LA TROBE UNIVERSITY

NCSEHE
National Centre for Student Equity in Higher Education

Curtin University

Towards an *Inclusive Analytics* for Australian Higher Education

Bret Stephenson, Andrew Harvey, Qing Huang

2022

Make tomorrow better.

ncsehe.edu.au

Towards an *Inclusive Analytics* for Australian Higher Education

2022

Dr Bret Stephenson, La Trobe University

Prof. Andrew Harvey, Griffith University, formerly La Trobe University

Dr Qing Huang, La Trobe University

National Centre for Student Equity in Higher Education
Tel: +61 8 9266 1743
Email: ncsehe@curtin.edu.au
ncsehe.edu.au
Building 602 (Technology Park)
Curtin University
Kent St, Bentley WA 6102
GPO Box U1987, Perth WA 6845

Acknowledgements

The authors acknowledge the funding of the National Centre for Student Equity in Higher Education (NCSEHE).

For his early feedback and assistance with the original grant application, the authors would also like to thank Professor Damminda Alahakoon, Director, Research Centre for Data Analytics and Cognition, La Trobe Business School, La Trobe University.

Table of contents

Abbreviations

AI	Artificial Intelligence
AIEd	Artificial Intelligence in Education
ATAR	Australian Tertiary Admission Rank
CRISP-DM	Cross-Industry Standard Process for Data Mining
DESE	Department of Education, Skills and Employment
DVC	Deputy Vice-Chancellor
EA	Educational Analytics
EDM	Educational Data Mining
FATML	Fairness, Accountability and Transparency in Machine Learning
HESP	Higher Education Standards Panel
IA	Institutional Analytics
IDS	Indigenous Data Sovereignty
IS	Information Services
ICT	Information and Communication Technology
ML	Machine Learning
NLP	Natural Language Processing
NPE	Non-Participating Enrolments
PLA	Predictive Learning Analytics
RIO	Responsible Innovation Organisation(s)
RL	Reinforcement Learning
RPA	Robotic Process Automation
SES	Socio-Economic Status
SIS	Student Information System
STEM	Science, Technology, Engineering and Mathematics
UA	Universities Australia
WAM	Weighted Average Mark
XAI	Explainable Artificial Intelligence

Executive summary

Artificial intelligence (AI) and machine learning (ML) applications now quietly power countless automated decision-making, and predictive processes, across university business areas and throughout the student lifecycle. The recent challenge of the COVID-19 crisis, and the emergency shift to online learning, has also notably increased institutional interest in the adoption of AI/ML "business solutions." While advanced data analytics techniques can be responsibly deployed to advance student equity interests, if adopted uncritically they can also amplify social inequalities and historical injustice, often by stealth. Moreover, it is increasingly difficult for non-specialist university leaders and decision-makers to anticipate how the AI/ML applications their institutions adopt may be working to undermine their own strongly held commitments to student equity and diversity. The proprietary nature of commercial analytics-powered products and services can also serve to frustrate a university's attempts to audit the impact of these processes on equity students and equity interests more broadly.

In this report we identify the potential benefits of advanced analytics for student equity, and the institutional and cultural changes required for such potential to be fulfilled. We also argue, however, that the growing use of analytics involves risks and threats to student equity, further underlining the importance of institutional change, including educative and regulatory reform. We begin this report by providing a brief overview of the uses of advanced analytics within the higher education context. Analytic techniques now inform vast areas of the university and traverse the whole student lifecycle: from the recruitment and admissions of prospective students, through to the building of employability profiles of graduates. Part 1 also reviews many of the important conceptual and practical challenges involved in the quantification or datafication of equity, and equity cohorts, within the Australian context.

In Part 2 of this report, we outline the potential of analytics to protect and advance student equity. We highlight at least three related ways in which improved outcomes might be delivered for marginalised students, and for the broader cause of equity overall. First, analytics can help us to discover discrimination, including within historical processes such as admissions and course guidance. Second, analytics can help us to identify individual disadvantage, and move beyond our reliance on group membership assumptions. In the Australian context, this opportunity potentially enables institutions to move beyond the six conventional equity groups to consider dynamic behavioural indicators at the individual level. Proponents of analytics have long touted this capacity for a more sophisticated, individuated understanding of risk and disadvantage. We note, however, that this is a complex and contested area, and frequently requires consideration of protected characteristics, e.g., equity group membership, stereotype risks, and the perils of what has been called a "colour blind" approach. Third, we argue that analytics provides an opportunity to assess emergent and contingent forms of disadvantage, such as the impacts of COVID-19.

While the effective use of advanced analytics can demonstrably improve student equity, its uncritical adoption, and a failure to maintain effective oversight, can result in a dramatic undermining of equity goals. The report outlines these potential perils, across the four stages of the machine learning lifecycle: 1) data collection and preparation; 2) model building/learning or algorithmic training; 3) model evaluation and verification; and 4) model deployment. We consider various forms of "data bias" including historical bias, representation bias, measurement bias and aggregation bias, which can have the effect of reifying stereotypes, misrepresenting individuals, and exacerbating inequity. Further risks include "algorithmic bias" and interventions based on predictive analytics, such as tailored communications to "at risk" student. These analytics-driven interventions may be either ineffective or even counter-productive, in some cases leading to self-fulfilling prophecies of failure. Threats to privacy are also rising and remain priority areas for ethical and equitable implementation.

Collectively, our analysis reveals that advanced analytics is widely used but not closely monitored for its equity impacts throughout the university and its many business areas. Where equity protections do exist, these are typically limited to the teaching and learning functions of the university.

In the final section, we address the Fairness, Accountability and Transparency in Machine Learning (FATML) movement and provide a brief account of its constituent elements and limitations. While FATML typically focusses on technical aspects of machine learning, such as mathematical definitions of algorithmic fairness, we argue that this focus is necessary but insufficient to support student equity within the university. Defining, conceptualising, and prioritising equity requires an understanding of existing structural inequity and the broader frameworks in which advanced analytics operate. Such understanding itself requires engagement with philosophical, political and policy questions, and the expertise of staff outside the technology and analytics domains.

Ultimately, we highlight a pressing need for institutions to embed greater data literacy and equity consciousness across their organisations. Harnessing the potential of analytics to improve student equity requires a comprehensive institutional approach, and a range of sophisticated strategies and practices. In our discussion we briefly address the need for broader education and professional learning among both academic and professional staff. While not every staff member needs to understand the technical processes of advanced analytics, this knowledge needs to be distributed across the university, including within areas responsible for many of the analytics-informed interventions, e.g., student support staff, academic progression staff, and equity practitioners. Further initiatives could include greater monitoring and evaluation of the deployment of advanced analytics and of data-informed interventions. Such oversight could be partly provided by regulatory committees that embody specialist knowledge and that might operate similarly to existing ethics committees. Greater oversight and regulation should not, however, be an excuse for stifling institutional innovation. A program of inclusive analytics should instead provide the appropriate safeguards within which innovation can be leveraged to benefit all students in a spirit of inclusivity. Advanced analytics provides both an opportunity and a threat to student equity. Active engagement, education, and oversight are required to ensure that the emancipatory promise of technology can be fulfilled in the present age of advanced data analytics.

Recommendations

1. That universities develop data analytics policies and procedures that protect equity interests throughout the full student lifecycle and across all business areas.
2. That universities broaden distribution of *analytic expertise*, particularly within the DVC (Academic) divisions.
3. That universities broaden distribution of *equity and ethics expertise*, particularly including within data analytics (institutional research and performance), Information Services, and ICT divisions of the university.
4. That universities increase professional education of staff, including academics, engaged with analytics projects at each stage of the development and deployment process.
5. That universities establish in-house regulatory structures and professional expertise to ensure equity and fairness are protected through the deployment of advanced data analytics, e.g., standing committees to oversee analytics, similar to ethics committees.
6. That universities ensure that analytics-informed interventions are tailored, based on behavioural factors, and designed to reduce self-fulfilling prophecies based on immutable characteristics.
7. That universities regularly monitor and evaluate the analytics project lifecycle for impact on equity and “fairness” interests.
8. That universities work towards benchmarking/collective agendas, potentially involving Universities Australia (UA) leadership.
9. That universities conduct and facilitate further interdisciplinary research into the intersection of equity in higher education and advanced data analytics as an urgent priority.

Part 1: Background - Advanced analytics in the Australian higher education equity context

1.1 The proliferation of advanced analytics throughout the modern university

The effects of advanced data analytics processes are now felt by students long before they have applied for university admission. Primary and secondary school students are now exposed to countless commercial algorithms, data collection regimes, and data-hungry educational technologies well before any decisions are made concerning post-secondary education options (Williamson & Hogan, 2020). Among the modern university's many business functions, marketing departments are among the more eager adopters of data analytics tools for the purpose of the digital customer, or future student tracking, and the deployment of targeted marketing campaigns (MacMillan & Anderson, 2019). This is despite the well-known threats these practices pose to equity and diversity interests by targeting advertising – thereby impacting student aspirations for particular cohorts – to only those who embody a selected or "maximized" subset of personal, often demographic, attributes (Speicher et al., 2018).

Additionally, universities are increasingly using advanced data analytics processes to inform admissions decisions (Pangburn, 2019) and scholarship allocation (Kurniadi et al., 2018). Once a student is admitted, advanced data analytics techniques further power all manner of learning and teaching activities, including for example:

- intelligent or adaptive tutoring systems (Graesser et al., 2018)
- automated essay and exam marking (Ge & Chen, 2020; Stephens, 2001; Wang et al., 2021)
- predictive "risk" modelling for student failure and attrition (Anagnostopoulos et al., 2020; Foster & Siddle, 2020)
- automated academic dishonesty detection (Afuro & Mutanga, 2021; Jaramillo-Morillo et al., 2020)
- automated advising or recommender systems for course, unit, or career selection and sequencing (Engina et al., 2014; Lin et al., 2018).

Each of these applications comes with its own unique set of challenges, and potential harms, to equity and diversity interests.

The power and reach of these advanced analytics applications, for both good and ill, are further amplified by the growth in business partnerships between universities and well-known commercial data aggregators and vendors of cloud computing, digital advertising and machine learning (ML) and artificial intelligence (AI) products – e.g. Microsoft, Amazon (AWS), Google, Facebook and IBM (Lacity et al., 2017; Microsoft Australia Education, 2018). These partnerships, as Phan et al. (2021) have described, are increasingly putting university researchers in the field of AI Ethics, and affiliated fields, in an ethically "curious position" (p. 9). Wedged between the "Big Capital" of "Big Tech" – which funds much of the current university research in this area – and their own social, ethical, and professional commitments, university-based researchers must work to carefully reconcile the values of academe with the commercial interests of Big Tech that frequently support university research budgets (p. 10).

