. Qualitative descriptions of how an individual acted in particular circumstances provide more information about an individual's teamwork skills. The AI decision-making process should be understood as a more limited procedure than a human decision-making process by virtue of its limited capacity to capture qualitative data and the context behind this data.

There another important difference between the human and the AI decision-making processes. Human decision-makers determine what facets of the person they consider through choosing the decision-making criteria. In contrast, AI decision-making processes create what Luke Stark calls the 'scalable subject.'[74] The system purports to model the individual but in fact reflects correlations between characteristics present within a group that may not apply to the individual in question.[75] Annamaria Carusi elaborates that the model represents the individual in a reductive way[76] and that the approach to representation contains particular values.[77] The lack of granular information can result in the model making generalisations that are unfair to the individual.[78] To illustrate, since the model groups students based on past examinations performance for the purpose of making a prediction about future results, it would group students who performed poorly irrespective of the reason for the low results. This may result in the AI system falsely predicting a low grade for a student whose performance in the past had been influenced by an illness but who recovered later.

Patrick Allo comments that the model depicting the patterns in the data represents a proxy for what one is trying to predict rather than the actual state of the world.[79] A model that predicted the student's performance on an examination that tested how well the student memorised the material is not a reliable proxy for the student's aptitude. While the model provides indirect information about an individual's memory capacity, it tells little about the student's aptitudes, such as problem-solving capacity and creativity. The differences between human and AI decision-making processes do not preclude considering the automation of decisions in terms of an AI decision-making process. In fact, the term process highlights the fact that human beings construct the decision-making procedure within the machine. What is more, this term makes it possible to demarcate at what point the decision-making commences and ends. Crucially, the definition that focuses on the process rather than the system prevents misrepresentation. The term artificial intelligence system can create a misleading impression that the system has capabilities that correspond to human intelligence or that the decision of computer scientists relating to the variable to be predicted had no impact on the outcome for the applicants.

The question is what elements comprise the AI decision-making process. The proposed definition does not cover situations where a human decision-maker uses the analytics the AI system generates as a sole or partial basis for reaching a decision. By combining definitions of the Council of Europe Committee of Experts and the Australian Human Rights Commission one can arrive at a suitable definition of an AI decision-making process. What needs to be included in addition is the element of human decision-making involved in formulating a problem to be solved and how to construct the system to achieve this goal. The AI decision-making process should be understood as starting with the computer scientist formulating the problem to be solved and the goals to be achieved. It encompasses the collection, cleaning, labelling, aggregation, analysis, manipulation and processing of data. These steps are carried out in order to predict future performance and to produce a decision affecting the right or entitlement of an individual to a positive decision. The definition includes the application of a template for determining whether an individual should be granted a positive decision. This definition is appropriate even though the process of creating a model of the environment as a basis for making a prediction is a separate stage from the decision-making procedure for determining an individual's entitlement to an affirmative de-

[70] Provost and Fawcett, *Data Science for Business* 339

[71] Friedler, Scheidegger and Venkatasubramanian, 'On the (Im)possibility of Fairness' 3

[72] Educational Testing Service, 'About the GRE General Test' (*Educational Testing Service*, 2019) <https://www.ets.org/gre/revised_general/about> accessed 22 July 2019

[73] Robert J Sternberg and Wendy M Williams, *Does the Graduate Record Examination Predict Meaningful Success in Psychology* (Yale University 1994), quoted in Robert J Sternberg, 'Myths, Countermyths, and Truths about Intelligence' (1996) 25 Educational Researcher 11, 14

[74] Luke Stark, 'Algorithmic Psychometrics and the Scalable Subject' (2018) 48 Social Studies of Science 204, 207

[75] Ibid 213

[76] Annamaria Carusi, 'Beyond Anonymity: Data as Representation in E-research Ethics' (2008) 37 International Journal of Internet Research Ethics 37, 61

[77] Ibid 42

[78] Frederick Schauer, *Profiles, Probabilities and Stereotypes* (Belknap Press 2006) 3

[79] Patrick Allo, 'Mathematical Values and the Epistemology of Data Practices' in Irina Baraliuc and others (eds), *Being profiled: Cogitas Ergo Sum* (Amsterdam University Press 2018) 27

cision.[80] A broad definition encompassing all the elements that bear on how the algorithmic process measures and predicts future performance as well as how it produces a decision is needed. This approach ensures adequate protection of individuals. This approach leaves scope for the fact that a computer scientist could embed a variety of decision-making procedures for arriving at a decision. For instance, an AI decision-making process could allocate the resources to individuals with the highest score for predicted performance.[81] It could take into account other considerations. To illustrate, Aditya Krishna Menon and Robert C Williamson designed an algorithmic decision-making procedure for an AI system that they argue achieves the best trade-off between accuracy and fairness.[82]

Why is an expansive definition for the term AI decision-making process necessary? The process of mapping the data onto the model and of predicting an individual's performance based on the model has influence on whether the individual will receive a positive decision. The proposed definition is designed to capture the fact that computer scientists make subjective decisions in the course of creating the architecture that enables the AI decision-making process to collect, aggregate and analyse data.[83] The choices computer scientists make affect how the AI decision-making process produces decisions and what kind of decision an individual receives.[84] Often, the decisions of computer scientists are hidden and reflect a particular understanding of the world.[85] For example, computer scientists make assumptions when deciding how to represent a person in a model.[86] Since individuals are multidimensional and cannot be described exhaustively, it is in theory possible to create an infinite number of snapshots of the individual depending on what combination of characteristics one inputs. For instance, a candidate can be described as a female with a score of eighty percent for mathematics and a score of fifty percent for English language. Alternatively, the same candidate can be designated as a female candidate who is enroled in a school located in an underfunded district. She learns in an overcrowded classroom due to a shortage of English teachers. Depending on what characteristics one chooses as being relevant for the purpose of generating a model, one can get a different snapshot of the person. What is more, since inequalities are structurally embedded in society, groups will be represented in a distorted manner when mapped onto the model.[87] The researchers cite the fact that the verbal section of the standardised American university admission test SAT functions differently for the African-American individuals.[88] It follows that there is a discrepancy between the real world and how the AI decision-making process maps the world onto a geometrical space as part of generating a model of the world.[89]

Computer scientists can steer the data analysis process by framing for what metric the AI decision-making process formulates the predictions and by choosing a particular approach to data analysis.[90] In the context of AI and human decision-making processes the choice of characteristics to denote merit for the purpose of selecting students shapes whether individuals have an equal opportunity to be admitted to university. Some selection criteria appear neutral but in fact hide the fact that the decision-making procedure creates admission barriers for children from poor socioeconomic backgrounds. For instance, the computer scientist can set a good candidate performance for admissions to a university in terms of excelling at playing a musical instrument, painting, playing professional sports or winning a dance contest. This approach to student admissions resembles how the highest ranked universities in the United States select students.[91] Children have unequal access to participation in extracurricular activities. Anna Bull examines how complex factors lead to children from middle-class and upper-class families being more likely to play a musical instrument.[92] The reasons include the cost of music lessons and the fact that the approach to teaching music reflects the nature of interactions prevalent in middle-class teaching settings.[93] This example shows that the criteria computer scientists embed into the AI decision-making process and the metrics by which the program generates the prediction will shape whether individuals have equal access to university education. Accordingly, the AI decision-making process should be defined to incorporate all stages of system development and operation beginning with formulation of the problem to be solved and ending with a decision output.

2. Introducing the theoretical framework: the vulnerability theory

The values one holds amongst others will determine how one evaluates the AI decision-making process. For instance, those who value efficiency will ask questions such as whether the use of the AI decision-making process cuts costs or shortens

[80] Provost and Fawcett, *Data Science for Business* 25

[81] Friedler, Scheidegger and Venkatasubramanian, 'On the (Im)possibility of Fairness' 3

[82] Menon and Williamson, 'The Cost of Fairness in Binary Classification' 2

[83] Friedler, Scheidegger and Venkatasubramanian, 'On the (Im)possibility of Fairness' 3

[84] Ibid

[85] Ibid

[86] Ibid 6-7

[87] Ibid 7

[88] Ibid 8; Maria Veronica Santelices and Mark Wilson, 'Unfair Treatment? The Case of Freedle, the SAT, and the Standardisation Approach to Differential Item Functioning' (2010) 80 Harvard Educational Review 106, 126

[89] Friedler, Scheidegger and Venkatasubramanian, 'On the (Im)possibility of Fairness' 7

[90] Felix Stalder, 'From Inter-subjectivity to Multi-subjectivity: Knowledge Claims and the Digital Condition' in Irina Baraliuc and others (eds), *Being profiled: Cogitas Ergo Sum* (Amsterdam University Press 2018) 136

[91] Ilana Kowarski, 'How Colleges Weigh Applicants' Extracurricular Activities' (*US News*, 2018) <https://www.usnews.com/education/best-colleges/articles/2018-10-25/how-colleges-weigh-applicants-extracurricular-activities> accessed 14 May 2019

[92] Anna Bull, 'Reproducing Class? Classical Music Education and Inequality' (*Discover Society*, 2014) <https://discoversociety.org/2014/11/04/reproducing-class-classical-music-education-and-inequality/> accessed 14 May 2019

[93] Ibid

the deliberation time.[94] Economists define efficiency as 'a situation where each good is produced at the minimum cost and where individual people and firms get the maximum benefit from the use of the resources.'[95] Equity is a different type of value in comparison to efficiency.[96] Equity concentrates on whether there is fairness and justice.[97] Individuals who value equity will ask different types of questions than economists when assessing the desirability of using AI decision-making processes. How one evaluates AI decision-making processes will differ depending on how one defines fairness. Different people have a different understanding of what constitutes fairness.[98] Additionally, there is a difference between how scholars[99] and the general population define fairness.[100] In discussing fairness it is important to acknowledge the value of pluralism and cultural diversity. Respect for individuals is contingent on a recognition of their opinions and value systems. The representation of a plurality of views is conducive to informed discussions about what constitutes a good life. It widens the array of arguments and introduces new vistas from which to assess propositions.

Roger Brownsword argues that to gain legitimacy regulators should adopt instruments that capture common values and concerns while leaving room for local difference.[101] The present article uses the vulnerability theory as a lens for evaluating AI decision-making processes because the theory captures how citizens conceive of core components of fairness and justice. Martha Albertson Fineman formulated 'vulnerability theory' as an 'an alternative to theories of social justice and responsibility that focus on achieving formal equality.'[102] The term 'social justice' focuses on the position of many individuals within a society.[103] Traditionally, advocates of social justice called for a just distribution of resources and of the fruit of economic production amongst the individuals.[104] What differentiates Fineman's approach to understanding how to advance social justice is that she examines the impact of legally constructed social institutions and relationships on the lives of individuals.[105] The vulnerability theory reflects how citizens understand fairness by focusing on the way in which the state constructs relationships between individuals and institutions.[106] A study found that individuals use the terms fairness and justice interchangeably to refer to violations of equity and equality.[107] Individuals understand fairness to include both how individuals are positioned in relation to other individuals in relationships as well as how individuals are positioned in relation to institutions in society.[108] The employment of the vulnerability theory allows one to assess what impact the use of AI decision-making processes has on individuals and society. The present article evaluates a number of ways in which the use of the AI decision-making processes impacts on the individuals and society from the perspective of social justice. It is beyond the scope of this work to evaluate the AI decision-making process from the vantage point of all theories of justice and fairness across cultures. Neither is it possible to examine in a comprehensive manner all the ways in which the cumulative use of the AI decision-making process in different domains will transform society.

The use of the vulnerability theory approach avoids drawing an arbitrary distinction between the private and the public domains.[109] What becomes relevant for the analysis is how the employment of the AI decision-making processes affects the subject of the decision-making procedure and society rather than whether the inequity arose from the relationship with the state or with other individuals. In contrast, scholarly writings in political science and philosophy distinguish between public and private domains to demarcate when the state can intervene to regulate.[110] This distinction is arguably apparent from how some scholars contrast the terms justice and fairness.[111] John Rawls for example defines justice as relating to the institutional arrangements and practices that define rights, duties and offices.[112] He defines fairness as relating to the rights of persons arising in the course of their interaction with one another on an equal basis.[113] Individuals would agree on rules to ensure that they did not feel they were being taken advantage of in interpersonal relationships.[114] The drawing of a distinction between the relationships individuals have with each other and with the institutions for the purpose of assessing the impact of AI decision-making processes is undesirable. Fairness and justice are open-ended terms that society uses as heuristic devices to redress inequities. The content of the terms justice and fairness can be given different meanings[115] depending on the context to which individuals

[94] Vishal Marria, *The Future of Artificial Intelligence In The Workplace* (Forbes Media LLC 2019)
[95] John Sloman, *Economics* (6 edn, Prentice Hall 2006) 9
[96] Ibid 11
[97] Ibid
[98] Ibid
[99] Norman J Finkel, Rom Harré and José-Luis Rodriguez Lopez, 'Commonsense Morality Across Cultures: Notions of Fairness, Justice, Honor and Equity' (2001) 3 Discourse Studies 5, 5
[100] Ibid 21
[101] Roger Brownsword, 'Regulatory Cosmopolitanism: Clubs, Commons, and Questions of Coherence' (2010) 018/2010 TILT Law & Technology Working Paper 2, 4
[102] Nina A Kohn, 'Vulnerability Theory and the Role of Government' (2014) 26 Yale Journal of Law & Feminism 2, 6
[103] Martha Albertson Fineman, *Vulnerability and Social Justice* (Emory University 2019) 1
[104] United Nations, *Social Justice in an Open World: The Role of the United Nations* (United Nations 2006) 7
[105] Fineman, *Vulnerability and Social Justice* 2
[106] Ibid
[107] Finkel, Harré and Lopez, 'Commonsense Morality Across Cultures: Notions of Fairness, Justice, Honor and Equity' 21
[108] Ibid
[109] Martha Albertson Fineman, 'Injury in the Unresponsive State: Writing the Vulnerable Subject Into Neo-Liberal Legal Culture' in Anne Bloom, David M Engel and Michael McCann (eds), *Injury and Injustice: The Cultural Politics of Harm and Redress* (Cambridge University Press 2018) 19
[110] See for instance the work of Max Weber, Isaiah Berlin, Jürgen Habermas, Richard Rorty, Michael Walzer and John Stuart Mill. Raymond Geuss, *Public Goods, Private Goods* (Princeton University Press 2003) 10
[111] Finkel, Harré and Lopez, 'Commonsense Morality Across Cultures: Notions of Fairness, Justice, Honor and Equity' 5
[112] John Rawls, 'Justice as Fairness' (1958) 67 The Philosophical Review 164, 164
[113] Ibid 178
[114] Ibid
[115] Manuel Velasquez and Claire Andre, 'Justice and Fairness' (1990) 3 Issues in Ethics 1

apply these terms[116] and depending on the society's value system.[117] If one is to address inequities comprehensively, one should focus on all sources from which the inequities may arise.

The vulnerability theory demonstrates why it is important to focus both on how the operation of the AI decision-making process constructs relationships between individuals, and between individuals and institutions in analysing the impact of these systems from the perspective of social justice. According to the vulnerability theory, individuals are situated in different economic, social, cultural and institutional relationships.[118] These relationships cannot be clearly demarcated as being either private or public.[119] The position of the individual within these relationships determines whether the institutional arrangements create opportunities or impediments.[120] These institutions form a system that determines the resilience of the individual.[121] The term resilience refers to the individual's ability to recover from life's setbacks and to take advantage of opportunities.[122] There are five types of resources that the institutions provide that are crucial for human flourishing.[123] First, material goods determine the individuals' quality of life and allow them to accumulate additional resources.[124] Second, individuals derive support from social networks.[125] Third, human assets, such as education and employment, place the individuals in a position to develop their capabilities.[126] Fourth, individuals benefit from having access to existential and aesthetic resources, such as religion, culture and art.[127] Fifth, individuals need ecological assets, such as the natural environment, to maintain physical well-being.[128] Access to interpersonal resources of support, such as family[129] and social networks, constitute relationships that one typically views as private. In practice, such private relationships cannot be separated from relationships with the state.[130] For instance, laws prohibiting harassment, bullying and discrimination play a crucial role in creating inclusive spaces where individuals can engage in interpersonal relationships. Consequently, it is artificial to distinguish between justice and fairness based on whether the relationship is public or private in nature.

For the purpose of this article, it appears desirable to use the vulnerability theory rather than theories of fairness, which focus on the treatment of an individual for the purpose of evaluating the AI decision-making processes. The school allocation system in New York City is a case study that illustrates how a focus on the impact of the employment of the AI decision-making process on the individual can result in failing to detect both sources of social injustice and unfairness for the individual. Currently, the authorities in the New York City use an algorithm to place children into high schools.[131] Children provide a list of twelve school choices to the authorities.[132] The algorithm allocates children to a school by selecting a pool of candidates with the highest grade.[133] This means that in practice each school will have students within a particular grade range. Schools that are in the highest demand will have a pool of students with top grades. Schools with a lesser demand will have a pool of students who have grades in the mid or low range. This approach to using an algorithmic process to allocate children to schools results in segregation. Children with high grades study in different buildings and are geographically separated from the children with low grades. This finding should be viewed against the backdrop that schools worldwide have racial and socioeconomic segregation.[134] James A Allen expresses a broader concern that the use of AI decision-making processes perpetuates and reinforces existing segregation.[135]

The focus on whether the selection procedure the AI decision-making process utilised is fair for a particular student in terms of merit occludes the wider social justice concerns. When a thirteen-year-old Jimmy (not his real name) voiced his opposition to being rejected by five of his top preference schools, he was told that his grade of eighty five did not qualify him for admission.[136] The cut-off point for admission to those schools was a grade of ninety.[137] A focus on whether Jimmy performed better in comparison to another student precludes a more in-depth enquiry. The grade has an appearance of being an objective marker that measures the students'

[116] Michael Adler, 'Fairness in Context' (2006) 33 Journal of Law and Society 615, 638
[117] Kenneth A Rasinski and Leslie A Scott, 'Culture, Values, and Beliefs About Economic Justice' (1990) 4 Social Justice Research 307, 320
[118] Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' 22
[119] Fineman, 'Injury in the Unresponsive State: Writing the Vulnerable Subject into Neo-Liberal Legal Culture' 19
[120] Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' 23
[121] Ibid 22
[122] Fineman, 'Equality and Difference–the Restrained State' 622-23
[123] Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' 22-23
[124] Ibid 22
[125] Ibid 23
[126] Ibid
[127] Ibid
[128] Ibid
[129] Margaret Thornton, 'The Cartography of Public and Private' in Margaret Thornton (ed), *Public and Private: Feminist Legal Debates* (Oxford University Press 1995) 2
[130] Fineman, 'Injury in the Unresponsive State: Writing the Vulnerable Subject into Neo-Liberal Legal Culture' 19
[131] Alvin Roth, 'Why New York City's High School Admissions Process Only Works Most of the Time' (*Chalkbeat*, 2015) <https://www.chalkbeat.org/posts/ny/2015/07/02/why-new-york-citys-high-school-admissions-process-only-works-most-of-the-time> accessed 15 May 2019
[132] Ibid
[133] Ibid
[134] Thomas Toch, 'The Lottery That's Revolutionizing D.C. Schools' (*The Washington Post*, 2019) <https://www.washingtonpost.com/news/magazine/wp/2019/03/20/feature/the-lottery-thats-revolutionizing-DDD-c-schools> accessed 15 May 2019
[135] James A Allen, 'The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining' (2019) 46(2) Fordham Urban Law Journal 219, 234
[136] Alvin Roth, 'Why New York City's High School Admissions Process Only Works Most of the Time' (*Chalkbeat*, 2015) <https://www.chalkbeat.org/posts/ny/2015/07/02/why-new-york-citys-high-school-admissions-process-only-works-most-of-the-time> accessed 15 May 2019
[137] Ibid

intelligence and the effort they put into studying. There is an impression that the state treated Jimmy fairly in the sense of just deserts. The just desert approach to justice stipulates that the entitlement of individuals to resources depends on their performance and expended effort.[138] This approach ignores how the relationships of individuals to institutions result in individuals having an unequal ability to navigate the social institutions.

The focus on whether an individual received her just desert fails to account for the fact that citizens regard the role of relationships between the individual and the state as being relevant to the assessments of justice. The importance of the state creating a just relationship between the individual and the social institution has a particular significance in the context of the provision of education. Access to education is intimately linked to the ability of individuals to develop resilience, to accumulate resources and to take advantage of opportunities in the future.[139] When a state does not provide adequate education for individuals, their ability to secure employment and to maintain adult family relationships suffers.[140] An individual may find it impossible to compensate for or to recover from such deprivations.[141] Studies show that there is a relationship between inequality and the geographical location where the individual lives.[142] For instance, in the United States a school in one district can receive three times less funding than a school located in a more affluent district within the same city.[143] This stems from the fact that property taxes provide half of the funding budget for the schools and from the fact that properties have different values.[144] An evaluation of whether Jimmy received just treatment from the vantage point of social justice would require that one ask whether Jimmy's school had the same quality of teaching as that of children who received admission offers from Jimmy's top school choices. On the application of the vulnerability theory, Jimmy did not receive just treatment if the admitted students attended better-funded schools.

Another problem with the individual centred approaches to evaluating justice is that there is no enquiry into whether some students received a higher grade because they benefited from their parents being able to afford to pay for private tutors. For instance, in the United Kingdom a quarter of children benefit from having a private tutor.[145] Lee Elliot Major is a member the Sutton Trust dedicated to advancing social mobility.[146] Lee comments: 'You are four times more likely to get a private tutor if you are in the top fifth of the income range, so we are worried about the gap outside the school gates.'[147] Since students who benefit from having tutors have an advantage over students who do not, one cannot call a school allocation decision-making process fair if it does not account for the differential access of students to private tutors. A vulnerability theory analysis opens the window to enquire into how unequal access to human assets, such as tutoring, hinders the ability of students from lower income families to gain admission to top performing schools.

The vulnerability theory allows one to consider a greater array of concerns than some theories of justice, such as the just desert theory. Unlike vulnerability theory, the just desert approach to justice does not enquire into whether it is just to segregate individuals because it focuses on the actions of the individual. From the perspective of vulnerability theory, the process of segregation creates unjust relationships. The process of segregation denies children access to well-resourced schools and equal access to learning opportunities.[148] It precludes students from building a network with peers from different backgrounds and from developing their full capability. Students have a diminished capacity to acquire cultural capital[149] in the form of knowledge that individuals in the possession of power designate as high culture.[150] If a society values the possession of cultural capital, then the lack of socialisation between children from different socioeconomic backgrounds adversely affects the ability of these children to succeed.[151] According to Fineman, asset-conferring institutions operate concurrently, interactively and 'can be sequential.'[152] Institutional arrangements can have inter-generational effects, particularly where society structured the group's identity in a manner that differentiates it from other groups.[153] Segregation produces unjust social relationships by symbolically demarcating some groups as inferior and[154] by enacting hierarchies.[155] On Fineman's analysis, segregation leads to individ-

[138] David Schmidtz, *The Elements of Justice* (Cambridge University Press 2006) 31

[139] Fineman, 'Equality and Difference–the Restrained State' 623

[140] Ibid 624

[141] Ibid

[142] Alwyn Young, 'Inequality, the Urban-Rural Gap and Migration' (2013) 128 The Quarterly Journal of Economics 1727, 1750

[143] Cory Turner and others, 'Why America's Schools Have A Money Problem' (*United States National Public Radio*, 2016) <https://text.npr.org/s.php?sId=474256366> accessed 14 May 2019

[144] Ibid

[145] Tracy McVeigh, 'Are Private Tutors for Children Just the Latest Educational "Arms Race"?' *The Guardian* (London, 4 October 2015) <https://www.theguardian.com/education/2015/oct/04/private-tutors-arms-race-schools-parents> accessed 15 May 2019

[146] The Sutton Trust, 'About Us: Our Cause' (*The Sutton Trust*, 2019) <https://www.suttontrust.com/about-us/> accessed 14 May 2019

[147] McVeigh, 'Are Private Tutors For Children Just the Latest Educational "Arms Race"?'

[148] Sonia Lehman-Frisch, 'Segregation, Spatial (In)Justice, and the City' (2011) 24 Berkeley Planning Journal 70, 78

[149] Pierre Bourdieu, 'Les Trois États du Capital Culturel' (1979) 30 Actes de la Recherche en Sciences Sociales 3, 3

[150] Pierre Bourdieu, *Distinction: A Social Critique of the Judgment of Taste* (Harvard University Press 1984) 245

[151] Annette Lareau and Elliot B Weininger, 'Cultural Capital in Educational Research: a Critical Assessment' (2003) 32 Theory and Society 567, 598; Mads Meier Jæger, *Does Cultural Capital Really Affect Academic Achievement?* (Working Paper Series CSER WP No 0001, 2010) 26

[152] Martha Albertson Fineman, 'Beyond Identities: The Limits of an Antidiscrimination Approach to Equality' (2012) 92 Boston University Law Review 1713, 1757

[153] Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' 21

[154] Lehman-Frisch, 'Segregation, Spatial (In)Justice, and the City' 79

[155] William T Bielby and James N Baron, 'Men and Women at Work: Sex Segregation and Statistical Discrimination' (1986) 91 American Journal of Sociology 759, 761

uals finding themselves in a vicious cycle where individuals remain in poverty.