Importantly, the recent disruption caused by the COVID-19 crisis has also worked to increase the rate of adoption of advanced analytic tools and methods. As the crisis threatens to disrupt traditional admissions pathways, many universities are considering alternative predictive analytic models, and novel data sources, upon which admissions decisions may be based (Barnard, 2020). Additionally, the radical shift to online learning has further stoked the existing institutional imperative to digitally capture, machine analyses, and predict student learning behaviours in online environments (Williamson et al., 2020). The severe financial fallout of the crisis is also likely to quicken existing institutional interest in adopting cost-reduction strategies that leverage the process-automation afforded by advanced analytics techniques.

While ML/AI applications can potentially work in the student's own learning interests, they also present a demonstrable and, perhaps, insidious threat to the project of student equity (Eubanks, 2018; MacMillan & Anderson, 2019; O'Niel, 2016). Moreover, it is increasingly difficult for non-specialist university leaders and decision-makers to anticipate how the ML/AI applications their institutions adopt, may be working to undermine their own strongly held commitments to student equity and diversity. The proprietary nature of commercial ML/AI products and services can also serve to frustrate a university's attempts to audit the impact of these processes on equity students and equity interests more broadly (Zuboff, 2019). Notwithstanding this breadth and depth of activity, the deployment of analytics also occurs within a broader policy context, and understanding this context is also central to strengthening student equity.

1.2. Advanced analytics within the equity policy context

Concern for disadvantaged groups, or what are commonly referred to as "equity" groups, in Australian higher education has a long history (Anderson, 1983; Gale & Tranter, 2011; La Nauze, 1940) and can be traced back to the founding charters of many Australian universities, including its first, The University of Sydney (Brett, 2018, p. 18). However, the last 30 years of this history have been all but exclusively defined in terms first set out in the Australian government's policy statement, *A Fair Chance for All* (Commonwealth of Australia, 1990). Born out of the wider "Dawkins reforms" of the late 80s and early 90s, *A Fair Chance for All* (also referred to as the Framework) has had the effect of firmly concretizing the way we categorise and quantify equity groups up to today. Moreover, the Framework has set the tone for how subsequent Australian governments, and the university sector as a whole, interpret and understand the related although often underdefined terms of "disadvantage", "equity", and "social justice" in Australian universities.

As Harvey et al. (2016) have observed, *A Fair Chance for All* embodied an "explicitly stated belief in the need for higher education to promote fairness and social inclusion, interpreted as *proportional representation* at the level of class, gender and race" (p. 4, italics added). As the policy statement indicates:

> *The overall objective for equity in higher education is to ensure that Australians from all groups in society have the opportunity to participate successfully in higher education. This will be achieved by changing the balance of the student population to reflect more closely the composition of society as a whole.*
>
> (Commonwealth of Australia, 1990, p. 2)

Based upon broad sector-wide consultation, the policy document identified six now familiar equity groups that were determined to be "significantly under-represented in higher education:

- People from socio-economically disadvantaged backgrounds
- Aboriginal and Torres Strait Islander people
- Women, particularly in non-traditional courses and postgraduate study

- People with disabilities
- People from non-English-speaking backgrounds
- People from rural and isolated areas" (p. 10).

A Fair Chance for All was also careful to acknowledge that the six identified equity groups were not mutually exclusive and that "[m]any people suffer from multiple disadvantage…" (p. 10). Finally, the Framework encouraged institutions to explore and address other potential equity groups that may be incorporated into individual institutional equity plans. Going forward, however, the six identified groups would become the exclusive focus of a national data reporting, monitoring, and governance framework. It is important to note both the extraordinary longevity of the Framework and its limitations, both of which we have addressed elsewhere (Harvey et al., 2016).

In the decades since its establishment though, policy makers have largely side-stepped the politically and philosophically difficult task of theorising "equity" and "disadvantage" and focussed instead on the statistical methodologies implicated in a view of equity as *proportional representation*. In this way, equity in Australian higher education is fundamentally about data and quantification, or a counting of students from identified groups and analysing their proportional representation within universities. Equity is, in brief, a matter of counting students and seeking statistical balance – it is fundamentally a challenge of "seeing" the student population through datafication and analysis.

1.3. Challenges to quantifying equity in the age of advanced analytics

In addition, there are several important ways in which disadvantage or under-representation – when conceived as *proportional representation* – may fail to be adequately registered or quantitatively "seen". In other words, there are numerous critically important ways in which equity status may fail to be registered, quantified or "datafied" – that is, translated from "reality" into an accurate form of data that is useful for the purpose of governance and counting (Van Dijck, 2014). These conceptual and quantitative shortcomings in our operational definition of equity in Australian higher education are likely to be further amplified as advanced data analytics techniques and processes continue to proliferate within universities. Analytic processes are fundamentally and inescapably powered by underlying data. This means that we must remain cognisant of the fact that analytics rely on the always imperfect datafication of human behaviours and the equally troublesome (or political) processes of categorising and naming particular human cohorts. These challenges of "seeing" disadvantage through a process of datafication are numerous, but here we delineate just four of the main challenges which a project of inclusive analytics must grapple with.

First, the quantitative or statistical approach to equity has produced the negative effect of rendering, as statistically invisible, small groups of students who may experience disadvantage. As a particular group's size reflects a smaller and smaller portion of the overall population, it becomes increasingly difficult to establish statistical measures of proportionality that are significant or meaningful. This has been identified as an issue for numerous student groups such as: South Sea islanders (Martin, 1994, p. 7): care leavers (Harvey et al., 2015); caregivers (O'Shea, 2015); students from refugee backgrounds (Harvey & Leask, 2020); military veterans (Harvey et al., 2020); and prisoners (Willems et al., 2018). In this way, even when robust data may be available, relatively small groups are difficult to statistically represent within a proportional representation classification regimen.

Second, the data necessary for identifying disadvantage or under-representation may simply go uncollected, poorly collected, or undisclosed. In the Australian context, for example, Allen (2021) has described the way in which census data are collected has served to critically diminish our view of the country's ethnic diversity. She argues that "data and related

infrastructure should not create nor perpetuate unfairness and inequalities, especially in overlooking diversity" (p. 5). The same could be said of our university student population data which has been shaped largely by minimum requirements for government reporting.

Third, individuals and groups may be reluctant to engage with institutions or mobile device applications, both state and private, that are known to collect personal data for bureaucratic or commercial purposes – or what is increasingly described as "dataveillance" (Van Dijck, 2014). Brennan (2019), for example, has detailed the growing culture(s) of "opting out of digital media" also described as the "digital temperance movement". For our purposes, and for the purpose of the project of inclusive analytics, we must recognise that datafication, or the increasing digital data capture of our daily lives, does not offer researchers a neutral or objective paradigm for understanding our diverse social world (Van Dijck, 2014). For a variety of reasons, whole groups and sub-groups of people may go unrepresented, or partially represented, in the "big" data sets of universities, governments, or of corporate data aggregators. Their absence, in many important cases, should not be read as a deficit caused by a "digital divide", but as an act of personal agency.

Finally, an inclusive analytics paradigm must engage with emerging claims to Indigenous Data Sovereignty (IDS), and this may include various forms and instances of opting out of data collection and analysis. The growing Indigenous Data Sovereignty movement represents a critically important challenge to the logic of datafication – frequently rendered via culturally insensitive expropriation – and its presumed statistical objectivity which "is the data lens by which Indigenous Peoples are made visible" (Walter et al., 2021, p. 2). Indigenous Data Sovereignty researchers have worked to challenge how government data, like that collected via national census (Andrews, 2018), are frequently used, and problematically so, to relentlessly depict Indigenous deficits (Drew et al., 2016; Kukutai & Taylor, 2016; Kukutai & Walter, 2021; Rainie et al., 2019; Walter, 2016; Walter et al., 2021; Walter & Suina, 2019).

1.4. Aims and outline of this report: Towards and Inclusive Analytics

In this report, we reviewed and engaged with extant research from a variety of disciplines. We aimed to construct an interdisciplinary and conceptual analysis of both the potential benefits, and likely pitfalls, of machine learning (ML) and artificial intelligence (AI) applications as they relate to student equity interests and objectives . The report further considers what steps universities should make towards ensuring the equitable deployment of advanced data analytics throughout the institution – this is what we are calling a program of *Inclusive Analytics*. Towards this end, the report considers and investigates three central research questions:

- ***RQ1 How can advanced data analytics techniques and processes be deployed in the interest of supporting and advancing student equity interests in Australian higher education?***
- ***RQ2 In what ways does the adoption of advanced data analytics techniques and processes threaten to undermine student equity interests?***
- ***RQ3 What can universities and equity practitioners do to protect and promote equity interests in the age of advanced analytics?***

Understanding the equity policy context in which advanced analytics programs operate within Australian higher education is essential to advancing student equity goals in the age of advanced analytics. In Part 2 of this report, we delineate what we mean by "advanced analytics" and argue that a broad suite of data analytic practices could help to strengthen the existing equity policy framework and provide important mechanisms for its reform. In particular, advanced analytics provides the potential to discover discrimination; identify

individual disadvantage, and relatedly reduce stereotype threat and deficit assumptions; and address emergent and contingent forms of disadvantage in new, dynamic ways. These potential advantages could be leveraged to strengthen and advance existing policies concerning equity in higher education. Such progress will ultimately require deeper consideration, however, of the conceptual and theoretical nature of equity, fairness, and disadvantage as they are understood and defined within Australian higher education.

In Part 3 we describe and evaluate the emerging movement for Fairness, Accountability, and Transparency in Machine Learning (FATML). We further describe the critical limitations of the "fairness movement" within the computer and data sciences and delineate its important contributions to our own project of *Inclusive Analytics*. Ultimately, we conclude that the emerging tools and techniques that have been developed within the FATML movement are necessary, although insufficient in our effort to advance equity interests in the age of advanced analytics.

Part 3 further concludes with a suite of recommendations for the Australian university sector. We argue that institutions need to embed greater data literacy and equity consciousness across their organisations as a matter of priority. Harnessing the potential of analytics to improve student equity requires a comprehensive institutional approach, and a range of sophisticated strategies and practices. In this section we briefly address the need for broader education and professional learning among both academic and professional staff. While not every staff member needs to understand the technical processes of data collection and training, advanced analytics knowledge needs to be distributed across the university, including within areas responsible for many of the related interventions, e.g., student support staff, academic progression staff, and equity practitioners. Further initiatives include greater monitoring and evaluation of the deployment of advanced analytics and of data-related interventions. Such oversight could be partly provided by regulatory committees that might operate similarly to existing ethics committees.

In summary, we argue that universities will need to increase their efforts to oversee and harness advanced analytics, in order to ensure equity, efficiency, ethics, and effectiveness. Collectively, our analysis reveals that advanced analytics is widely used but not well understood, nor is it well scrutinised for equity impacts across the university. This gap is particularly problematic because analytics is no more "neutral" than the data it collects. Further research is required into the equity implications of interventions and initiatives based on advanced analytics, and deeper engagement is also required with the conceptual framework of equity. The existing Australian student equity framework, premised on the six identified equity groups, is increasingly problematic in an age of advanced analytics. Framework issues include a focus on proportional representation to the exclusion of deeper notions of disadvantage, including across the student life cycle; the reification of identities; the challenges of ascribed versus self-defined categories; limitations of a static, point-in-time approach to equity; and the marginalisation of some "invisible" groups and forms of contingent and emergent disadvantage. In addition to the recommendations to institutions outlined within this report, we argue that greater conceptual and policy work is also required at a national level.

Part 2: The promise and peril of advanced analytics applied to equity interests

We began this report by providing a brief overview of advanced analytics within the context of Australian higher education, including the use of artificial intelligence, machine learning and educational data mining. Analytics models now inform vast areas of the university and traverse the whole student lifecycle from the recruitment and admissions of prospective students to success and progression strategies for current students, right through to the production of employability profiles of graduates. The ubiquity of analytics presents new opportunities for the advancement of student equity, though it also presents significant risks and challenges.

In Part 2, we turn our focus to delineating how a program of inclusive analytics may help to advance student equity interests and equity research in the Australian context. We further outline the many potential pitfalls, such as the potential for the stealthy introduction of "bias" or "unfairness" throughout the analytics lifecycle that are persistent dangers within advanced analytics processes: machine learning (ML), educational data mining (EDM), and artificial intelligence (AI).

In sections 2.1 - 2.3 below, we argue that there are at least three related ways in which the use of analytics might improve outcomes of marginalised students, and the broader cause of equity overall. First, analytics can help us to discover discrimination, including within historical processes such as admissions and course guidance. Second, analytics can help us to identify individual disadvantage, beyond reliance on group membership assumptions. In the Australian context, this opportunity potentially enables institutions to move beyond the six identified equity groups to consider dynamic behavioural indicators at an individual level. Proponents of analytics have long touted this capacity for a more sophisticated, individuated understanding of risk and disadvantage. We note, however, that this is a complex and contested area, and requires consideration of protected characteristics, e.g., equity group membership, stereotype risks, and the perils of what has been called a "colour blind" approach. Third, we note that analytics provides an opportunity to assess emergent and contingent forms of disadvantage, such as the impacts of COVID-19.

While the effective use of advanced analytics can demonstrably improve student equity, analytics can also be used to maintain and even exacerbate inequity. In section 2.4 we outline the potential perils which advanced analytic practices pose in relation to equity interests. Our examination runs across the four primary stages of the machine learning lifecycle: 1) data collection and preparation; 2) model building/learning or algorithmic training; 3) model evaluation and verification; 4) model deployment. We consider risks here including historical bias, representation bias, measurement bias and aggregation bias, which can have the effect of reifying stereotypes, misrepresenting individuals, and exacerbating inequity. Further risks include interventions based on predictive analytics, such as tailored communications to "at risk" students, which may be either ineffective or even counter-productive by instigating self-fulfilling prophecies of failure. Threats to privacy are also rising and remain priority areas for the ethical and equitable implementation of analytics throughout the modern university.

2.1. Can advanced analytics revolutionise equity efforts in Australian higher education?

Given the widespread and longstanding recognition of the imperfect nature of our conventional Australian higher education equity metrics, data collection and cohort identification processes, it is little surprise that many have seen a possible solution in "learning analytics", "educational analytics", or advanced data analytics more generally

(Coates et al., 2017; Institute for Social Science Research, 2018; Naylor et al., 2016; Naylor & James, 2016; Zacharias & Brett, 2019).

The *Review of Identified Equity Groups*, a broad consultative research report conducted by the Institute for Social Science Research (2018), advocated for the introduction of "additional and more granular indicators to improve the monitoring of equity in Higher Education" (p. 131). They further recommended that:

> HE participation data could be *supplemented by drawing on universities' administrative data and learning analytics (e.g., indicators of learning outcomes) to pinpoint pockets of disadvantage. This would enable a more nuanced and more precise targeting of people who are disadvantaged in HE. Some stakeholders argued this move would make the current group-based approach redundant.*
>
> (p. 312)

This view that "learning analytics" can and should be deployed to further enrich the equity picture has also been advanced by Naylor and James (2016) and Naylor et al. (2016). Both have argued that universities now capture tremendous amounts of potentially useful student data that extend well beyond simple demographic indicators of group membership. As Naylor and James (2016) argue:

> *It is now possible for under-prepared or educationally disadvantaged students to be identified upon enrollment, and their engagement with subject materials, peers and teaching staff…and learning management systems can be tracked throughout their university studies.*
>
> (p. 10)

Naylor et al. (2016) argue that analytics may be used in a renewed equity program aimed at capturing and identifying patterns of what they call individual "hyper-intersectionality". Leaving the program of group identification behind, they argue for the

> *… idea of using intersecting vectors of quantitative metrics to account for differences in the numerous identity criteria… Using algorithms to connect student admissions data, educational analytics can predict student performance in desirable student outcomes such as grades, persistence, and retention… New typologies predicated on data beyond demographics information will need to be created.*
>
> (p. 270)

What Naylor et al. (2016) propose is a program closely akin to the interests of researchers in the fields of Educational Data Mining (EDM) and Learning Analytics (LA). This project, they argue, will require new data sources that will help to delineate more refined pictures of individual disadvantage. Although recognising that this raises both ethical and privacy concerns, Naylor et al. (2016) argue for the need to supplement institutional data with "data from government systems beyond higher education and likely also more data from the mobile/social technologies that play such a vital role in how people intersect with social and institutional systems" (p. 271).

In a more delimited application, Campbell et al. (2019) have argued that advanced data analytics techniques could be used to address the challenge of "undermatching" among low SES students. In the UK context, undermatching describes a frequently observed situation where disadvantaged, or low SES students, decline to apply to more selective or prestigious courses and universities, even though their A level results would likely win them admission. In an effort to address this challenge, Campbell et al. (2019) argue for the use of automated course recommendation systems that are "[b]ased on the idea of targeted advertising, as used by many popular websites such as Amazon and Facebook". These data-driven

marketing systems would then be used to identify and target low SES students with strong A level results. Then, via targeted electronic advertising, these systems would recommend particular courses "based on their (predicted, or preferably, actual A level (or equivalent)) subjects and grades" and thereby nudge students towards fully "spending", so to speak, their A level achievements (p. 79).

The remainder of Part 2 of this report may be read as an extended appraisal of these calls for renewing Australian higher education equity efforts through the application of what we will refer to as "advanced analytics" techniques. Our aim is to outline both the promise and peril of such a program with an eye towards further refining what an inclusive analytics program should minimally entail. Our guiding research questions in Part 2 are:

- ***RQ1 How can advanced data analytics techniques and processes be deployed in the interest of supporting and advancing student equity interests in Australian higher education?***
- ***RQ2 In what ways does the adoption of advanced data analytics techniques and processes threaten to undermine student equity interests?***

2.2. What do we mean by "advanced analytics"?

In this report, we use the term "advanced analytics" to embody a wide range of data analytics fields of research and professional practice. For the purposes of outlining the shape of our proposed program of *inclusive analytics*, we cast a wide net across the many disciplinary fields and sub-disciplines that make up the broad interrelated fields of data analytics and data science. In other words, and for the purposes of this report, we believe it is unhelpful to get mired in disciplinary boundary disputes between what constitutes the countless fields of study – frequently expressed as initialisms – that now develop and deploy advanced data analytic techniques and practices in the university setting: Learning Analytics (LA), Educational Data Mining (EDM), Educational Analytics (EA), Institutional Analytics (IA), Artificial Intelligence (AI), Machine Learning (ML), and Artificial Intelligence in Education (AIEd) to name just a few.

Even within a single field of advanced data analytics, such as Artificial Intelligence, there exists a surprising level of disagreement concerning the field's disciplinary boundaries and the fundamental definition of what constitutes "artificial intelligence". In a sweeping review of AI definitions from across the research literature, Wang (2019) offered this diagnosis of the discord. "The current field of AI", Wang argues, "is actually a mixture of multiple research fields, each with its own goal, methods, applicable situations, etc., and they are called 'AI' mainly for historical, rather than theoretical, reasons" (p. 28).

Therefore, rather than participate in disciplinary disputes, we use the term "advanced analytics" to refer to the broad application of data analytics techniques and methodologies to the sphere of higher education as a whole. As we demonstrate in Part 3 of this report, our interest in developing a program of *inclusive analytics* necessarily extends well beyond "learning analytics" – or to learning and teaching functions alone – to the application of advanced analytics techniques and processes throughout the whole of the student life cycle, including the business functions of the university.

Given that this report is not intended to be a highly technical exposition of the many allied fields which we have included under the banner of "advanced analytics", we will now provide a basic outline of its two fundamental building blocks: 1) the various forms of data that now power advanced analytic processes, and 2) the typical analytic techniques and methodologies that are applied to these data for the purposes of, for example, prediction, classification, and automated decision making.

Advanced analytics: the data

In our current "information age", the new data sources now collected and available for analysis within, and outside of, the modern university are potentially vast. This vastness is typically referred to as "Big Data", or what Laney (2001) first described as the "three Vs" of "volume'", "velocity" and "variety" of data that is now digitally mediated and made available for analytic and predictive purposes.[1] The difference between old and new analytics approaches, as Bichsel (2012) has indicated, is that the "data used now are becoming much more extensive and automatic, and the processes used to extract and analyze data are becoming repeatable" (p. 7).