Social justice is chosen as a lens through which to conduct the analysis of the impact of using AI decision-making processes because it is more expansive than the requirement of equality.[156] Approaches to justice based on equality and discrimination are individualistic in their focus.[157] Fineman argues that a concentration on equality does not address social, economic, political and structural inequalities.[158] A focus on differential treatment based on the protected characteristics makes it difficult to achieve substantive equality.[159] The equality analysis only focuses on the moment of the injury the discriminatory treatment inflicts rather than on how historical, systemic and institutional structures made it possible for someone to impose the injury.[160] Assessing the situation using formal equality as a benchmark can justify the inequalities in question.[161] Fineman cites the *Parents Involved in Community Schools v Seattle School District No 1* to substantiate this claim.[162] In this case the court ruled that the initiative of the school districts in Seattle to allocate children to schools so as to have a better racial representation in the school bodies[163] was unlawful.[164] The school did not meet the test of necessity for intervention to justify correcting racial isolation through a measure of affirmative action.[165] The school districts did not demonstrate that there was a history of intentional discrimination in the districts.[166]

The case study of the admissions process in New York schools using an algorithmic decision-making process[167] corroborates Fineman's argument. This discussion showed how a purportedly neutral treatment of students based on their grades conceals inequality in social relations. The inequality in the parents' income and unequal educational facilities place hurdles on children from poorer socioeconomic backgrounds to gaining admission to the schools of their choice. Broader evidence for Fineman's position is found in the fact that in the United States the elimination of affirmative action policies 'would reduce the number of black students in the most selective schools of law and medicine to less than 1 percent of all students.'[168] According to a university admissions officer, very few individuals from poor socioeconomic backgrounds meet the admissions criteria of top American universities.[169] This example illustrates the value of using the vulnerability theory as a framework for interrogating how the use of AI decision-making processes affects individuals and society. It opens new avenues for thinking about how the employment of AI decision-making processes bears on human diversity and whether the types of values these processes embed in society are conducive to protecting human diversity.

3. Thinking about AI decision-making processes as an institution that brings about transformative effects

The employment of AI decision-making processes is part of a larger trend of digital technologies transforming society.[170] There is a view that digital technologies are ushering in a Fourth Industrial Revolution.[171] The vulnerability theory is a fruitful lens for better understanding the AI decision-making processes and what kind of values these processes enact. It sheds light on some of the institutional and societal changes the employment of AI decision-making processes introduces from the perspective of social justice. As a result, one gains insight into how the cumulative use of automated decision-making processes is likely to impact on individuals and society at large from the vantage point of social justice.

Section 3.1 will investigate how the employment of AI decision-making processes impacts on individuals through changing the relationships individuals have to each other and the institutions. Marlies van Eck studied how the Dutch administrative bodies use systems[172] that combine information from different government databases[173] and autonomously determine[174] whether an individual is entitled to payment.[175] These systems do not use big data,[176] that lack self-learning capabilities[177] and execute procedures that lack a discretionary decision-making element.[178] She posits that the use of automated decision-making processes to make assessments regarding the entitlement of individuals to receive a payment from the state changes the relationship the citizens have with the administrative body.[179] Furthermore, they amend the relationship between administrative bodies.[180] This argument should be expanded and modified to fit the context of the AI decision-making process. It will be put forward that the AI

156 Fineman, 'Equality and Difference–the Restrained State' 609
157 Fineman, *Vulnerability and Social Justice* 15
158 Fineman, 'Equality and Difference–the Restrained State' 609
159 Ibid 610
160 Ibid 612
161 Ibid
162 Ibid 610-11; *Parents Involved in Community Schools v Seattle School District No 1* 551 US 701 (2007)
163 *Parents Involved in Community Schools v Seattle School District No. 1* 551 US 701 709-710 (2007)
164 Ibid 797-98
165 Ibid
166 Ibid
167 Roth, 'Why New York City's High School Admissions Process Only Works Most of the Time'
168 William G Bowen and Derek Bok, *The Shape of the River : Long-Term Consequences of Considering Race in College and University Admissions Twentieth Anniversary* (20 edn, Princeton University Press 2019) 282
169 Jerome Karabel, *The Chosen: the Hidden History of Admission and Exclusion at Harvard, Yale and Princeton* (Houghton Mifflin Company 2005) 537
170 Rabeh Morrar, Husam Arman and Saeed Mousa, 'The Fourth Industrial Revolution (Industry 4.0): A Social Innovation Perspective' (2017) 7 Technology Innovation Management Review 12, 13
171 Ibid
172 Eck, 'Geautomatiseerde Ketenbesluiten & Rechtsbescherming: Een Onderzoek Naar de Praktijk van Geautomatiseerde Ketenbesluiten Over een Financieel Belang in Relatie Tot Rechtsbescherming 42
173 Ibid 45
174 Ibid 42-43
175 Ibid 45
176 Ibid 47
177 Ibid 46
178 Ibid 45
179 Ibid
180 Ibid 204

decision-making process should be conceived of as an institution. This institution changes how the individual is embedded in a set of relationships with individuals and institutions. The sets of relationships the operation of the AI decision-making processes gives rise to are complex and interlocking. The employment of AI decision-making processes may be seen as bringing about a social order[181] and as modifying a societal architecture. Numerous transformations the AI decision-making processes bring about which are detrimental for human diversity will be discussed. Section 3.2 will assess what values the AI decision-making process enacts and how it bears on human diversity. It will engage with the concerns Mireille Hildebrandt and Karen Yeung raised relating to how the collection and analysis of data from many sources will impact on humanity[182] and social life.[183]

3.1. *AI decision-making processes reconfigure human, institutional and social relationships*

The present analysis is confined to the employment of AI decision-making processes to create a representation of the individuals, to estimate with a margin of error their performance and to reach a conclusion concerning an individual's entitlement to a positive decision. Since the attention is on selecting students for admission to a university, the enquiry excludes many applications of the process from the discussion, such as the use of AI processes to diagnose diseases[184] and to allocate public service resources.[185] The analysis centres on how the use of AI decision-making processes impacts on individuals who are most affected by being situated in unequal relationships. The findings can be extrapolated beyond the case study of the university admissions to the employment context and to contexts where a holistic assessment of the applicant and the exercise of discretion facilitate the attainment of socially just outcomes. The vulnerability theory lens makes it possible to identify some of the ways in which the use of AI decision-making processes will affect the relationships of individuals with each other and societal institutions. At the core of the vulnerability theory is a shift of focus from the individual and the individual's autonomy to how the state organises relationships and societal institutions.[186] Under the vulnerability theory framework, the AI decision-making process should be thought of as an institution. Technology in the course of its operation constitutes relationships between citizens, devices and infrastructures.[187] These relationships can be thought of as a network.[188] One of the relationships the use of the AI decision-making process gives rise to is between the applicant and the university. The AI decision-making process mediates this relationship by allocating candidates to positive and negative decision quadrants. It determines what combinations of characteristics entitle a candidate to be allocated a space at a university and channels how the candidates may communicate their qualities to the university.

The employment of the AI decision-making process gives rise to a relationship of subjugation between the applicant and the university through a process of erasure. Mathematical processes model the world as black and white.[189] The process of representing an individual as a cluster of quantifiable characteristics using a vector[190] pushes individuals into categories.[191] When the AI decision-making process evaluates the individual against a template of an optimum combination of characteristics, it is unable to account for the individuality of people, the uniqueness of their experiences as well as the full scope of the contribution they can make to organisations and society as a result of having these qualities. The AI decision-making process withholds resources from individuals who either have difficulty fitting the algorithmically constructed profile of a good candidate or whose aptitudes algorithmic processes cannot detect.

The proposed use of the AI decision-making process to score student essays as part of the university admissions process[192] illustrates that the analysis of data using algorithmic processes provides an incomplete representation of the person.[193] This representation fails to capture important information about the individuals and thereby impedes the recognition of their aptitudes. This problem is deeper than technologies misreading or failing to read certain bodies.[194] Everyone is affected. The lyrics of the Queen song 'I want to

---

[181] Florian Eyert, Florian Irgmaier and Lena Ulbricht, 'Algorithmic Social Ordering: Towards a Conceptual Framework' in Günter Getzinger (ed), *Conference Proceedings of the 17th STS Conference Graz 2018* (Technischen Universität Graz 2018) 48

[182] Mireille Hildebrandt, 'Slaves to Big Data. Or Are We?' (2013) 16 IDP Revista De Internet, Derecho Y Política 1, 2

[183] Karen Yeung, *A Study of the Concept of Responsibility for Artificial Intelligence Decision- making Systems with Human Rights Framework* (Council of Europe 2018) 22

[184] Jun Wu, 'AI and Medical Diagnosis' (*Medium*, 2019) <https://medium.com/@junwu_46652/ai-and-medical-diagnosis-261218de33a0> accessed 25 June 2019

[185] Sahil Jain, Naufal Khan and Anusha Dhasarathy, 'When Governments Turn to AI: Algorithms, Trade-offs, and Trust' (*McKinsey & Company*, 2019) <https://www.mckinsey.com/industries/public-sector/our-insights/when-governments-turn-to-ai-algorithms-trade-offs-and-trust> accessed 26 June 2019

[186] Fineman, 'Equality and Difference–the Restrained State' 617-18

[187] Mireille Hildebrandt, 'Legal and Technological Normativity: More (and Less) Than Twin Sisters' (2008) 12 Techné: Research in Philosophy and Technology 1, 5

[188] Lucy Suchman, 'Located Accountabilities in Technology Production' (2002) 14 Scandinavian Journal of Information Systems 91, 92

[189] Raad van State, *Advies W04.18.0230/I: Ongevraagd Advies Over de Effecten van de Digitalisering Voor de Rechtsstatelijke Verhoudingen* 4

[190] John Johnston, *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI* (MIT Press 2008) 312

[191] Seeta Peña Gangadharan and Jędrzej Niklas, *Between Antidiscrimination and Data: Understanding Human Rights Discourse on Automated Discrimination in Europe*, 2018) 21

[192] Tina Nazerian, 'Can AI Help Students—and Colleges—Determine the Best Fit?' *EdSurge* (Burlingame, 11 April 2018) <https://www.edsurge.com/news/2018-04-11-can-ai-help-students-and-colleges-determine-the-best-fit> accessed 3 May 2019

[193] Corien Prins and Lokke Moerel, *Privacy for the Homo Digitalis: Proposal For a New Legal Framework For Data Protection in the Light of Big Data and the Internet of Things* (Tilburg University 2016) 36

[194] Linnet Taylor, 'What is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally' (2017) 4 Big Data & Society 1, 5

break free'[195] performed by Freddie Mercury is chosen as a case study because it is a text where the singer both expresses his identity and communicates to society. In this song Mercury talks about his desire to 'break free' from society's lies, about being in love for the first time and about the need to 'make it on my own.'[196] The reader needs to know the context behind the text in order to understand the communication. The context pertains to society using categories to designate sexual identity[197] and society historically stigmatising homosexuality.[198] Furthermore, the capacity for empathy is a pre-requisite for understanding the entire communication in the text of the song, namely Mercury's feelings of anguish and resentment. Mercury talks about his love for his partner and the need to 'break free'[199] as an allusion to the decision to begin a homosexual relationship notwithstanding society's approbation of such conduct. The song is an expression of Mercury's individuality because it encapsulates his feelings, opinions, life experiences and struggles.

How the programmer designs an AI decision-making process will determine what words the system will designate as relevant for the analysis, how it constructs links between the words[200] and how it scores a particular combination of words. The AI decision-making process evaluates the similarity between words based on their distance in geometrical space.[201] Because it is difficult to use a mathematical model to represent context,[202] the AI decision-making processes cannot link the text to the societal context underlying the communication.[203] Consequently, it lacks the capacity to derive meaning from the text of Queen's song. In the course of mapping the song as a set of linkages between disparate words[204] the AI decision-making process excises from the text what the program does not allow it to detect. The AI decision-making process erases meaning, human expression, individuality and the narration of lived experiences. It precludes individuals from communicating the diversity of their experiences in essays and to explain what contributions they can make to society if admitted to the university programme. For instance, the AI decision-making process will not detect that the song 'I want to break free'[205] demonstrates the writer's creativity and capability to advocate for a more inclusive society. This concern may become attenuated if computer scientists find a way to imbue the AI decision-making processes with human qualities, such as the capacity for empathy,[206] abstract thought and the capacity to link the emotions to the communication content. These qualities facilitate the ability of individuals to understand the meaning behind text where individuals communicate their personal experiences and needs. The real question is whether society wishes to have synthetic persons displace living decision-makers.

Another way in which the use of the AI decision-making process subjugates individuals is arguably through pushing them to adopt particular patterns of behaviour.[207] According to Mireille Hildebrandt, individuals will adjust their behaviour in anticipation of how algorithmic processes operate in order to improve their chances of receiveing a favourable decision.[208] Individuals will write essays for the purpose of university admissions in a manner that reflects how the AI decision-making process carries out analysis on the data and in a manner that increases their chances of receiving a high score. This inhibits self-expression and the ability of individuals to communicate holistically about how they can make a contribution to society if selected for the university programme. Of course, some university admissions officers create barriers for candidates by discriminating against them on the basis of their identities.[209] However, universities can implement measures to reduce explicit and implicit bias[210] in order to facilitate a holistic assessment of candidates during the decision-making process. The problem with employing fully autonomous AI decision-making processes is that there is no human being to read the essay and other application materials holistically, to detect how the decision-making process may disadvantage a candidate and to report a problem. Human oversight may not be a practicable solution. The use of AI decision-making processes is redundant when a human decision-maker reads and carefully considers the application materials.

The way in which the AI decision-making process creates a model of human diversity reinforces existing systems of classification.[211] This fails to capture individual identities[212] and

[195] Queen, 'I Want to Break Free' (*STANDS4 LLC*, 2019) <https://www.lyrics.com/lyric/27041093/I+Want+to+Break+Free> accessed 20 June 2019

[196] Ibid

[197] Martha Albertson Fineman, 'Introduction: Feminist and Queer Legal Theory' in Martha Albertson Fineman, Jack E Jackson and Adam P Romero (eds), *Feminist and Queer Legal Theory: Intimate Encounters, Uncomfortable Conversations* (Routledge 2016) 4

[198] Siobhan Fenton, 'LGBT Relationships are Illegal in 74 Countries, Research Finds' *The Independent* (London, 17 May 2016) <https://www.independent.co.uk/news/world/gay-lesbian-bisexual-relationships-illegal-in-74-countries-a7033666.html> accessed 9 June 2019

[199] Queen, 'I want to break free'

[200] Tolga Bolukbasi and others, 'Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings' (2016) ARXIV 1, 1

[201] Ibid 3

[202] Clare Ann Gollnick, 'Induction is Not Robust to Search' in Irina Baraliuc and others (eds), *Being profiled: cogitas ergo sum* (Amsterdam University Press 2018) 147

[203] Hila Gonen and Yoav Goldberg, 'Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But Do not Remove Them' (2019) 1903.03862v1 arXiv 1, 5

[204] Bolukbasi and others, 'Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings' 1

[205] Queen, 'I Want to Break Free'

[206] Jörg Wellbrink, 'Roboter Am Abzug' (Talk delivered at Hotel Aquino, Berlin, 4 September 2013)

[207] Mireille Hildebrandt, 'Learning as a Machine: Crossovers Between Humans and Machines' (2017) 4 Journal of Learning Analytics 6, 7

[208] Ibid

[209] Wood, '"The Wolf of Racial Bias": the Admissions Lawsuit Rocking Harvard'

[210] Quinn Capers IV, 'Rooting Out Implicit Bias in Admissions' *Association of American Medical Colleges News* (Washington DC, 5 February 2019) <https://news.aamc.org/diversity/article/rooting-out-implicit-bias-admissions> accessed 12 June 2019

[211] Femke Snelting, 'Uniformity vs Diversity' (Robot Love Lectures "The Relation Between Universality and Diversity," Eindhoven, 2018)

[212] Ibid

institutes a relationship of subordination. Technologic design can reinforce hierarchies based on race,[213] gender[214] and colonial history[215] even when on the face of it the design purports to achieve fair outcomes. IBM proposes to designate individuals who have historically experienced discrimination into groups so as to allocate to them additional points during the decision-making process to ensure fairness.[216] What this approach misses is the problematic nature of designating individuals into categories using a system of classification. Queer legal theory rejects the use of categories to designate sexual orientation[217] as a means to end subordination of homosexuals to heterosexuals.[218] Queen's song 'I want to break free'[219] illustrates that homosexual individuals are just like the rest of humanity in their experience of love and quest for relationships. The use of AI decision-making processes subjugates groups by perpetuating a hierarchical system of classification and by shifting attention away from how the construction of societal relationships disadvantages individuals.

If subordination is to be ended, then a vulnerability analysis is preferable. A related concern with achieving affirmative action policies using AI decision-making processes can be gleaned from the observation of Sandra Wachter. Wachter explains that there are reasons why individuals may wish not to disclose that they have characteristics protected by the prohibition of discrimination.[220] This means that the initiatives by companies such as IBM to confer additional points to individuals with protected characteristics are likely to be ineffective. The focus should be on the vulnerabilities individuals experience as a result of being human[221] and on how the state can restructure social institutions[222] to ensure that all individuals have equal access to resources crucial to their flourishing.[223] This approach shifts the analysis to the root causes of inequality and on what steps the state should take to ensure that individuals are equally situated in relationships. Furthermore, the focus should be on developing frameworks to protect individuals from discrimination without individuals having to disclose that they have characteristics protected by non-discrimination statutes. Wachter for instance proposes that equality protection can be achieved without individuals disclosing their protected characteristics by prohibiting differential treatment on the grounds of there being an association between an individual and a group that has protected characteristics.[224]

The use of AI decision-making processes reconfigures relationships between individuals. Tobias Blanke maintains that geometric rationality underpins algorithmic decisions.[225] Algorithmic processes measure social relations.[226] The distance between data points representing individuals in a geometrical space designates the degree of similarity between individuals.[227] Blanke's proposition should be extended to argue that the employment of AI decision-making processes brings about a relationship between applicants. The decisions the computer scientist makes when gathering the data and building the AI decision-making infrastructure determine the contours of the relationship. The computer scientist's choices determine what similarities exist between individuals,[228] what resemblances become amplified and what similarities are attenuated. The nature of the relationship the computer scientist constructs between applicants depends on which personal attributes of the students the computer scientist includes in the decision-making process and what weight the computer scientist attaches to the attributes. The more weight the computer scientist puts onto the student's socioeconomic status, the less the relationship between students is characterised by individual autonomy. There are stronger elements of interdependence and solidarity in the relationship.

To illustrate, if the computer scientist models the students by only looking at their examination grades, the individuals who have lower examination grades due to experiencing social barriers will be grouped together. Students, such as young mothers who lack sufficient support, have lower high school grades.[229] The AI decision-making process will group many young mothers together because it focuses on detecting a similarity in the past performance on examinations. Since the AI decision-making process will designate these mothers as being poor candidates for university admissions, their unequal position within social institutions becomes solidified. These women will have diminished employment opportunities as a result of lacking a university degree. The use of the AI decision-making process weakens the relationship of solidarity, connection and interdependence between young mothers with insufficient support and students

[213] Ali Breland, 'How White Engineers Built Racist Code–and Why It's Dangerous for Black People' *The Guardian* (London, 4 December 2017) <https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police> accessed 15 June 2019

[214] Rachel N Weber, 'Manufacturing Gender in Commercial and Military Cockpit Design' (1997) 22 Science, Technology & Human Values 235, 235

[215] Lilly Irani and others, 'Postcolonial Computing: a Lens on Design and Development' (2010) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 1311, 1315

[216] IBM, 'AI Fairness 360 Open Source Toolkit' (*IB.*, 2019) <http://aif360.mybluemix.net/> accessed 17 June 2019

[217] Janet Halley, 'Queer Theory by Men' in Martha Albertson Fineman, Jack E Jackson and Adam P Romero (eds), *Feminist and Queer Legal Theory: Intimate Encounters, Uncomfortable Conversations* (Routledge 2016) 27

[218] Ibid 14

[219] Queen, 'I Want to Break Free'

[220] Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' (2019) 35(2) Berkeley Technology Law Journal 1, 9

[221] Fineman, 'Beyond Identities: The Limits of an Antidiscrimination Approach to Equality' 1752-53

[222] Ibid 1760

[223] Ibid 1761

[224] Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' 66-68

[225] Tobias Blanke, 'The Geometric Rationality of Innocence in Algorithmic Decisions' in Irina Baraliuc and others (eds), *Being Profiled: Cogitas Ergo Sum* (Amsterdam University Press 2018) 93

[226] Ibid

[227] Ibid

[228] Provost and Fawcett, *Data Science for Business* 157-58

[229] National Conference of State Legislatures, 'Not Making the Grade: Academic Achievement Difficult for Teen Parents' (*National Conference of State Legislatures*, 2019) <http://www.ncsl.org/research/health/teen-pregnancy-affects-graduation-rates-postcard.aspx> accessed 17 June 2019

with no care commitments. In societies that recognise the interdependence of individuals and the dependence of individuals on societal structures,[230] the government has a duty to intervene to redress unjust institutional relationships.[231] Its duty is to strengthen the resilience of individuals by providing them with resources.[232] The use of the AI decision-making process that focuses on choosing candidates with optimal predicted performance deepens unjust societal relationships because it constructs relationships characterised by individual autonomy. These relationships include those between candidates, and between the candidate and the university.

Consider now a case where the AI decision-making process incorporates information about factors that result in individuals being unequally situated in institutional relationships in order to account for social justice concerns. Navid Tanzeem and colleagues created an artificial intelligence software which as part of predicting the students' grades[233] depicts the correlation between the socioeconomic, psychological and educational background of the student.[234] There are practical and technical difficulties with constructing an AI decision-making process that accurately captures the disadvantage individuals experience by virtue of being situated in unequal relationships. For instance, partially sighted students recount how attitudes that they differ from the general population,[235] a lack of resources such as recorders,[236] and teachers making inadequate usage of assistive technology devices create learning barriers.[237] Lack of universal classroom and curriculum design[238] as well as societal biases result in partially sighted students being in less equal relationships than their peers with other students, teachers, the school and other social institutions. If the student in addition to being partially sighted comes from a one-parent family or has a parent who is ill then this student will experience unequal social relationships created by inadequate state financial, social and healthcare support.

Kimberle Crenshaw coined the concept of 'intersectionality' to explain that each person experiences discrimination and disadvantage differently.[239] Institutional structures, such as sexism and racism, impact on individuals in an intersecting[240] and multi-layered manner to disadvantage individuals.[241] Factors, such as poverty and lack of job skills, that are the result of class, gender and other forms of oppression compound the disadvantaged position of individuals.[242] The cumulative disadvantage individuals experience is greater than the sum of the factors giving rise to the disadvantage.[243] The concept of intersectionality[244] suggests that to understand the disadvantaged position of an individual one needs to consider the fluid manner in which all personal and institutional relationships in which an individual is situated interact.

A hurdle to employing AI decision-making processes is that it is difficult to translate how unequal relationships inhibit the ability of individuals to succeed onto a model using quantifiable variables. It is unclear how one can quantify the disadvantage a white female teenager experiences due to being bullied at school for having a bisexual orientation and due to having caring responsibilities for a relative. Neither can one compare with precision how the disadvantage this female teenager experiences compared to the barriers a male teenager of colour with a chronic health condition encounters. The inclusion of a set of factors into the geometrical representation of the individual and the giving of mathematical weight to these factors fails to accurately represent the experiences of individuals.

The testimonies of civil society organisations support the assertion that the use of mathematical formulas precludes the decision-making process from providing appropriate weight to the unjust societal relationships in which an individual is situated. Civil society organisations oppose a decision-making procedure where human decision-makers tick boxes and work like a computer.[245] They worry that algorithmic decision-making processes cannot take relevant factors into account concerning an individual and give appropriate weight to such factors.[246] While the process of quantifying unjust societal relationships strengthens a recognition that individuals are interdependent, it replaces socially embedded relationships with symbolic mathematical boundaries. This occurs because the AI decision-making process transforms individuals into a set of factors that designate social barriers. As the AI decision-making process allocates individuals into groups[247] and decision quadrants it erects symbolic boundaries. Societal relationships become partitioned and mechanised. This discussion points to the need to preserve human decision-making where the decision relates to evaluating the capability of the individual. Such decisions incorate deliberation how to intervene in order to remedy the

[230] Fineman, 'Equality and Difference–the Restrained State' 622
[231] Fineman, 'Beyond Identities: The Limits of an Antidiscrimination Approach to Equality' 1762
[232] Fineman, 'Equality and Difference–the Restrained State' 624
[233] ATM Shakil Ahamed, Navid Tanzeem Mahmood and Rashedur M. Rahman, 'An Intelligent System to Predict Academic Performance Based on Different Factors During Adolescence' (2017) 1 Journal of Information and Telecommunication 155, 158
[234] Ibid 157-58
[235] Kerri Janae Johnson-Jones, 'Educating Students with Visual Impairments in the General Education Setting ' (PhD thesis, The University of Southern Mississippi 2017) 81
[236] Ibid 82
[237] Ibid 83
[238] National Disability Authority, 'What is Universal Design: the 7 Principles' (*National Disability Authority,* 2014) <http://universaldesign.ie/What-is-Universal-Design/The-7-Principles/> accessed 26 June 2019
[239] Kimberle Crenshaw, 'Demarginalising the Intersection of Race and Sex: a Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics' (1989) 1 University of Chicago Legal Forum 139, 149
[240] Kimberle Crenshaw, 'Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color' (1991) 43 Stanford Law Review 1241, 1244
[241] Ibid 1245-46
[242] Ibid
[243] Crenshaw, 'Demarginalising the Intersection of Race and Sex: a Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics' 140
[244] Ibid 149
[245] Gangadharan and Niklas, *Between Antidiscrimination and Data: Understanding Human Rights Discourse on Automated Discrimination in Europe* 20
[246] Ibid
[247] Provost and Fawcett, *Data Science for Business* 24

unequal position of individuals in social and institutional relationships. Of course, if computer scientists succeed in replicating human capabilities fully in AI as well as in providing the AI with knowledge of the social processes, then the AI decision-making process will approximate human decision-making more closely. The closer the AI decision-making process replicates human decision-making, the fewer concerns there will be with delegating decision-making to machines.