What is new in the era of Big Data is the relatively low costs and general ease with which massive amounts of data can be captured, stored, and analysed. Following Fischer et al. (2020), it is useful to think of these data as belonging to three high levels of categorisation: microlevel, mesolevel, and macrolevel. Microlevel data consist of "fine-grained interaction data with seconds between actions that can capture individual data from potentially millions of learners" (Fischer et al., 2020, p. 132). Often described as "digital dust", "digital footprints", or "digital exhaust" (Neef, 2015), microlevel data are typically automatically collected as students interact with university systems, but most commonly within online learning management systems (LMS). These microlevel interactions produce reams of individual digital data traces that are often collected for the purpose of maintaining system security – detecting system errors or hacking efforts – but also for the application of advanced analytics techniques described below.

Mesolevel data, according to Fischer et al. (2020) includes "computerized student writing artifacts systematically collected during writing activities in a variety of learning environments ranging from course assignments to online discussion forum participation, intelligent tutoring systems, and social media interactions" (p. 132). These largely textual mesolevel data are frequently analysed using a sub-class of artificial intelligence called natural language processing (NLP). Mesolevel data can be used to produce insights into students':

> *(a) cognitive processes (e.g., cognitive functioning, knowledge, and skills), (b) social processes (e.g., discourse and collaboration structures), (c) behavioral processes (e.g., learner engagement and disengagement), and (d) affective pro-cesses (e.g., sentiment, motivation).*
>
> (Fischer et al., 2020, p. 140)

Macrolevel big data are then inclusive of the more traditional data that have been traditionally collected at the institutional level for many years and that are used largely for administrative purposes and government reporting. These data are typically digitally stored within an institution's student information system (SIS) and include "student demographic and admission data, course enrolment and grade records, course schedule and course descriptions…" (Fischer et al., 2020, p. 143).

The educational big data typology described by Fischer et al. (2020) provides a useful high-level delineation of the types of data collected by higher education institutions, but it does not capture the full and *global* extent of data that are potentially available for analysis

We may wish to call this "global big data", or data external to the institution that might include data from government such as census data, tax data, or individual student data that is collected by government across multiple higher education providers – identified through the Commonwealth Higher Education Student Support Number (CHESSN). Finally, we must also acknowledge the availability of the truly staggering amounts of data that may be

[1] IBM would later add "veracity" as a fourth characteristic "V" IBM. The four v's of big data. http://www.ibmbigdatahub.com/infographic/four-vs-big-data.

commercially available through social media and data aggregation companies such as Microsoft, Facebook, Google, and Amazon. To give just one example, data relating to the subsequent employment outcomes of a university's alumni may be purchased from LinkedIn (Microsoft) or other big data aggregators.

Taken as a whole, this is what is new about today's big data landscape: volume, velocity, and variety. It is a data collection landscape that, in many instances, may be accurately described as a state of "digital surveillance" or "dataveillance" (Clarke, 1988; Degli Esposti, 2014; Zuboff, 2019). In contrast to the data available to the creators of the original *A Fair Chance for All framework* in 1990 – which was limited to relatively poor macrolevel data – today's data landscape is, by comparison, far richer and more varied. This does not mean, however, that these new forms of data are immune to the dangers of classification, mismeasurement, and opting out that we outlined in Part 1.

Advanced analytics: techniques and methodologies

While there are numerous forms of advanced analytic methodologies and new emerging sub-fields, we will focus on just two broad techniques/methodologies that are most prominent in educational settings: artificial intelligence (AI) and machine learning (ML).

Given the seaming ubiquity of the term "artificial intelligence" (AI) in academia, journalism and popular culture, there remains a surprising level of disagreement among experts in relation to its fundamental definition and current capabilities. Most of the discord can be traced to disagreements about what constitutes "intelligence" in relation to computer generated AI as compared to conventional notions of human intelligence (Emmert-Streib et al., 2020). Fundamentally, however, AI consists of the ability for computer systems to perform automated tasks – such as decision making, image recognition, or forms of automatic adaptation – that are perceived to be reflective of human intelligence. While AI applications are rapidly gaining ground in educational contexts, as evidenced by the growing AIEd field of study (Chaudhry & Kazim, 2021), for the purposes of this report we believe it is most important that we focus on the AI sub-field of machine learning (ML).

The term "machine learning", is frequently used interchangeably with the terms "data science", "data mining" and "analytics" more broadly (Baker, 2021, p. 45). Born out of the need to analyse increasing amounts of big data in a largely automated fashion, machine learning can be defined, according to Murphy (2012) "as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty…" (p. 1). While there are many forms of ML, there are two broad methodologies worth noting, even in a non-technical report such as this one: supervised and unsupervised learning.

In supervised ML, for example, an algorithm may be "trained" on a historical dataset of student records to find patterns, or mappings, that predict or infer a specific known outcome or "target variable" – such as predicting the binary (categorical) outcome of whether a student will pass a given subject (yes/no). The "machine" part of "machine learning" refers to the algorithm's ability to automatically "learn" to produce what is called a "model" "of the relationship between a set of descriptive features and a target feature based on a set of historical examples, or instances" (Kelleher et al., 2015, p. 3). A "target variable", or predicted variable, is very similar to the dependant variable in more traditional inferential statistical processes. Target variables can be categorical, resulting in classification models, or continuous numeric variables such as a student's next semester average mark (0-100). In the case of continuous target variables, numerous linear regression techniques can be applied. We then typically speak broadly of either classification algorithms or continuous/numeric regression algorithms for supervised ML processes.

While perhaps less common than supervised learning, in unsupervised learning the process of automated pattern discovery does not have the benefit of a known outcome or target variable. Instead, the goal of an unsupervised ML process is "to discover 'interesting structure' in the data; this is sometimes called knowledge discovery" (Murphy, 2012, p. 9). This "interesting structure" is typically found through processes of "clustering" or "cluster discovery". As Baker and Siemens (2014) explain:

> *In clustering, the goal is to find data points that naturally group together, splitting the full dataset into a set of clusters. Clustering is particularly useful in cases where the most common categories within the dataset are not known in advance.*
>
> (p. 258)

In this report we focus primarily on the predictive methodologies deployed in supervised ML and the pattern discovery methodologies deployed within unsupervised ML processes.[2] Importantly, ML methodologies that are typically deployed within predictive analytics and data mining processes are significantly different to inferential research methodologies. While traditional statistics are concerned with testing hypothesised causal relationships between variables, or describing the general tendencies within a dataset (Baepler & Murdoch, 2010), machine learning "treats prediction as sacrosanct – it is not important why a given set of covariates are predictive, so long as they are" (Barabas et al., 2018, p. 8). It is often said, therefore, that ML uses the brute-force of massive computational power and "big" data to find *useful* correlations and patterns in datasets – although not necessarily *meaningful* – with which predictions can be made (Johnson, 2018).

We will further delineate the typical analytics and ML lifecycle in section 2.4 below when we describe the potential sources of hidden bias throughout the process. But first, we attempt to describe how big data and advanced analytics methodologies may serve to further equity interests in Australian higher education.

2.3. The promise of big data and advanced analytics to promote equity interests

If we were to, for the moment, disregard our concerns regarding poor equity theorisation and poor data quality – and further postpone our concerns about the introduction of bias throughout the analytic lifecycle itself (2.4 below) – we can identify what we see as a number of important ways in which new "big data" sources and advanced analytics may be used to profitably advance equity interests in Australian higher education.

Analytics for discrimination discovery

The full student life-cycle – from pre-application aspirations through to graduation – is littered with instances where a student's fate is left in the hands of institutional decision making that is often rules-based. These moments of decision making, whether they be human judgements or automated (algorithmic) decisions, present a multitude of moments where discrimination and bias may be introduced even without the decision makers' full knowledge. Such moments are likely to be found in, for example:

1. the advice students receive during year-12 VCE subject selection and university course applications
2. university and course admissions decisions, but particularly direct applications
3. scholarship award decisions
4. decisions relating to recognition of pre-existing credit
5. professional placement, or work integrated learning (WIL), decisions

[2] We could also mention here natural language processing (NLP) and reinforcement learning (RL) as additional commonly used forms of machine learning, but given the space limitations we have not covered them in detail for this report.

6. study abroad application decisions
7. internal course transfer application decisions
8. academic progression and academic integrity decisions.

It is generally thought that data mining and machine learning practices can only add or compound discriminatory outcomes in high-stakes decisions, and this is certainly possible as we will outline in section 2.4 below. But researchers have increasingly shown that advanced analytic methods can be deployed for the purpose of pursuing highly refined projects of "discrimination discovery" within historical decision records and contemporary, or real-time, decision processes (Bonchi et al., 2017; Hajian et al., 2016; Kahneman et al., 2021; Qureshi et al., 2020; Ruggieri et al., 2010; Zhang et al., 2016). Where robust historical data records of these equity-sensitive decisions exist, advanced data analytics methods, particularly data mining, can be used to help identify patterns and instances of discrimination. Sophisticated analytic methodologies have been developed for the detection of both direct and indirect discrimination. As (Zhang et al., 2016) explain:

> *Direct discrimination occurs when individuals receive less favorable treatment explicitly based on the protected attributes. An example would be rejecting a qualified female applicant in applying to university just because of her gender. Indirect discrimination refers to the situation where the treatment is based on apparently neutral non-protected attributes but still results in unjustified distinctions against individuals from the protected group. A well-known example of indirect discrimination is redlining, where the residential Zip code of the individual is used for making decisions such as granting a loan*.
>
> (p. 1)

Importantly, the interest in "indirect discrimination" has resulted in the creation of discrimination discovery methods that focus, not only on singular protected groups or attributes, but also on intersecting or multiple categories of equity/disadvantage whose compounding effect may be otherwise obscured (Ruggieri et al., 2010).
We further note that with the rapid expansion of Robotic Process Automation (RPA) throughout the business functions of the university (Razak et al., 2021; Turcu & Turcu, 2019), widescale and automated discrimination discovery and monitoring is becoming increasingly important. As van der Aalst et al. (2018) have described, "RPA is an umbrella term for tools that operate on the user interface of other computer systems in the way a human would do. RPA aims to replace people by automation…" (p. 269). RPA and other business process automation techniques are now becoming commonplace in universities and are frequently overtaking the human role in decision making processes we described at the beginning of this sub-section. RPA can, for example, be used for automating the review of student applications for admission on a rules-based, or algorithmic, automation system. Again, as automated decision making of this kind becomes more commonplace in universities, the more pressing the need for discrimination discovery and monitoring processes.