In the course of restructuring relationships between students, and between the students and the university the operation of the AI decision-making process brings about transformations at societal level that influence the ability of individuals to realise themselves as human beings. Florian Eyert, Florian Irgmaier and Lena Ulbricht posit that AI decision-making processes structure social processes and generate a particular type of social order.[248] Abeba Berhane is of the view that AI decision-making processes 'restructure the very fabric of the social world.'[249] Social orders arise from individuals adopting particular beliefs, common interpretations of events and acceptable ways of behaving.[250] Such conduct gives rise to patterns of interaction[251] and produces a social system.[252] The cumulative use of AI decision-making processes to determine admission to educational institutions, selection for employment and whether to extend a loan to an individual produces a particular type of social order.[253] This social order amplifies the scientific tradition of elevating rationality above emotions, experience and the power of imagination.[254] AI decision-making processes structure how individuals navigate their relationships with each other. There is a danger that the operation of AI decision-making processes will act as a divisive force. As a result, there will be a deepening of segregation.[255] Individuals will adapt their behaviour in order to increase their chances of receiving a positive decision from the AI decision-making process.[256] Parents may encourage or nudge their children to associate with other children who have those habits or lifestyles that make it more likely that they will meet the criteria of optimal performance as designated by the AI decision-making process. Since the AI decision-making process associates optimal performance with a set of particular characteristics shared by a group,[257] parents and children may avoid individuals who lack such characteristics. Further support for this argument is found in the scholarship of Wachter. Sandra Wachter draws attention to the fact that the AI decision-making processes may profile people who have an association or 'affinity' with a group who are protected by antidiscrimination statutes by virtue of having shared interests.[258] For instance, the individual can have interests or engage in activities that have a linkage to the characteristics of the protected group[259] or to characteristics that functions as an indicator of a protected characteristic.[260] To illustrate, there may exist statistical evidence that members of an ethnic group like to consume Carribean food and listen to jazz music.[261] This discussion illustrates that the use of AI decision-making processes has wide-ranging secondary effects that legislators should take into account. Human decision-making is crucial for discerning the full array of human expression and creativity.

3.2. *The problematic impacts of AI decision-making processes on the advancement of human diversity*

The term human diversity is closely linked with how society makes sense of the lived human experiences. Diversity relates to the fact that there are a multitude of ways in which human beings can define their identity and that of other individuals.[262] Increasingly there is a recognition that identities entail complex representations rather than conceptualisations in terms of particular characteristics, such as one's ethnicity, gender or political opinions.[263] It follows from this that one of the ways of understanding the impact the use of AI decision-making processes has on the protection of human diversity is an analysis of how these processes bear on how society conceives of human beings. A relevant consideration is what values the operation of the AI decision-making processes embeds into society by virtue of prioritising certain values.[264] This aspect is relevant because how individuals conceive of people's identities bears on the position of individuals within relationships with other individuals and institutions. In turn, such positioning influences their life opportunities.[265] What is more, what values individuals are exposed to through social interactions shapes their perceptions and how they treat their fellow human beings. Data scientists purport to capture individuals and their lives. The process of modelling individuals in a technical system transforms them. Kevin D Haggerty and Richard V Ericson use Gilles Deleuze's and Felix Guattari's concept of

[248] Eyert, Irgmaier and Ulbricht, 'Algorithmic Social Ordering: Towards a Conceptual Framework' 48
[249] Abeba Birhane, 'Algorithmic Injustices-Towards a Relational Ethics' (BIAS in Artificial Intelligence and Neuroscience Transdisciplinary Conference, Nijmegen, 2019) 2
[250] Lawrence K Frank, 'What Is Social Order?' (1944) 49 American Journal of Sociology 470, 473
[251] Ibid
[252] Ibid 474
[253] Eyert, Irgmaier and Ulbricht, 'Algorithmic Social Ordering: Towards a Conceptual Framework' 48
[254] Maria Mies, 'Feminist Research: Science, Violence and Responsibility' in Vandana Shiva, Maria Mies and Ariel Salleh (eds), *Ecofeminism* (Zed Books 2014) 47
[255] James A Allen, 'The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining' (2019) 46(2) Fordham Urban Law Journal 219, 230
[256] Hildebrandt, 'Learning as a Machine: Crossovers Between Humans and Machines' 7
[257] Provost and Fawcett, *Data Science for Business* 81
[258] Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' 5-6
[259] Ibid
[260] Ibid 8
[261] Ibid 39
[262] Richard Crisp, 'Introduction' in Richard Crisp (ed), *The Psychology of Social and Cultural Diversity* (Blackwell Publishing 2010) 1
[263] Ibid
[264] Council of Europe Committee of Ministers, 'Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes Decl(13/02/2019)1' (1337th meeting of the Ministers' Deputies, Council of Europe 2019) <https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b> 15 February 2019
[265] Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' 23

a 'process of becoming'[266] to theorise how systems of surveillance operate on the body.[267] This argument can be extended to the context of an AI decision-making process. The use of AI decision-making processes initiates what Deleuze and Guattari call a 'process of becoming' or transformation into an alternative entity through the process of being represented.[268] Deleuze and Guattari illustrate the process of becoming by giving the following example. When a painter paints a bird, the bird becomes 'something else, a pure line and pure colour.'[269]

Haggerty's and Ericson's description of how the surveillance systems operate on the body resonates with how AI decision-making processes represent the individual as part of the decision-making process. According to Haggerty and Ericson, surveillance systems 'abstract' 'the human bodies from their territorial settings' and separate them into a series of flows.[270] They then reassemble the individuals through a series of data flows[271] into 'data doubles.'[272] This process transforms the body[273] and the individual[274] into a new type of entity that is 'pure information.'[275] The purpose of the representation is to enable decision-makers to distinguish between individuals rather than to generate an accurate portrayal of the individual.[276] Haggerty's and Ericson's scholarship suggests that individuals should be distinguished from their data and from the model the AI decision-making process generates. Support for this argument is found in feminist scholarship. Donna Harraway writes that individuals can only achieve partial knowledge.[277] The perspective of the observer shapes what the observer sees and fails to detect.[278] Harraway has written about how the application of scientific knowledge entails a process of translation.[279] The process of translating the real world into data representations using technology disembodies the subject[280] and produces a particular reality.[281] Vandana Shiva echoes Harraway when she points out that science renders individuals reductive and dispossesses them of their potential.[282] Science superimposes a narrative of objectivity on this process.[283] In fact, science is not universal and reflects particular values.[284]

The case study of using an AI decision-making process to select candidates for admission to a university based on their grades illustrates how AI decision-making processes have embedded within their architecture institutional mechanisms that construct differences between groups through defining the norm. Proponents of the AI decision-making processes argue that this technology can predict with a degree of accuracy a future performance on the chosen metric.[285] The grade is an example of a metric that educational institutions regard as an indicator of the student's intellectual ability and capacity to do well on the programme.[286] Educational institutions introduced tests as a means of promoting fairness by reducing the subjectivity of teachers' judgments about children.[287] Jeff Howard argues that intelligence is socially constructed rather than a fixed quality.[288] Fred P Pestello similarly emphasises the need to view grades as being a product of an organisational process rather than as an indicator of achievement.[289] It will be shown that intelligence and grades are social constructs that purport to capture something real but which in fact capture skills that have commercial or organisational value within a given society. When an AI decision-making process generates a representation of the student's capability and predicts the student's grade, the process measures how closely the student fits a social construct. In doing so the AI decision-making process perpetuates a current system of differentiation between groups that disadvantages certain groups of individuals.

David J Hand claims that the choice of criterion the data scientists create to distinguish between groups is subjective and arbitrary.[290] There is value in Hand's observation that AI decision-making processes that use metrics, such as the predicted grade, as a means to allocate university places have an element of subjectivity and arbitrariness. Society's understanding of intelligence is subjective. Social processes, such as the nature of available symbolic and material tools, play an important role in defining what is intelligence.[291] According to Catherine D'Ignazio and Lauren Klein, societies use clas-

---

[266] Gilles Deleuze and Félix Guattari, *A Thousand Plateaus: Capitalism and Schizophrenia* (University of Minnesota Press 1987) 304

[267] Kevin D Haggerty and Richard V Ericson, 'The Surveillant Assemblage' (2000) 51 British Journal of Sociology 605, 613

[268] Deleuze and Guattari, *A Thousand Plateaus: Capitalism and Schizophrenia* 304

[269] Ibid

[270] Haggerty and Ericson, 'The Surveillant Assemblage' 606

[271] Ibid 611

[272] Ibid 606

[273] Ibid 613

[274] Ibid 614

[275] Ibid 613

[276] Ibid 613-14

[277] Donna Haraway, 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective' (1988) 14 Feminist Studies 575, 586

[278] Ibid 587

[279] Ibid 580

[280] Ibid 581

[281] Ibid 589

[282] Vandana Shiva, 'Reductionism and Regeneration: A Crisis in Science' in Vandana Shiva, Maria Mies and Ariel Salleh (eds), *Ecofeminism* (Zed Books 2014) 24

[283] Haraway, 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective' 583

[284] Shiva, 'Reductionism and Regeneration: A Crisis in Science' 22

[285] Havan Agrawal and Harshil Mavani, 'Student Performance Prediction using Machine Learning' (2015) 4 International Journal of Engineering Research & Technology 111, 113

[286] Robin Matross Helms, *University Admission Worldwide* (International Bank for Reconstruction and Development 2008) 3

[287] Sternberg, 'Myths, Countermyths, and Truths about Intelligence' 14

[288] Jeff Howard, 'Getting Smart: the Social Construction of Intelligence' (1992) 6 The International Network of Principals' Centers 3, 9

[289] Fred P Pestello, 'The Social Construction of Grades' (1987) 15 Teaching Sociology 414, 414

[290] David J Hand, 'Classifier Technology and the Illusion of Progress' (2006) 21 Statistical Science 1, 10

[291] Ashley E Maynard, Patricia M Greenfield and Kaveri Subrahmanyam, 'Technology and the Development of Intelligence: From the Loom to the Computer' in Robert J Sternberg and David D Preiss (eds), *Intelligence and Technology: the Impact of Tools on the Nature and Development of Human Abilities* (Routledge 2011) 29

sification systems to construct categories.[292] Categories, such as male/female and white/racialised, have social, cultural and political values embedded within them.[293] Larger social structures, such as legal frameworks, social structures, and cultural values operate to create and entrench the classification systems.[294] Their argument can be extended to the categories of intelligence and intellectual disability. The grade in an examination serves as a system of classification that empowers some groups while disadvantaging others. Psychologists define intelligence as having an ability to reason, to solve problems, to think abstractly, to learn from experience and to learn quickly.[295] The definition of intelligence is complemented by the category of its lack, namely an intellectual disability. The American Association on Intellectual and Developmental Disabilities defines an intellectual disability in terms of a significant limitation in intellectual functioning and adaptive behaviour.[296] There is significant limitation in an individual's cognitive capability whenever an individual scores two standard deviations below the mean on the intelligence quotient (hereinafter IQ) test.[297] Intelligence tests require individuals either to use different types of reasoning to solve problems or to carry out a multitude of tasks that require an individual to employ different types of cognitive capabilities.[298]

Social constructionists maintain that the culture of testing creates a disability.[299] Society labels some individuals as handicapped through applying to them a series of practices, procedures and policies of classification.[300] Consequently, the distinction between citizens with and without an intellectual disability is artificial. In agrarian societies that do not prioritise literacy, intellectual knowledge and having a knowledge-based economy, all individuals can participate fully in society.[301] The type of activities that society prizes has an impact on how that society defines intelligence. Psychologists who create intelligence tests fail to account for the fact that the Arabic numbers engender a system of thinking that human beings created.[302] IQ tests measure how well individuals can apply skills that they learn through school instruction. The skills refer to how well an individual can engage with societally created symbols and images, such as numbers and letters.[303] Written scripts and numerals restructure intellectual activity[304] and channel how individuals think.[305] The systems of testing occlude the fact that non-academic activities, such as farming, require intelligence. Farmers need to know under what conditions a plant thrives.[306] They have to find solutions to overcome unfavourable environmental conditions[307] and to develop new farming practices to respond to environmental changes.[308] The selection processes that focus on identifying individuals with highest intelligence based on examination scores reproduces a classification system that creates hierarchies between groups. The classification system is intimately intertwined with a knowledge and technology-driven economy.

The process of designing clothes is another example illustrating the socially constructed nature of human intelligence and the arbitrary process of conferring recognition for talent. Fashion designer Alexander McQueen made a dress for the Voss collection number S/S 2001 by layering blood-red ostrich feathers with red glass microscope slides.[309] His goal was to give women the beauty of a bird.[310] Social values and culture determine whether society recognises a dress made from microscope slides as original and beautiful. Similarly, fashion commentators, mass media and marketers play a role in shaping the public perceptions regarding whether such a dress enhances female beauty. The example of fashion design illustrates that social constructivists are correct in their claim that culture, social structures and social processes play a role in determining what society perceives as exhibiting creativity, intelligence and merit. One of the reasons why society does not recognise the imagination and the creativity of individuals it labels as having an intellectual disability is because society does not create a market where the products

[292] Catherine D'Ignazio and Lauren Klein, 'Data Feminism' (*MIT Press Open,* 2019) chapter 3 <https://bookbook.pubpub.org/data-feminism> accessed 3 January 2019

[293] Ibid

[294] Ibid

[295] Linda S Gottfredson, 'Mainstream Science on Intelligence: An Editorial With 52 Signatories, History, and Bibliography' (1997) 24 Intelligence 13, 13

[296] American Association on Intellectual and Developmental Disabilities, 'Definition of Intellectual Disability' (*American Association on Intellectual and Developmental Disabilities,* 2019) <https://aaidd.org/intellectual-disability/definition> accessed 6 June 2019

[297] Ibid; Robert L Schalock, 'The Evolving Understanding of the Construct of Intellectual Disability' (2011) 36 Journal of Intellectual & Developmental Disability 223, 226

[298] Roberto Colom and others, 'Human Intelligence and Brain Networks' (2010) 12 Dialogues in Clinical Neuroscience 489, 202

[299] Dan Goodley, *Disability Studies: An Interdisciplinary Introduction* (SAGE Publications Limited 2010) 58

[300] Shelley Tremain, 'Foucault, Governmentality and the Critical Disability Theory Today: a Genealogy of the Archive' in Shelley Tremain (ed), *Foucault and the Government of Disability* (University of Michigan Press 2015) 14

[301] Tom Shakespeare, *Disability Rights and Wrongs Revisited* (Routledge 2013) 61

[302] Preiss and Sternberg, 'Technologies for Working Intelligence' 189

[303] Ibid 184

[304] Jerome Seymour Bruner, *Toward a Theory of Instruction* (Harvard University Press 1966) 112

[305] Preiss and Sternberg, 'Technologies for Working Intelligence' 199

[306] Secretariat of the Convention on Biological Diversity, 'Biodiversity, Food and Farming for a Healthy Planet' (*Secretariat of the Convention on Biological Diversity*, 2019) <https://www.cbd.int/ibd/2008/youth/farmers/> accessed 10 June 2019

[307] The University of Minnesota Department of Horticultural Science, 'Pests and Diseases: An Introduction' (*The University of Minnesota,* 2019) <https://smfarm.cfans.umn.edu/pests-and-diseases> accessed 11 June 2019

[308] Paulina Moses, 'Onyaanya Farmers Adopt New Practices' *New Era Newspaper* (Windhoek, 3 May 2019) <https://neweralive.na/posts/onyaanya-farmers-adopt-new-practices> accessed 18 June 2019

[309] Ariane Fennetaux, 'Birds of a Feather: Alexander McQueen's Victorian Bestiary' (2018) 88 Cahiers Victoriens et Édouardiens par 12

[310] Chloe Fox, *Vogue on Alexander McQueen* (Quadrille Publishing 2012) 105

of the self-expression of these individuals have commercial value.

Catherine D'Ignazio and Lauren Klein suggest that binary distinctions erase the experiences of certain groups and conceal systems of power that position groups in unequal relationships.[311] The binary distinction between intelligence and intellectual disability does not reflect how a variety of individuals experience the reasoning process and creativity. These definitions confer value on the skills of dominant groups while designating other individuals as lacking cognitive capacity. Students who are labelled as having an intellectual disability have greater difficulty charting a life path of their choice and accessing opportunities.

Given that human intelligence and grades are social constructs, AI decision-making processes are not just technologies for predicting performance with a degree of accuracy. The AI-based analytics entrenches existing social mechanisms that define the norm and deviation from the norm and which produce inequality between groups. One of the perceived values of the AI decision-making process is that it enables optimisation.[312] The AI decision-making process identifies characteristics associated with optimal or best performance in the course of its operation.[313] It achieves this by finding correlations in the data for how different combinations of characteristics are linked to attaining the highest grades in examinations. By focusing on identifying individuals who perform best in examinations and on characteristics that are linked to top performance, the AI decision-making processes further entrenches a mechanism that designates some groups as inferior. The narrative of a human being with an optimum set of characteristics is ableist in nature. Fionna Kumari Campbell defines ableism as a network of beliefs, processes and practices that produce the notion of a perfect human body of the human species.[314] The characteristics that the AI decision-making process identifies as corresponding to top performance come to designate who has the perfect body and who does not. This corroborates the assertion of Miro Griffiths that artificial intelligence technology is rooted in ableism.[315]

The notion that some individuals are top performers by virtue of possessing a set of optimum characteristics has a degree of similarity with eugenics. Eugenicists used biometrics to promote the idea that some groups were superior to others.[316] They measured physical and mental attributes,[317] and analysed what correlations existed between characteristics.[318] Eugenicists used their calculations to argue that poverty stemmed from genetically inherited traits.[319] While data scientists do not pursue a goal of privileging one group over another, the use of AI decision-making processes to create predictive models can lead to groups being designated as inferior. AI decision-making processes create a model that associates groups with particular combinations of characteristics. This stems from the fact that an AI decision-making process generates a model by finding patterns in the data[320] and by grouping individuals together on the basis of similarity.[321] The process of judging some groups as superior to others on the basis of having a particular set of characteristics has a degree of similarity with eugenics.

Traditional decision-making processes where human beings select students for university admissions differ from AI decision-making processes. Admissions officers rank candidates based on their grades. They do not purport to create a report about the individual's capability and about the relative capability of groups. On the other hand, AI decision-making processes communicate a range of narratives about individual candidates and groups in the course of generating a model. The model constructs a narrative about the characteristics of an individual, how the individual compares to other individuals, what characteristics of each candidate contributed to either success or failure as well as how shared group characteristics are related to top performance. The fact that societies already designate some individuals into stigmatising categories suggests that societies will associate individuals and groups with inferiority on the basis of the predictive models the AI decision-making processes generate. Consequently, the use of AI decision-making processes is likely to result in stigmatisation of groups[322] and in exacerbating unjust processes of differentiation.

The vulnerability theory draws attention to the fact that how we define a good applicant through implementing a decision-making process matters. The vulnerability theory states that being human[323] means being a vulnerable, embodied being who is susceptible to internal and external forces.[324] This is a very different narrative from that of optimal performance and ableism inherent in AI decision-making processes.

---

311 D'Ignazio and Klein, 'Data Feminism' chapter 3.

312 James Manyika and others, 'What's Now and Next in Analytics, AI, and Automation' (*McKinsey Global Institute,* 2017) <https://www.mckinsey.com/featured-insights/digital-disruption/whats-now-and-next-in-analytics-ai-and-automation> accessed 11 June 2019

313 Kerryn Kohl, 'The Rise of AI in Talent Management' (*The Talenttalks Africa,* 2019) <https://www.talenttalks.net/rise-ai/> accessed 11 June 2019

314 Fiona AK Campbell, 'Inciting Legal Fictions-Disability's Date with Ontology and the Ableist Body of the Law' (2001) 10 Griffith Law Review 42, 44 footnote 5

315 Miro Griffiths, *Miro Griffiths on AI and Technology "Rooted in Ableism" and Social Inequalities*, 1:40-1:45 accessed 23 October 2019, https://www.youtube.com/watch?v=sI99ZoE444M.

316 Francisco Louçã, 'Emancipation Through Interaction–How Eugenics and Statistics Converged and Diverged' (2009) 42 Journal of the History of Biology 649, 655

317 Alain Desrosières, *The Politics of Large Numbers–a History of Statistical Reasoning* (Harvard University Press 1998) 329

318 Louçã, 'Emancipation Through Interaction–How Eugenics and Statistics Converged and Diverged' 655

319 Donald MacKenzie, 'Statistical Theory and Social Interests: A Case-Study' (1978) 8 Social Studies of Science 35, 59

320 Provost and Fawcett, *Data Science for Business* 25

321 Ibid 24

322 Bart Jan van Ettekoven and Corien Prins, 'Data Analysis, Artificial Intelligence and the Judiciary System' in Vanessa Mak, Eric Tjong Tjin Tai and Anna Berlee (eds), *Research Handbook in Data Science and Law* (Edward Elgar Publishing 2018) 442

323 Fineman, 'Equality and Difference–the Restrained State' 614

324 Fineman, 'Beyond Identities: The Limits of an Antidiscrimination Approach to Equality' 1753

The vulnerability theory recognises that social relations and conventions construct differences.[325] They exclude individuals from participation and mark them as inferior.[326] The use of vulnerability theory leads the analyst to treat deviations from the average, such as delayed speech development, as part of what it means to be human.[327] This lens of analysis makes it unnecessary to refer to the fact that Albert Einstein did not speak before he turned four years old[328] to normalise the experience of children who start to speak later than their peers. The vulnerability theory calls for the decision-maker to explore the multitude of positive qualities that each individual has and to consider individuals in their entirety rather than as a set of disjointed characteristics. This approach is responsive to the fact that individuals who have disadvantaged backgrounds[329] want the decision-makers to recognise their humanity, to understand their life in context and to connect to them on a personal level.[330] Given the current state of scientific knowledge, AI decision-making processes lack the capacity to evaluate an applicant holistically. The vulnerability analysis calls for a preservation of human decision-making where a holistic assessment of individuals is crucial for evaluating their capabilities, contributions and how unjust societal relationships prevent individuals from realising their potential.

The vulnerability theory puts on the agenda the significance of the issue of how AI decision-making processes bear on the experience of individuals of being human. Eubanks views the process of describing an individual as a numerical value as dehumanising.[331] Reuben Binns and colleagues report that the individuals they interviewed expressed a concern that the use of statistical inference methods to make decisions about individuals reduces them to a percentage.[332] Mireille Hildebrandt argues that the massive collection and analysis of data from digital and non-digital sources will have an impact on what kind of humans we will become.[333] Karen Yeung raises the concern that artificial intelligence technologies simulating human behaviour will dehumanise social life[334] by displacing human interaction and compassion.[335] The AI decision-making process produces a technocratic vision of what it means to live a human life that excises the lived experiences of the individuals. The vulnerability theory analysis illustrates why this conception of a human life is harmful, inaccurate and unrealistic. Crucially, the conceptualising of individual identities using mathematical methods impairs the protection of human diversity. It is desirable to preserve human decision-making processes in order for the decision-making procedure to fully accommodate the lived experiences of individuals. One of the contexts in which human decision-making should be preserved is when the decision relates to evaluating the individuals as in terms of their capabilities. More broadly, the representation of individuals in geometric space undermines the protection of human diversity.