We believe there is considerable scope for Australian equity researchers to submit historical and contemporary datasets to "discrimination discovery" techniques drawn from the related fields of advanced analytics. The hope being that where discrimination in decision making can be clearly defined, it can also be discovered for the purposes of remediation and ongoing monitoring.

There are at least two significant hurdles that must be overcome if this is to be achieved. First, analytic discrimination discovery techniques still require a working definition of what constitutes unacceptable discrimination, or as the fields of analytics now commonly describe it, "fairness" or "unfairness" (Barocas et al., 2019). This is an important hurdle and one that cannot itself be submitted to automation. It requires engagement with the often difficult

ethical and political negotiations that should take place among domain experts. Secondly, a program of this kind would, of course, be limited by the many types of data distortion that we have described earlier in this report. It is also likely to be most useful in instances where long standing equity group membership has been consistently identified in the historical data record. In this way, discrimination discovery is likely to be most effective when applied to conventional equity categories and groups for whom we have longstanding, although imperfect, data.

Analytics for individual "disadvantage"

Perhaps the most frequently expressed hope for big data and advanced analytics, among Australian equity researchers and practitioners, has been the expectation that these tools might allow the sector to move beyond, at least in part, the conventional paradigm of equity as membership within an under-represented group (Institute for Social Science Research, 2018; Naylor et al., 2016; Naylor & James, 2016; Zacharias & Brett, 2019). There are many motivations behind such a desire, but the problems of deficit thinking, and the limitations of government recognised equity categories, are frequently cited. The hope expressed has been that new data sources and new analytic techniques might allow for the emergence of an equity paradigm that is more individualised and more authentically intersectional (Institute for Social Science Research, 2018; Naylor et al., 2016; Willems, 2010) – that is one that can register the compounding effect of multiple disadvantages, or competing power structures, that are experienced by marginalised individuals (Collins & Bilge, 2020).

There are, perhaps, two primary ways in which advanced analytics might help to better individualise – assuming, for the moment, that this is a desirable pursuit – our view of disadvantage within higher education. First, predictive analytics methods can help to shift equity work from a process of group identification to one of early identification of academically struggling and disengaged students. Second, and much like the process of discrimination discovery described above, data mining techniques such as "knowledge discovery" may find success in identifying new, more individualised, forms of disadvantage through the "mining" of new and emerging data sources. We have already mentioned knowledge discovery methods above and will here focus on predictive analytics methods.

We believe there is an important sense in which these hopes are, to a limited degree, already being realised through the deployment of predictive analytics in Australian universities. To accept this claim would require a remapping of our conventional understanding of disadvantage *as equity group membership*, to a view of disadvantage as, for example: 1) under-preparation for higher education study; 2) the increased risk or demonstration of academic under-performance or attrition; or 3) disengagement from the institution and from learning activities. Taking an international view, these are the typical applications towards which predictive and/or learning analytics processes have been applied in relation to student achievement and progress (Anagnostopoulos et al., 2020; Dawson et al., 2017; Herodotou et al., 2020; Ranjeeth et al., 2020; Seidel & Kutieleh, 2017; Wolff et al., 2014). Moreover, it appears that the equity researchers we quoted in section 2.1 (above) have articulated an argument for a remapping of equity of this kind. Again, we take Naylor and James (2016) as representative when they write that, through the leveraging of advanced analytics techniques it "is now possible for under-prepared or educationally disadvantaged students to be identified upon enrollment" (p. 10), not by equity group membership, but through the many forms of data institutions now collect throughout an individual student's studies. What equity researchers such as Naylor and James (2016) and others (Institute for Social Science Research, 2018; Naylor et al., 2016) have described is something akin to a program of predictive analytics.

Predictive analytics is one of the most prominent methodologies to be found in the fields of Learning Analytics (LA) and Educational Data Mining (EDM) (Baker & Siemens, 2014) and deserves a far more detailed treatment than we can provide in this report. Sometimes

referred to as Predictive Learning Analytics (PLA) (Herodotou et al., 2020), predictive analytic processes are now deployed within many Australian universities, for a variety of purposes, and may be conducted by internal analytics units, by small third-party consultancies, or by multinational commercial data analytics providers. Given the relative ubiquity of predictive analytics and related practices throughout the sector, its footprint within the Australian research literature remains relatively small by comparison (Dawson et al., 2017; Gasevic, Shum, et al., 2016; Pardo et al., 2016; Seidel & Kutieleh, 2017).
The paucity of published research on predictive analytics programs, but particularly those that explore the effectiveness of subsequent student support interventions in Australian higher education, may speak to several factors, including: 1) claims to commercial privacy and intellectual property on the part of third-party analytics providers; 2) difficulties in securing human ethics committee approvals; and 3) the fact that evaluating such programs, and their interventions – particularly in the absence of randomised control trials which can be difficult to pass ethics committee scrutiny – is particularly difficult (Dawson et al., 2017). At least one Australian research study has, however, described an interesting partnership between internal institutional researchers and a third-party analytics consultancy (Seidel & Kutieleh, 2017). In still other cases, the results of predictive analytics programs – more specifically, the results of the student support interventions they inform – simply go unpublished due to the fact that they failed to deliver successful outcomes through their chosen intervention strategy (Clow et al., 2016).

This raises an important distinction which we believe equity researchers and practitioners should bear in mind in relation to the deployment of predictive analytics techniques. There is a distinction to be drawn between predictive analytics outputs – for instance, their "accuracy", which itself can be a notoriously slippery term – and the interventions that institutions decide to pursue based on these predictions. While a discussion of the technological "neutrality thesis" is beyond the scope of this report (Johnson, 2018; Miller, 2021), we note that an accurate set of predictions may not result in the "inclusive" or effective use of those predictions.

Protected attributes, stereotypes, and "colour blindness"

The need for more refined understandings of a student's individual "risk" profile – determining risk of academic underperformance (failure) and/or attrition that go beyond simply registering equity group membership – have been demonstrated in numerous studies. For example, many studies which deploy a multivariate analysis have now shown that conventional equity group membership is not, in itself, a strong factor or attribute that puts students at increased risk for non-completion of their course (Harvey & Luckman, 2014; Li & Carroll, 2017; Walker-Gibbs et al., 2019). Li and Carroll (2017), for example, found that after controlling for weighted average mark (WAM), students from equity cohorts were less likely to leave their course in the following year than were students who were not identified as belonging to an equity group. Moreover, studies that do not have the benefit of using a student's individual academic achievement results, such as WAM, have found that factors unrelated to equity status – such as institutional factors, course study load, and admissions factors – reveal themselves as much stronger predictors of course non-completion (DESE, 2020; HESP, 2018; Norton et al., 2018). Taken together, these studies suggest that

> *While equity group membership alone is generally not a risk factor for attrition or non-completion, members of equity groups are often more likely to be at risk. In other words, risk factors such as unit failure and low academic achievement, low ATAR, and part-time study mediate the relationship between equity group membership and attrition and non-completion.*
>
> (Stephenson et al., 2021, p. 9)

These studies, which deploy a more traditional inferential statistical methodology, raise important questions concerning the use of equity and demographic data features as "inputs" within machine learning, or predictive analytics processes.

Researchers in the fields of learning analytics (LA) and educational data mining (EDM) – which again employ significantly different methodologies as compared to conventional inferential and descriptive statistics – have asked whether the inclusion of "protected attributes", or immutable equity group characteristics, as model input features causes greater discrimination by unfairly increasing the risk profile for under-represented groups. If more traditional multivariate inferential statistical studies frequently find that these equity categories are not predictive of, for example attrition, should we still include equity features (or variables) in machine learning or predictive analytic processes?

In one of the few Australian research studies on the application of machine learning for predicting university student attrition, Seidel and Kutieleh (2017) found that of the 50 variables/features they tested for predicting attrition, none of the many demographic or equity variables trialled were of sufficient predictive power to be included in the final models. Behavioural data, or data relating to observed academic outcomes and (dis)engagement, were far more useful to the predictive modelling than were static or immutable demographic characteristics/variables. As the authors explain:

> *The inclusion and relative strength of behavioural data in the predictive models meant that students were identified and contacted based on their patterns of behaviour and performance rather than their profile, thus avoiding potential stigmatisation through stereotyping. These results allowed students within equity groups, known to have elevated levels of attrition, to be considered equally to all other students who tend to have lower levels of attrition across the population but may have elevated predicted attrition risks based on their personal behavioural characteristics*.
>
> (Seidel & Kutieleh, 2017, p. 213)

In a UK study, Foster and Siddle (2020) came to a similar conclusion, but for slightly different reasons. They tested the efficacy of using a "learning analytics" system to produce "no-engagement" alerts – based on (dis)engagement data collected from various institutional electronic systems – to accurately identify students at-risk of academic underperformance. The researchers found that while students from low socioeconomic backgrounds had higher rates of academic underperformance, targeting interventions around observed disengagement behaviours was more efficient at identifying truly at-risk students than were demographic attributes. While the authors provide no details on the analytic processes involved, they conclude that this approach "creates a more neutral framework for any" interventions and "avoids the potential risk of stigmatising a student's background…" (Foster & Siddle, 2020, p. 852).

In a similar study from a large U.S. research university, Yu et al. (2021) tested predictive analytics models aimed at predicting student attrition after one year. Their research sought to determine if including four "protected attributes" – gender, first-generation college student, underrepresented minority, and high financial need – would improve prediction accuracy and further improve "algorithmic fairness". We return to the issue of algorithmic fairness below, but here we may indicate that the researchers were concerned to test whether the predictions were "fair", or equally predictive and error prone, for all demographic groups. They found that "including protected attributes does not impact the overall prediction performance and it only marginally improves the algorithmic fairness of the predictions" (Yu et al., 2021, p. 91). Perhaps counterintuitively, the researchers do not conclude that "protected attributes" should be dropped from ML predictive process, thereby making our predictive process unaware, or "blind" as they say, of demographic characteristics – in other

words, simply hiding or dropping these sensitive variables from the dataset. Instead, they argue that the ML processes that deploy "socio-demographic-aware models" can better

> *…capture structural inequalities in society that disproportionately expose members of minoritized groups to more adverse conditions. In addition, the deliberate exclusion of protected attributes from dropout prediction models can be construed as subscribing to a "colorblind" ideology, which has been criticized as a racist approach that serves to maintain the status quo*.
>
> (Yu et al., 2021, p. 98)

Žliobaitė and Custers (2016) have made a similar, although more forceful, argument that sensitive or protected attributes *must* be collected and used in predictive modelling processes to adequately detect and correct discriminatory outcomes. In sum, they argue for an analytics process that is fully "aware" of protected, demographic, or immutable characteristics. In a practical sense, this means that "collecting sensitive personal data is needed in order to guarantee fairness of algorithms, and law making needs to find sensible ways to allow using such data in the modelling process" (Žliobaitė & Custers, 2016, p. 199).