Conclusions

Given that the cumulative use of AI decision-making processes is transforming society[336] and embedding a set of values, the citizens and governments should give careful thought to what kind of society they want to have. It is important that the perspectives of individuals who have historically experienced disadvantage and discrimination have particular weight in the discussion. Some individuals will find it easier to adapt to societal changes brought about by technological development than others. The discussion showed how the distinction between individuals, such as intelligence and intellectual disability,[337] arises because individuals have unequal possibilities for adapting to the demands of a technological society. The present discussion underscores the importance of preserving human decision-making processes for certain types of decisions. The state should adopt legislation specifying the contexts in which there is a requirement for human decision-makers to reach decisions. One such context is where organisations use the AI decision-making processes to evaluate the capabilities of individuals. Examples of contexts where the decision-maker assesses the capabilities of the individual are education, employment and banking. The extent to which the same concerns apply to a situation where the human decision-maker uses the analysis an AI system generates depends on the purpose for which the AI system is designed.

There is a potential that the use of AI decision-making processes will create a barrier for citizens to understand how unjust social relationships inhibit their ability to succeed. Citizens may find it difficult to comprehend the technical design of the AI decision-making process. They will struggle to detect how the subjective choices computer scientists make at every stage of constructing the AI decision-making process[338] bear on whether they get a favourable outcome. As a result, individuals will blame themselves for not succeeding instead of challenging the AI decision-making process as an institution. Furthermore, the employment of AI decision-making processes obscures how the failure of the state to intervene to correct unjust institutional relationships leads to individuals being denied access to crucial opportunities, such as university education. The capacity of individuals to exert pressure on the

[325] Fineman, 'Equality and Difference–the Restrained State' 619
[326] Ibid
[327] Ibid 620
[328] Yiyun Li, 'Einstein Didn't Talk Until He was Four' *The Guardian* (London, 2 March 2005) <https://www.theguardian.com/lifeandstyle/2005/mar/02/familyandrelationships.features11> accessed 17 June 2019
[329] Eubanks, *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor* 168
[330] Ibid
[331] Ibid 152
[332] Reuben Binns and Others, 'It's Reducing a Human Being to a Percentage; Perceptions of Justice in Algorithmic Decision' (ACM Conference, Montreal, April 2018)
[333] Mireille Hildebrandt, 'Slaves to Big Data. Or Are We?' (2013) 16 IDP Revista De Internet, Derecho Y Política 1, 2
[334] Karen Yeung, *A Study of the Concept of Responsibility for Artificial Intelligence Decision- making Systems with Human Rights Framework* (Council of Europe 2018) 22
[335] Ibid
[336] Birhane, 'Algorithmic Injustices-Towards a Relational Ethics' 2
[337] Schalock, 'The Evolving Understanding of the Construct of Intellectual Disability' 225
[338] Friedler, Scheidegger and Venkatasubramanian, 'On the (Im)possibility of Fairness' 3

government to intervene to remedy unjust relationships will be affected. It is easier for individuals to use their life experience to communicate about social injustices than to construct a complex account of how their embeddedness in a relationship with the AI decision-making process intersects with their position in unjust societal relationships. The citizens will have a diminished capability to hold their government accountable and to press for appropriate state policies. This is deeply problematic from the perspective of advancing social justice. More broadly, as Martha Fineman points out, the state should be responsive to structural inequality and should reform institutions in a manner that allocates resources equally.[339] Instead of turning to purely technological solutions to tackle bias and unfairness the government should actively tackle social inequalities.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FURTHER READING

Adler M. Fairness in Context. J. Law Soc. 2006;33:615.

Agrawal H, Mavani H. Student performance prediction using machine learning. Int. J. Eng. Res. Technol. 2015;4:111.

Ahamed ATMS, Mahmood NT, Rahman RM. An intelligent system to predict academic performance based on different factors during adolescence. J. Inf. Telecommun. 2017;1:155.

Allen JA. The colour of algorithms: an analysis and proposed research agenda for deterring algorithmic redlining. Fordham Urban Law Journal 2019;46:219.

Allo P. Mathematical values and the epistemology of data practices. In: Baraliuc I, editor. Being Profiled: Cogitas Ergo Sum others. Amsterdam University Press; 2018.

American Association on Intellectual and Developmental Disabilities. Definition of intellectual disability. American Association on Intellectual and Developmental Disabilities; 2019 https://aaidd.org/intellectual-disability/definition accessed 6 June 2019.

Australian Human Rights Commission. Decision making and artificial intelligence. Australian Human Rights Commission; 2019 https://tech.humanrights.gov.au/decision-making-and-artificial-intelligence accessed 22 January 2020.

Bailey J, Steeves V. Introduction: cyber-Utopia? Getting beyond the binary notion of technology as good or bad for girls. In: Bailey J, Steeves V, editors. eGirls, eCitizens: Putting Technology, Theory and Policy into Dialogue with Girls' and Young Women's Voices. University of Ottawa Press; 2015. p. 1.

Barocas S, Selbst AD. Big data's disparate impact. California Law Review 2016;104:671.

Benthall S, Haynes BD. Racial Categories in Machine Learning. Association for Computing Machinery; 2019. p. 1.

Bielby WT, Baron JN. 'Men and women at work: sex segregation and statistical discrimination. Am. J. Sociol. 1986;91:759.

Binns R. Imagining Data, Between Laplace's demon and the rule of succession. In: Baraliuc I and others, editors. Being profiled: cogitas ergo sum. Amsterdam University Press; 2018a. p. 180.

Binns R and others. It's reducing a human being to a percentage: perceptions of justice in algorithmic decision. ACM Conference, Montreal, April 2018b.

Birhane A. 'Algorithmic injustices-towards a relational ethics. BIAS in Artificial Intelligence and Neuroscience Transdisciplinary Conference, Nijmegen, 2019.

Blanke T. The geometric rationality of innocence in algorithmic decisions. In: Baraliuc I, editor. Being Profiled: Cogitas Ergo Sum. Amsterdam University Press; 2018. p. 93.

Bolukbasi T. and others, 'Man is to computer programmer as woman is to homemaker? Debiasing Word Embeddings' (2016) ARXIV 1607.06520v1 [cs.CL] 1

Bourdieu P. Les Trois États du Capital Culturel. Actes Rech. Sci. Soc. 1979;30:3.

Bourdieu P. Distinction: A Social Critique of the Judgment of Taste. Harvard University Press; 1984.

Bovens M, Zouridis S. From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. Public Adm. Rev. 2002;62:174.

Breland A. 'How white engineers built racist code–and why it's dangerous for black people. The Guardian; 4 December 2017 London https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police accessed 15 June 2019.

Brownsword R., 'Regulatory cosmopolitanism: clubs, commons, and questions of coherence' (2010) 018/2010 TILT Law & Technology Working Paper 2.

Bruner JS. Toward a Theory of Instruction. Harvard University Press; 1966.

Bull A. 'Reproducing class?. Classical Music Education and Inequality. Discover Society; 2014 https://discoversociety.org/2014/11/04/reproducing-class-classical-music-education-and-inequality/ accessed 14 May 2019.

Butterly J. 7 Admissions Officers Share the Things They Never Tell Applicants. Insider Inc; 2018 https://www.businessinsider.com/7-things-college-admissions-officers-wish-every-applicant-knew-2018-2?international=true&r=US&IR=T accessed 26 June 2019.

Campbell FAK. Inciting legal fictions-disability's date with ontology and the ableist body of the law'. Griffith Law Rev. 2001;10:42.

Capers Q IV. 'Rooting out implicit bias in admissions. Association of American Medial Colleges News; 5 February 2019 Washington DC https://news.aamc.org/diversity/article/rooting-out-implicit-bias-admissions/ accessed 12 June 2019.

Carusi A. Beyond Anonymity: data as Representation in E-research Ethics. Int. J. Int. Res. Ethics 2008;37:37.

Colom R. Human intelligence and brain networks. Dialogues Clin. Neurosci. 2010;12:489.

Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence. Draft Recommendation of the Committee of Ministers to Member States On the Human Rights Impacts of Algorithmic Systems. Council of Europe; 2019.

Crenshaw K. Demarginalising the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ. Chicago Legal Forum 1989;1:139.

Crenshaw K. Mapping the margins: intersectionality, identity politics, and violence against women of color. Stanford Law Rev. 1991;43:1241.

[339] Fineman, 'Equality, Autonomy and the Vulnerable Subject in Law and Politics' 26

D'Ignazio C, Klein L. Data Feminism. MIT Press Open; 2019 https://bookbook.pubpub.org/data-feminism accessed 3 January 2019.
Deleuze G, Guattari F. A Thousand Plateaus: Capitalism and Schizophrenia. University of Minnesota Press; 1987.
Desrosières A. The Politics of Large Numbers–A History of Statistical Reasoning. Harvard University Press; 1998.
Dokania E. Can AI help humans overcome bias?. The Seattle Globalist; 22 May 2019 Seattle https://www.seattleglobalist.com/2019/05/22/can-ai-help-humans-overcome-bias/83957 accessed 3 March 2019.
Eck BMA. Geautomatiseerde Ketenbesluiten & Rechtsbescherming: Een Onderzoek Naar de Praktijk van Geautomatiseerde Ketenbesluiten Over een Financieel Belang in Relatie tot Rechtsbescherming PhD thesis. Tilburg University; 2018.
Educational Testing Service. About the GRE general test. Educational Testing Service; 2019 https://www.ets.org/gre/revised_general/about accessed 22 July 2019.
Ettekoven BJ, Prins C. Data analysis, artificial intelligence and the judiciary system. In: Mak V, editor. Research Handbook in Data Science and Law. Edward Elgar Publishing; 2018. p. 425.
Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St Martin's Press; 2018.
Eyert F, Irgmaier F, Ulbricht L. Algorithmic social ordering: towards a conceptual framework. In: Getzinger G, editor. Conference Proceedings of the 17th STS Conference Graz 2018. Technischen Universität Graz; 2018. p. 48.
Fennetaux A. 'Birds of a feather: Alexander McQueen's Victorian Bestiary. Cahiers Victoriens et Édouardiens 2018;88:1.
Fenton S. LGBT Relationships are Illegal in 74 countries, research finds. The Independent; 17 May 2016 London https://www.independent.co.uk/news/world/gay-lesbian-bisexual-relationships-illegal-in-74-countries-a7033666.html accessed 9 June 2019.
Fineman MA. Beyond identities: the limits of an antidiscrimination approach to equality. Boston Univ. Law Rev. 2012;92:1713.
Fineman MA. Equality, autonomy and the vulnerable subject in law and politics. In: Grear A, Fineman MA, editors. Vulnerability: Reflections on a New Ethical Foundation for Law and Politics. Ashgate Publishing Limited; 2013. p. 13.
Fineman MA. Equality and difference–the restrained state. Ala Law Rev 2015;66:609.
Fineman MA. Introduction: feminist and queer legal theory. In: Fineman MA, editor. Feminist and Queer Legal Theory: Intimate Encounters, Uncomfortable Conversations. Routledge; 2016. p. 1.
Fineman MA. Injury in the unresponsive state: writing the vulnerable subject into neo-liberal legal culture. In: Bloom A, editor. Injury and Injustice: The Cultural Politics of Harm and Redress. Cambridge University Press; 2018. p. 50.
Fineman MA. Vulnerability and Social Justice. Emory University; 2019.
Finkel NJ, Harré R, Lopez J-LR. Commonsense morality across cultures: notions of fairness, justice, honor and equity. Discourse Stud. 2001;3:5.
Fox C. Vogue On Alexander Mcqueen. Quadrille Publishing; 2012.
Frank LK. What is social order? Am. J. Sociol. 1944;49:470.
Friedler S.A., Scheidegger C. and Venkatasubramanian S., 'On the (Im)possibility of Fairness' (2016) 1609.07236v1 arXiv 1.
Frys L, Staat C. University admission practices-France. Matching in Practice; 2016 http://www.matching-in-practice.eu/university-admission-practices-france/ accessed 1 August 2019.
G BW, Bok D. The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions Twentieth Anniversary. 20 edn. Princeton University Press; 2019.
Gangadharan SP, Niklas J. Between Antidiscrimination and Data: Understanding Human Rights Discourse On Automated Discrimination in Europe. London School of Economics and Political Science; 2018.
Geuss R. Public Goods, Private Goods. Princeton University Press; 2003.
Goertzel B, Wang P. Introduction: aspects of artificial general intelligence. In: Goertzel B, Wang P, editors. Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms Proceedings of the AGI Workshop 2006. IOS Press; 2007. p. 1.
Gollnick CA. Induction is not robust to search. In: Baraliuc I, editor. Being Profiled: Cogitas Ergo Sum. Amsterdam University Press; 2018. p. 147.
Gonen H. and Goldberg Y., 'Lipstick on a pig: debiasing methods cover up systematic gender biases in word embeddings but do not remove them' (2019) 1903.03862v1 arXiv 1.
Goodley D. Disability Studies: An Interdisciplinary Introduction. SAGE Publications Limited; 2010.
Gottfredson LS. 'Mainstream science on intelligence: an editorial with 52 signatories, history, and bibliography. Intelligence 1997;24:13.
Government of Canada. Directive on automated decision-making. Government of Canada; 1 April 2019. https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592, accessed 23 January 2019.
Griffiths M. Miro Griffiths on AI and technology rooted in ableism and social inequalities; 2019 https://www.youtube.com/watch?v=sI99ZoE444M, accessed 23 October 2019.
Greene D, Hoffman AL, Stark L. Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning. Hawaii: The Proceedings of the 52nd Hawaii International Conference on System Sciences; 2019.
Haggerty KD, Ericson RV. The Surveillant Assemblage. Br. J. Sociol. 2000;51:605.
Halley J. Queer Theory by Men. In: Fineman MA, editor. Feminist and Queer Legal Theory: Intimate Encounters, Uncomfortable Conversations. Routledge; 2016. p. 9.
Hand DJ. Classifier technology and the illusion of progress. Statis. Sci. 2006;21:1.
Haraway D. Situated knowledges: the science question in feminism and the privilege of partial perspective. Feminist Stud. 1988;14:575.
Hardt M. How Big Data is Unfair: Understanding Unintended Sources of Unfairness in Data Driven Decision Making. Medium Corporation; 2014.
Helms RM. University Admission Worldwide. International Bank for Reconstruction and Development; 2008.
Hildebrandt M. Legal and technological normativity: more (and less) than twin sisters. Techné 2008;12:1.
Hildebrandt M. Slaves to big data. or are we? IDP Revista De Internet, Derecho Y Política 2013;16:1.
Hildebrandt M. 'Learning as a machine: crossovers between humans and machines. J. Learn. Anal. 2017;4:6.
Howard J. Getting smart: the social construction of intelligence. Int. Netw. Principals' Centers 1992;6:3.
IBM. AI Fairness 360 Open Source Toolkit. IBM; 2019 http://aif360.mybluemix.net/ accessed 17 June 2019.
Irani L. Postcolonial computing: a lens on design and development. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Pages; 2010. p. 1311.
Jæger MM. Does Cultural Capital Really Affect Academic Achievement?; 2010 Working Paper Series CSER WP No 0001.
Jain S, Khan N, Dhasarathy A. When governments turn to ai: algorithms, trade-offs, and trust. McKinsey & Company; 2019