We can see then that predictive analytics may aide us in moving from conventional notions of equity as group membership to more behavioural indicators of under-preparedness or academic underperformance and attrition. The actual methodology of predictive analytics, as we review further below, can smuggle discrimination back into the ML process at many points throughout the lifecycle. It is for this reason that Australian equity researchers and practitioners may wish to retain conventional and other less-conventional equity group data within educational datasets. This also raises interesting questions regarding the cost-benefit calculous of both being "seen" and "unseen" within datasets of consequence. It may well be in an individual's or group's interest to remain "seen" within datasets, if only to better empower efforts towards discrimination discovery. We return to this issue briefly in section 2.4 below.

We can summarise the "promise" of predictive analytics for equity interests as follows. First, it may help us to move beyond dated notions that equity group membership necessarily indicates under-preparedness for university or increased risk of failure and attrition. In this way, predictive analytics, when followed by very well-considered interventions, may go some distance towards challenging deficit notions of equity. This requires, however, a redefinition of our conventional understanding of equity as static group characteristics, and instead focus our attention on under-prepared and under-performing students more broadly. Second, by moving beyond static or immutable student characteristics to real-time diagnostics based on observed behavioural and performance data, we can again go some distance towards avoiding the "self-fulfilling prophecy" that deficit models all too frequently stoke. But this too is a matter for further research and debate as we indicate below. Third, predictive analytics, when done well, allows us to move beyond notions of "risk" or "underperformance" as static phenomenon that are inherent to certain demographic groups of students. From the view of an ongoing predictive analytics program, a student may move from "high-risk" to "low-risk" of academic underperformance – or vice versa – within a short span of time, no matter what their "static" characteristics might be.

Ethical considerations for analytics-driven interventions

This is not to suggest, of course, that there are no "perils" lurking in relation to programs of predictive analytics. The ethical and data privacy dangers have been well covered in the research literature and continue to be salient and hotly debated (Corrin et al., 2019; Dawson et al., 2019; Selwyn & Gašević, 2020; Slade & Prinsloo, 2013). There are also important ethical debates to be had concerning the intervention strategy, or operational goals, that predictive analytic outputs are used for. The now infamous "drown the bunnies" case at Mount St. Mary's University in the United States has become a stark warning of the dangers

of predictive analytics to equity and inclusion interests the world over (Jaschik, 2016). In this case, it is alleged that the university's president insisted on using predictive analytics to identify students who were most at risk of attrition and then, based on these predictions, counsel these students to leave the university in the first few weeks of the term – what was described as "drowning the bunnies". The motivation was to preserve the institution's national rankings which included metrics on attrition. Encouragingly, the more commonly adopted suite of interventions aimed at supporting students who have been identified as "at-risk", by advanced analytics processes, has been targeted academic advising and pastoral or peer support (Dawson et al., 2017; Seidel & Kutieleh, 2017), or to "support instructors in the provision of frequent and effective formative and personalized feedback" (Pardo et al., 2016).

Serious questions also remain regarding the effectiveness of interventions that are triggered based on predictive analytics. It may be that interventions where a student is made aware that an analytic system has deemed them to be "at-risk" is all it takes to initiate the "self-fulfilling prophecy" of deficit thinking within the student. For example, Jones (2019) has described an instance where professional academic advisers rejected the use of predictive analytics due, at least in-part, to the dangers of creating self-fulfilling prophecies in the students they advise (p.450). Moreover, fundamental legal, ethical, and moral principles regarding the transparent use of student data would dictate that the basis for the intervention – the data and analytics processes that led to, for example, the "at-risk" designation – should be declared and made clear to the student. As a core, and widely accepted, ethical principle within the analytics literature, the principle of transparency dictates that institutions should share information about how data are collected, the techniques used to analyse those data, and provide clear indications of how, when, and why the data are being used for a particular purpose (Askinadze & Conrad, 2018; Corrin et al., 2019, pp. 10-11).

More closely registering the compounding effect of intersectional or multiple equity group belonging, using various data analytics techniques, has been advocated by several Australian higher education equity researchers (Institute for Social Science Research, 2018; Naylor et al., 2016; Willems, 2010). Only two of these studies have proposed minimally detailed models for how intersectional equity may be analysed (Institute for Social Science Research, 2018; Willems, 2010), but both proposals use only the existing conventional equity categories, although Willems (2010) proposes some amendments. Naylor et al. (2016), on the other hand, call for the creation of new data sources that go beyond the conventional equity categories.

An analytics program aimed at the elucidation of intersectional disadvantage would certainly encounter all the concerns regarding self-fulfilling prophecy that we mentioned above. We might ask to what use would a refined picture of individual disadvantage be put to? If the goal is to create individual "learner profiles" that would describe intersectional risk profiles for individual students, we must ask who these profiles would be available to and for what purpose? Would they be shared and made available to academic teaching staff in the interest of tailoring feedback? Would they be made available to the student themselves? When thought of in this light, there is little comfort to be found in a move to more refined and individualised portraits of student disadvantage or "risk". The "Pygmalion", or Rosenthal effect clearly looms large in such considerations (Boser et al., 2014).

While we believe that these are important concerns that require further research and debate, it should also be noted that "predictive analytics" and "learning analytics" frequently include real-time, or close to real-time, observations of student behaviours, academic results, or indicators of (dis)engagement. In other words, it is misleading to think of predictive models as being purely *predictive*. In cases such as student "ghosting behaviours" and subsequent failure (Stephenson et al., 2021), early interventions based on real-time observed disengagement, not just predicted disengagement, could potentially save a student from academic failure as well as from accumulating sizable student debts. In some cases, the

student may even be unaware that they are enrolled in a unit if not for the real-time observations that "predictive" or "learning analytics" systems can recognise and flag. In fact, digital traces of a student's disengagement from an institution's administrative and learning systems are now an Australian government-recommended means for determining whether a student is in fact a "genuine student" (Stephenson et al., 2021, pp. 6-7).

Aside from ethical and privacy concerns, there are also many ways in which the related techniques of predictive analytics, machine learning and educational data mining may themselves introduce disadvantage, or repeat historical bias, throughout the analytics lifecycle. We further delineate this critical challenge in section 2.4 below.

Analytics for emergent and conditional disadvantage

The growth of "big data" and advanced data analytics has also raised the potential for Australian higher education to better recognise and respond to what we will call emergent and conditional forms of disadvantage. Forms of *emergent* disadvantage might include conditions where an identifiable people-group – perhaps geographically identifiable – haven't necessarily suffered historical disadvantage but are experiencing emergent or even transient forms of disadvantage. The effects of climate change, including drought and bushfire, for example, have already begun to create emergent forms of disadvantage throughout Australia. Moreover, these emergent environmental, economic and psychological pressures can often disproportionately impact on people from conventional Australian equity groups including Indigenous people (Green et al., 2009), regional and remote farming communities (Berry et al., 2011), and the socioeconomically disadvantaged (Fritze et al., 2008). Still other forms of emergent disadvantage might be found in localised industry collapse such as the closing of Victorian car manufacturing (Beer, 2018), or in the more generalised, but still unequal, economic hardship created by the COVID-19 pandemic (O'Sullivan et al., 2020).

Like the logic of intersectionality, advanced analytics may help us to discover what we might call *conditional* forms of disadvantage. For instance, where an individual or groups of students may not fall into any of the conventional equity categories, but still experience other unrecognised forms of disadvantage. Research powered by new data forms, and advanced analytic techniques, is beginning to reveal instances of otherwise hidden geographical disadvantage. For example, a student may live in a relatively well-off socioeconomic area, but it may be geographically situated a long way from the university and public transport may be exceptionally arduous and time consuming for the student. There are an increasing number of studies that are exploring these kinds of intersections of conventional advantage/disadvantage with geographical particularities including, for example, qualitative and quantitative studies of commuter students in the greater London area (Liz Thomas Associates Ltd, 2019; London Higher, 2019).

2.4. The perils of big data and advanced analytics for equity interests

We have already described many of the "perils" involved in the use or deployment of advanced analytics outputs as it relates to student equity interventions and programs in Australian higher education. In this section we seek to summarise the "perils" that big data and advanced analytics methodologies present as an inherent part of the analytic or machine learning (ML) lifecycle itself. We will summarise the ways in which discrimination may be introduced, reproduced, and even multiplied within a seemingly benign analytics lifecycle. Recognising that data analytics and data science are broad fields composed of countless sub-fields, our analysis here is limited largely to machine learning (ML) which frequently underwrites processes within the related fields of learning analytics (LA), educational data mining (EDM) and artificial intelligence (AI).

Sources of bias and harm throughout the analytics lifecycle

While it has become commonplace to acknowledge that advanced analytics are "biased" because the "data are biased" – in the well-known pattern of "garbage in, garbage out" – this is only part of the story. It is, however, a critical part of the story and given that data are the bedrock upon which advanced analytics operate, it is also the beginning of the story. It is also quite common to hear within popular public discourse that the "perils" of advanced analytics can be traced back to what is commonly called "algorithmic bias" (Obermeyer et al., 2021). Once again, while there is truth to this claim, "algorithmic bias" – or bias that is introduced by the algorithm alone – is just one among many sources of bias or discrimination within the full ML lifecycle.

The machine learning (ML) lifecycle is typically broken into four fundamental stages: 1) data collection and preparation; 2) model building/learning or algorithmic training; 3) model evaluation and verification; and 4) model deployment. We will now briefly summarise just some of the more common "perils" at each stage and attempt to provide examples that are relevant to Australian higher education equity interests. In what follows we will be largely guided by the "seven sources of harm" or bias in ML processes that have been summarised by Suresh and Guttag (2020) and Mehrabi et al. (2021).

Data collection and preparation stage

Most descriptions of the analytics lifecycle begin with an assumption that the data already exist and simply need to be collected, managed, or "cleaned". Following Suresh and Guttag (2020) we will then refer to what is typically taken as the first stage of the lifecycle as the *data collection* stage. Having developed an understanding of the problem to be addressed through the application of advanced analytics, data analysts will first set about collecting the necessary, or simply the available, data whether it be of micro, meso, macro or global levels of provenance (Fischer et al., 2020). The data collection stage presents multiple opportunities for the introduction of bias, discrimination, or harm. The extent or "type" of harms that are introduced at this and subsequent stages depends largely on the application to which the analytic or predictive "model" is put. Data analytics scholars typically refer to four basic types of bias that can be introduced by the data themselves – historical, representational, measurement, and aggregation bias – but many more have been identified throughout the research literature (Barocas et al., 2019; Barocas & Selbst, 2016; Mehrabi et al., 2021; Ntoutsi et al., 2020; Olteanu et al., 2019; Suresh & Guttag, 2020).

Historical bias refers to harms that may be introduced through the faithful – that is to say "accurate" – capturing of true inequalities that currently exist, or used to exist, and cause an analytic model to reproduce that historical, and possibly harmful, status quo. As Suresh and Guttag (2020) explain, "[s]uch a system, even if it reflects the world accurately, can still inflict harm on a population" (p. 5). An imagined example might be a university's automated course or unit recommendation system, which are becoming increasingly common (Christensen & Eyring, 2011; Crozier, 2020). Bias may be introduced if the ML-powered system is, as is typical, "trained" on an historical dataset of student preferences. Here, if the data scientists/analysts are not exceedingly cautious, historical patterns and dated societal conventions may be "learned" and repeated by the automated system. Female students may be more likely to see suggestions relating to, for example, allied health courses, while male students may see more suggestions relating to STEM fields. In a general sense, a model or automated system that is built upon historical examples of course selection, rather than a user's self-declared interests, runs the risk of reifying an historical state of affairs that may embody discrimination or dated social conventions.

Representation bias can be expressed in a number of ways, but fundamentally has to do with analytic processes built upon datasets that fail to adequately represent a full population and therefore fail to generalise (Suresh & Guttag, 2020, p. 5). In Part 1 of this report, we

described numerous key student groups that may go “unseen” in Australian higher education datasets for a number of reasons, including small group size or various forms of “opting out”. These groups would be at risk of representation bias within analytic processes along with groups that are at risk, due to their low numbers, of being statistically “insignificant” and “unseen” when compared to the full student population. Importantly, local institutional conditions regarding representation may vary greatly when compared to national or state statistics. For instance, while the representation of women in non-traditional study areas has improved, an individual institution may have very little representation of women in their datasets. The same could be true for men in study areas that are considered non-traditional. For this reason, ML projects at the institutional or program-level must take these local issues of representation into careful consideration.

There are also several important types of *measurement bias* which typically consists of instances where a “feature or label is a *proxy* (a concrete measurement) chosen to approximate some *construct* (an idea or concept) that is not directly encoded or observable” (Suresh & Guttag, 2020, p. 6). In the Australian equity Framework, the postcode proxy for measuring low SES status, along with regional and remote status, has been among the most frequently criticised for potential measurement bias. Suresh and Guttag (2020, p. 6) have also described an apt example of proxy measures that oversimplify an otherwise complex construct, such as the concept of a “successful student”. Universities, researchers, and government use a wide range of singular proxy measures for student “success” – e.g., average mark, institutional retention, sector-wide retention (or “adjusted retention), timely completion, employment outcomes, etc. – yet none of these on its own can fully capture complex notions of what constitutes a “successful student”. A form of measurement bias then arises when just one of these proxies for success is selected as a “target” variable that is now going to be “optimised” within, for example, a ML-driven predictive process aimed at selecting for admission only those students who are most likely to meet this one (mis)measure of “success”.

Aggregation bias is a family of biases or statistical fallacies – including the ecological fallacy and Simpson’s Paradox – that arise “when false conclusions are drawn about individuals from observing the entire population” (Mehrabi et al., 2021, p. 5). Within the equity framework, it is particularly low SES metrics that risk committing the ecological fallacy if “inferences are made about individuals based purely on characteristics of the area in which they live…” (Wise & Mathews, 2011, p. 1). Equally so, if a dataset were to, as is often the case, aggregate all the many potential forms of student disability to a single label of “disability” (yes/no), this too would risk creating aggregation bias.

Model learning or algorithmic training stage

We should recall here that “machine learning” refers in part to a computational or algorithmic process where data are used to “train” an algorithm(s) which automatically “learns” patterns and relationships within the data – either “supervised” or “unsupervised”. This is the stage of the ML lifecycle that is largely automated. We say “largely automated” since this stage still requires numerous important decisions to be made on the part of the (human) data scientist/analyst. These decisions can have a profound effect on the ML model’s tendency towards biased or unfair outcomes. We suggest then that there are at least two broad ways in which we can think of “algorithmic bias”.

First, algorithmic bias can be introduced through any of the many decisions and trade-offs made by the (human) data scientist, that concern the algorithmic training portion of the ML lifecycle:

> *The algorithmic design choices, such as use of certain optimization functions, regularizations, choices in applying regression models on the data as a whole or considering subgroups, and the general use of statistically biased estimators in algorithms, can all contribute to biased algorithmic decisions that can bias the outcome of the algorithms*.
>
> (Mehrabi et al., 2021, p. 7)

Many of these decisions relating to algorithmic tuning, or the selection of algorithms and hyperparameters, are made in conjunction with model evaluation and verification. An analyst will adjust the algorithm to optimise certain types of accuracy or error that are discovered during the model evaluation stage described further below.

Second, algorithms that are fully trained and then, as is frequently the case, deployed for the purposes of automated decision making, can continue to produce biased outcomes and even develop new or emergent biases based on user interactions. These are the types of bias that are frequently observed in web-based or app-based algorithms that suggest, present, or rank content for users to see (Mehrabi et al., 2021, p. 7). Search engines are perhaps the most common examples of deployed algorithms that "learn" from and respond to user interactions in this way. In an educational context, examples would include automated recommendation engines that suggest courses, subjects/units, or even careers to student users (Casuat et al., 2020; Esteban et al., 2020). In their survey of the many types of bias created by recommendation systems, Chen et al. (2020) argue that such systems are particularly prone to discriminating against, or failing to recognise, the preferences of under-represented groups.

Model evaluation and verification stage

Once an algorithm has been "trained" on a set of historical or "training data" it can then be evaluated or verified for its performance or prediction accuracy through a variety of methods – but typically with what is called "hold-out" data. In a real-world scenario, stage 2 and 3 are often carried out in conjunction. An analyst will typically evaluate the error/accuracy of multiple models to optimise the algorithm towards a particular desired outcome or balance between accuracy, error, speed, and even the interpretability of the overall model (Gilpin et al., 2018).

Applying these choices to a real-world example, we can think of a data analyst who is interested in predicting students who are likely to be "ghost students" or non-participating enrolments (NPEs) (Stephenson et al., 2021), i.e. those who remain enrolled in a subject/unit, but fully disengage and fail to submit any assessments. This is a very difficult prediction problem given that effective interventions need to happen early in the semester, but many students will not engage until later in the semester (Stephenson et al., 2018).

Depending on the intended intervention, the analyst will adjust the hyperparameters (tune the dials) on the algorithm to manipulate the trade-offs between accuracy and various types of error. In a binary classification problem, for instance, the analyst would understand that there is always going to be error in the predictions of perplexing human behaviours such as "ghosting". This means that we must understand and anticipate the likely extent of false-positives and false-negatives for each model we consider. Adjusting for these types of error during model evaluation is often a trade-off, and a design choice, that must be made in nearly all real-world ML applications. In sum, these choices of evaluation metrics, trade-offs in error/accuracy, will necessarily benefit some data profiles – individuals, groups, demographics – over others. In Part 3 of this report, we briefly discuss the tremendous efforts computer scientists are now making to help create tools and processes that will aid in this tricky decision-making process, which is broadly known as the "fairness in ML" movement (Barocas et al., 2019).

Model evaluation is a critical part of nearly all applications of machine learning (ML) and artificial intelligence (AI) and, as we discuss below, is an increasingly critical skillset for all manner of university decision makers. We further suggest that equity practitioners who are encountering the increasing number of ML and AI applications within their own universities would do well to learn more about model evaluation, model accuracy/error, and concerns for "fairness" evaluation within these applications (Barocas et al., 2019; Kelleher et al., 2015, Ch. 8).

Model deployment stage

As Suresh and Guttag (2020) explain, bias can also be introduced during the model deployment stage "when there is a mismatch between the problem a model is intended to solve and the way in which it is actually used" (p. 8). In the context of student equity, we can again imagine models aimed at predicting a student's "at-risk" status. In a case such as this, we must be very clear about how the programmer/analyst understood the term "at-risk"– "at-risk of what"? A model built to predict a student's risk of failing a single introductory business subject/unit, for example, should not be repurposed to predict institutional attrition. The use case and subsequent intervention should match the intended purpose for which the model was created. Another form of bias that can be introduced during model deployment is what Selbst et al. (2019) call the "framing trap". Here bias is introduced by the end users of an ML or AI process in a way that was not intended or anticipated by the makers of the model. The framing trap then consists of a failure to consider the full "sociotechnical frame" within which a model will be deployed.

Part 3: Towards an *Inclusive Analytics*: protecting equity interests in the age of advanced analytics

In Part 1 of this report, we focused on the Australian higher education equity policy setting within which the theorisation, quantification, and categorisation of equity, underrepresentation, and disadvantage takes place. In Part 2 we then described how advanced analytics may be leveraged to support and strengthen equity policy, and how they may also work to undermine equity interests, often via stealth. In this section we ask how a culture of inclusive analytics can be created and sustained in Australian universities. We outline and evaluate the recent explosion of interest in the field of Fairness, Accountability and Transparency in Machine Learning (FATML), which represents a broad effort on the part of researchers in the fields of advanced analytics to address "fairness" concerns. The report then concludes with a discussion of recommendations for several ways in which universities and equity advocates can work to protect equity interests in the age of advanced analytics.