https://www.mckinsey.com/industries/public-sector/our-insights/when-governments-turn-to-ai-algorithms-trade-offs-and-trust accessed 26 June 2019.
Johnson-Jones KJ. Educating Students with Visual Impairments in the General Education Setting PhD thesis. The University of Southern Mississippi; 2017.
Johnston J. The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI. MIT Press; 2008.
Jordan B. Minimising the Risks of Unconscious Bias in University Admissions: 2017 Update on Progress. Universities and Colleges Admissions Service; 2017.
Karabel J. The Chosen: the Hidden History of Admission and Exclusion at Harvard, Yale and Princeton. Houghton Mifflin Company; 2005.
Kohl K. The rise of AI in talent management. The Talenttalks Africa; 2019 https://www.talenttalks.net/rise-ai/ accessed 11 June 2019.
Kohn NA. Vulnerability theory and the role of government. Yale J. Law Feminism 2014;26:2.
Kowarski I. How Colleges weigh applicants' extracurricular activities. US News; 2018 https://www.usnews.com/education/best-colleges/articles/2018-10-25/how-colleges-weigh-applicants-extracurricular-activities accessed 14 May 2019.
Lareau A, Weininger EB. Cultural capital in educational research: a critical assessment. Theory Soc. 2003;32:567.
Lehman-Frisch S. Segregation, spatial (in)justice, and the city. Berkeley Plann. J. 2011;24:70.
Li Y. Einstein didn't talk until he was four. The Guardian; 2 March 2005 London https://www.theguardian.com/lifeandstyle/2005/mar/02/familyandrelationships.features11 accessed 17 June 2019.
Loe BS. The Facets of Artificial Intelligence: a Framework to Track the Evolution of AI. Stockholm: International Joint Conferences on Artificial Intelligence Organization; 2018.
Louçã F. Emancipation through interaction–how eugenics and statistics converged and diverged. J. Hist. Biol. 2009;42:649.
MacKenzie D. Statistical theory and social interests: a case-study. Soc. Stud. Sci. 1978;8:35.
Manyika J. What's now and next in analytics, ai, and automation. McKinsey Global Institute; 2017 https://www.mckinsey.com/featured-insights/digital-disruption/whats-now-and-next-in-analytics-ai-and-automation accessed 11 June 2019.
Marria V. The Future of Artificial Intelligence In The Workplace. Forbes Media LLC; 2019.
Maynard AE, Greenfield PM, Subrahmanyam K. Technology and the development of intelligence: from the loom to the computer. In: Sternberg RJ, Preiss DD, editors. Intelligence and Technology: the Impact of Tools on the Nature and Development of Human Abilities. Routledge; 2011. p. 29.
McCarthy J. What is artificial intelligence? Basic questions. Stanford University; 2007 http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html accessed 13 May 2019.
McVeigh T. Are private tutors for children just the latest educational "arms race"?. The Guardian; 4 October 2015 London https://www.theguardian.com/education/2015/oct/04/private-tutors-arms-race-schools-parents accessed 15 May 2019.
Menon AK, Williamson RC. The cost of fairness in binary classification. Proce. Mach. Learn. Research 2018;81:1.
Mies M. Feminist research: science, violence and responsibility. In: Shiva V, editor. Ecofeminism. Zed Books; 2014. p. 36.
Morrar R, Arman H, Mousa S. The fourth industrial revolution (industry 4.0): a social innovation perspective. Technol. Innov. Manag. Rev. 2017;7:12.
Moses P. Onyaanya farmers adopt new practices. New Era Newspaper; 3 May 2019 Windhoek https://neweralive.na/posts/onyaanya-farmers-adopt-new-practices accessed 18 June 2019.
National Conference of State Legislatures. Not making the grade: academic achievement difficult for teen parents. National Conference of State Legislatures; 2019 http://www.ncsl.org/research/health/teen-pregnancy-affects-graduation-rates-postcard.aspx accessed 17 June 2019.
National Disability Authority. What is universal design: the 7 principles. National Disability Authority; 2014 http://universaldesign.ie/What-is-Universal-Design/The-7-Principles/ accessed 26 June 2019.
Nazerian T. Can AI Help students—and colleges—determine the best fit?. EdSurge; 11 April 2018 Burlingame https://www.edsurge.com/news/2018-04-11-can-ai-help-students-and-colleges-determine-the-best-fit accessed 3 May 2019.
Pestello FP. 'The social construction of grades. Teach Sociol 1987;15:414.
Preiss D, Sternberg R. Technologies for working intelligence. In: Preiss D, Sternberg R, editors. Intelligence and Technology: the Impact of Tools on the Nature and Development of Human Abilities. Routledge; 2005. p. 183.
Prins C, Moerel L. Privacy for the Homo Digitalis: Proposal for a New Legal Framework for Data Protection in the Light of Big Data and the Internet of Things. Tilburg University; 2016.
Provost F, Fawcett T. Data Science for Business. O'Reilly Media Inc; 2013.
Queen. I want to break free. STANDS4 LLC; 2019 https://www.lyrics.com/lyric/27041093/I+Want+to+Break+Free accessed 20 June 2019.
Rasinski KA, Scott, 'Culture LA, Values. Beliefs about economic justice. Soc. Justice. Res. 1990;4:307.
Rawls J. Justice as fairness. Philos. Rev. 1958;67:164.
Roth A. Why New York city's high school admissions process only works most of the time. Chalkbeat; 2015 https://www.chalkbeat.org/posts/ny/2015/07/02/why-new-york-citys-high-school-admissions-process-only-works-most-of-the-time accessed 15 May 2019.
Santelices MV, Wilson M. Unfair treatment? The case of freedle, the sat, and the standardisation approach to differential item functioning. Harv. Educ. Rev. 2010;80:106.
Secretariat of the Convention on Biological Diversity. Biodiversity, food and farming for a healthy planet. Secretariat of the Convention on Biological Diversity; 2019 https://www.cbd.int/ibd/2008/youth/farmers/ accessed 10 June 2019.
Schalock RL. The evolving understanding of the construct of intellectual disability. J. Intellect. Dev. Disabil. 2011;36:223.
Schauer F. Profiles, Probabilities and Stereotypes. Belknap Press; 2006.
Schmidtz D. The Elements of Justice. Cambridge University Press; 2006.
Shakespeare T. Disability Rights and Wrongs Revisited. Routledge; 2013.
Shiva V. Reductionism and regeneration: a crisis in science. In: Shiva V, editor. Ecofeminism. Zed Books; 2014. p. 22.
Silipo R. What's in a name? Artificial intelligence or data science?. BetaNews Inc; 2019 https://betanews.com/2019/02/05/artificial-intelligence-or-data-science/ accessed 14 May 2019.
Sloman J. Economics. 6 edn. Prentice Hall; 2006.
Smith C. 'Introduction. The History of Artificial Intelligence. University of Washington; 2006.
Snellen ITM. Het automatiseren van beschikkingen bestuurskundig beschouwd. In: Franken H, editor. Beschikken

En Automatiseren, Preadviezen Voor De Vereniging Voor Administratief Recht. Samsom HD Tjeenk Willink; 1993. p. 51.
Snelting F. Uniformity vs diversity. Robot Love Lectures The Relation Between Universality and Diversity; 2018 Eindhoven.
Stalder F. From inter-subjectivity to multi-subjectivity: knowledge claims and the digital condition. In: Baraliuc I, editor. Being Profiled: Cogitas Ergo Sum. Amsterdam University Press; 2018. p. 132.
Stark L. Algorithmic psychometrics and the scalable subject. Soc Stud Sci 2018;48:204.
State Raad van State. Advies W04.18.0230/I: Ongevraagd Advies Over de Effecten van De Digitalisering Voor De Rechtsstatelijke Verhoudingen. Raad van State; 2018.
State Raad van State, 'The council of state' (*Raad van State*, 2019) <https://www.raadvanstate.nl/talen/artikel/> accessed 25 July 2019
Sternberg RJ. Myths, countermyths, and truths about intelligence. Educat. Res. 1996;25:11.
Sternberg RJ, Williams WM. Does the Graduate Record Examination Predict Meaningful Success in Psychology. Yale University; 1994.
Story of Song. Queen: I want to break free in the works. Story of Song; 2020 https://storyofsong.com/story/i-want-to-break-free, accessed 25 May 2020.
Suchman L. Located accountabilities in technology production. Scandinavian J. Inf. Syst. 2002;14:91.
Taylor L. 'What is data justice? The case for connecting digital rights and freedoms globally. Big Data Soc. 2017;4:1.
The Sutton Trust. About us: our cause. The Sutton Trust; 2019 https://www.suttontrust.com/about-us/ accessed 14 May 2019.
The University of Minnesota Department of Horticultural Science. Pests and diseases: an introduction. The University of Minnesota; 2019 https://smfarm.cfans.umn.edu/pests-and-diseases accessed 11 June 2019.
Thornton M. The cartography of public and private. In: Thornton M, editor. Public and Private: Feminist Legal Debates. Oxford University Press; 1995. p. 2.
Tremain S. Foucault, governmentality and the critical disability theory today: a genealogy of the archive. In: Tremain S, editor. Foucault and the Government of Disability. University of Michigan Press; 2015.
Turner Cothers. Why America's schools have a money problem. United States National Public Radio; 2016 https://text.npr.org/s.php?sId=474256366 accessed 14 May 2019.
United Nations. Social Justice in an Open World: The Role of the United Nations. United Nations; 2006.
Velasquez M, Andre C. Justice and fairness. Issues Ethics 1990;3:1.
Vetzo M, Gerards J, Nehmelman R. Algoritmes En Grondrechten. Boom Juridisch; 2018.
Wachter S. Affinity profiling and discrimination by association in online behavioural advertising. Berkeley Technology Law Journal 2019;35:1.
Weber RN. Manufacturing gender in commercial and military cockpit design. Sci. Technol. Human Values 1997;22:235.
Wood J. "The Wolf of racial bias": the admissions lawsuit rocking Harvard. The Guardian; 18 October 2018 London https://www.theguardian.com/education/2018/oct/18/harvard-affirmative-action-trial-asian-american-students accessed 10 March 2019.
Wu J. AI and medical diagnosis. Medium; 2019 https://medium.com/@junwu_46652/ai-and-medical-diagnosis-261218de33a0 accessed 25 June 2019.
Yeung K. A Study of the Concept of Responsibility for Artificial Intelligence Decision Making Systems with Human Rights Framework. Council of Europe; 2018.
Young A. Inequality, the urban-rural gap and migration. Q. J. Econ. 2013;128:1727.