In Part 3 our guiding research question is:

- ***RQ3 What can universities and equity practitioners do to protect and promote equity interests in the age of advanced analytics?***

3.1. The Fairness, Accountability, and Transparency in Machine Learning (FATML) movement

As we have described throughout this report, it is now widely recognised that the ML/AI production "pipeline" involves a series of critical choices by which discriminatory outcomes may be introduced at any stage, and often without the awareness of the data analysts/programmers: data collection and preparation stage, algorithmic training stage, model evaluation stage, and final deployment (Suresh & Guttag, 2019). Moreover, data scientists and researchers within the many broad fields of advanced data analytics are themselves undergoing a significant reckoning with their field's own complicity in perpetuating social discrimination and disadvantage through ML/AI processes (Barocas et al., 2019; Hajian et al., 2016; Ntoutsi et al., 2020; Selbst et al., 2019). This has led to an explosion of research in the emerging, and related fields, of Fairness, Accountability and Transparency in Machine Learning (FATML) and "responsible" or "explainable" artificial intelligence (XAI) (Arrieta et al., 2020).

It is important to note that "fairness" has emerged as a catch-all term within this broad body of FATML research and appears to incorporate notions of equality, discrimination and social justice, but according to some observers, has not yet fully grappled with issues of "equity" as distinct from "equality" (Mehrabi et al., 2021, p. 25). Work in FATML/XAI has been highly technical in nature and focuses largely on the production of mathematical and statistical definitions and diagnostics, for detecting what – as we have already seen – is frequently described as "data bias" or "algorithmic discrimination" within a given ML/AI process. These proposed methodologies are frequently being developed in the hope that tests for ML/AI "fairness" could be automated within the analytics systems and processes themselves.

The field of FATML has now produced well over 20 technical definitions of "fairness" and a host of additional methodologies for detecting latent discrimination (Bellamy et al., 2018; Donovan et al., 2018; Makhlouf et al., 2021; Narayanan, 2018; Verma & Rubin, 2018). Some FATML "fairness" taxonomies distinguish between more familiar notions of fairness including distributive and procedural fairness (Marcinkowski et al., 2020), as well as group and individual fairness (Binns, 2020). What these "fairness" definitions within the FATML

literature hold in common is a tendency to be highly technical and mathematical in nature. While a full exposition of these many technical definitions of "fairness" is beyond the scope of this report, we believe equity researchers and practitioners interested in refining their conceptions of equity/fairness are likely to find important, although insufficient, insights from the field of FATML. Moreover, we argue that it is important for equity advocates and practitioners to understand the basics of how "fairness" methods in ML work and what their shortcomings might be. Following Verma and Rubin (2018), we can briefly summarise the emerging definitions of "fairness" in the FATML research literature as consisting of three main types: statistical measures; similarity based measures; and causal reasoning. Taken as a whole, these definitions have been developed to be deployed within what are now called "fairness-aware" algorithmic processes.

The family of "statistical measures" of fairness, which Verma and Rubin (2018) describe, rely largely on measuring different classification accuracy/error tradeoffs within a typical statistical confusion matrix – including rates of true positives, true negatives, false positives, and false negatives. Additional model performance metrics can then be derived or calculated from confusion matrices, such as precision, recall, and F_1 scores (Murphy, 2012). Within this class of fairness measures we will find, for example, definitions of fairness that require a model's predictions and/or probabilities to be equally reliable – by any of several possible metrics – for both sensitive (protected) and non-sensitive classes/groups/attributes. Green and Hu (2018) helpfully refer to this class of "fairness" measures as "fairness as satisfaction of balanced statistical metrics" (p.2).

Take for instance a predictive model which makes university admissions decisions or recommendations based on the applicant's likelihood of "success", measured as, for example, on-time course completion. If the model's predictions are equally accurate – by some agreed statistical measure, such as F_1 scores – for both male and female applicants, the model might be declared "fair" regarding this particular statistical measure in relation to gender. As Verma and Rubin (2018, p. 5) explain, statistical measures of fairness "largely ignore all attributes of the classified subject except for the sensitive attribute". In our example the protected attribute is gender, but it could just as well be low SES, disability, or Indigenous status. There now exists an abundance of methods, code, and tools available for detecting and measuring the growing number of statistical measures for ML/AI fairness (Bellamy et al., 2018; Edizel et al., 2020; Hajian et al., 2016; Veale & Binns, 2017)

"Similarity-based measures", on the other hand, also employ non-sensitive attributes within the fairness formula. In this family of fairness measures, the model is made explicitly aware of all similarities and differences between individuals – sometimes called "fairness through awareness" – and judges a model to be fair when similar individuals are treated similarly (Ntoutsi et al., 2020, p. 5). Finally, the "causal reasoning" family of fairness measures (Kilbertus et al., 2017) are based largely on the causal graph research and theories of Judea Pearl (2019). This approach, sometimes called "counterfactual fairness", utilises causal graphs to build "fair" ML algorithms by measuring the effects of sensitive attributes and further ensures that only tolerable levels of discrimination are caused by those attributes (Verma & Rubin, 2018, p. 6).

"Accountability", the "A" in the FATML field of research and practice, has received less attention but embodies a broad recognition of ethical responsibility within the allied fields of advanced analytics (Shah, 2018). FATML researchers are interested in increasing and affirming accountability – among data analysts, institutions, and all those who utilise analytic outputs – concerning the real-world sociotechnical (including ethical) outcomes of advanced analytics workstreams. Accountability can be taken as a broad statement against the "technological neutrality thesis" which maintains that the products of data science and analytics are not, in themselves, "good" or "bad", thereby relieving the field of moral responsibility. "Accountability" is then a principle of assumed responsibility on the part of technologists, data scientists and the institutions and individuals that make use of these tools

(Piano, 2020). In sum, the broad fields of advanced data analytics are moving away from simplistic notions which consider analytic outputs to be value neutral, while claiming that it is their use alone that introduces moral responsibility and the need for accountability. Accountability, however, requires a sufficient level of transparency (Veale et al., 2018).

"Transparency", the "T" in FATML, refers largely to the understanding that certification of "fairness" requires transparency, and its corollary of "explainability" (Barredo Arrieta et al., 2020), in the advanced analytics processes of automated decision making. It is now widely understood that ML/AI algorithmic processes can be so deeply complex that they are, in fact, uninterpretable, and therefore unexplainable, by their human developers. This is the well-known "black-box" problem in ML/AI process (Aggarwal et al., 2019). This opacity of process and decision-making is particularly acute in "deep learning" neural network algorithms that incorporate "hidden layers" within the algorithmic process (Castelvecchi, 2016). The transparency in ML/AI movement has taken up this concern through the many efforts amongst analytics researchers to provide human-interpretable explanations for complex automated decisions, rankings, and predictions (Saha et al., 2020; Zarsky, 2013). The hope is to not only build public trust in the fairness of automated decision-making systems, but to also meet legal requirements for transparency in data use and decision making (Citron & Pasquale, 2014). Requirements for transparency, like fairness, can run up against concerns for data privacy. For instance, can we guarantee an automated system is fair for all protected classes/attributes if those attributes are not represented for all individuals within a given dataset, or does their inclusion risk breaches of privacy? (Askinadze & Conrad, 2018; Veale & Binns, 2017). Still others have proposed methods for including sensitive attributes but "hiding" them via encryption to potentially sidestep the issue (Kilbertus et al., 2018).

3.2. The growing critique of FATML: necessary but insufficient

Since rapidly gaining ground in the various fields of advanced analytics over the past decade, the theories, tools, and assumptions of FATML researchers have been subjected to a growing wave of critique. While some have argued over the technical definitions and statistical tests for "fairness" that FATML researchers have proposed, the more cutting critique has been socio-political in flavour, and seeks to question – if not fundamentally undermine – the theoretical and philosophical assumptions within the fair-ML movement (Corbett-Davies & Goel, 2018; Green, 2020; Green & Hu, 2018; Greene et al., 2019; Hoffmann, 2019; Selbst et al., 2019).

For example, Selbst et al. (2019) have described five "traps" that technical efforts towards producing fair-ML systems, like those described in the FATML and XAI literatures, can fall into. They are worth repeating here for their utility, insight, and brevity:

1. ***The Framing Trap*** *- Failure to model the entire system over which a social criterion, such as fairness, will be enforced*
2. ***The Portability Trap*** *- Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context*
3. ***The Formalism Trap*** *– Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms*
4. ***The Ripple Effect Trap*** *- Failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system*
5. ***The Solutionism Trap*** *- Failure to recognize the possibility that the best solution to a problem may not involve technology* (Selbst et al., 2019)

Selbst et al. (2019) propose that a consideration of the five "traps" could become a critical part of the ML/AI lifecycle, where stakeholders question the project's fundamentals via each potential trap.

The "framing trap" would encourage stakeholders to consider the full sociotechnical environment (the full frame) in which the "solution" may be deployed. We might ask: will "fairness" be preserved throughout the human-technical (heterogeneous) system, from design through to deployment? The "portability" trap is particularly well known in learning analytics (LA) research circles, where the hope for "one size fits all" solutions have been well and truly laid to rest (Gasevic, Dawson, et al., 2016). While this trap is well understood among the LA community, it is, perhaps, less well understood among other business areas throughout the university where third-party services are frequently adopted based on the belief that "it worked over at university X". The "formalism" trap – which assumes fairness can be reduced to a mathematical formalism – is among the most critical, in our estimation, and is one that is particularly salient to equity interests in Australian higher education. As we are increasingly confronted with questions concerning the inclusive use of advanced analytics in Australian universities, we will most certainly need to further refine and, in some cases, renegotiate our concepts of fairness, equity, and disadvantage. The "formalism" trap reminds us that pure maths will not deliver us from the need for difficult ethical and political negotiation.

The "ripple effect" trap is also critically important as the introduction of "a new technology may appear to alter an organization's dynamics but may in fact aid in reifying a pre-existing group's claim to power, while downplaying or downgrading other groups' authority" (Selbst et al., 2019, p. 65). Finally, the "solutionism" trap reminds us that "pre-packaged algorithms stamped with 'fairness'" (p. 66), and other automated low or no-labour "solutions", may not provide the optimal solution. We should recognise that the "Silicon Valley narrative" is very strong in higher education (Weller, 2015), along with its attendant "technological solutionism" (Williamson & Hogan, 2020). Moreover, as Greene et al. (2019) have found in their critical survey of business "values statements" and manifestos, relating to the ethical use of ML/AI applications, there is a strong sense that ethics are best addressed and solved "through technical and design expertise" (p. 2122) – to be certain, the spirit of technological solutionism runs strong. But again, critics maintain that appeals to mathematical "solutions" to the fairness question, "without broader analysis of social and moral context" are all but destined to fail (Green & Hu, 2018, p. 3).

The sum of these critiques and traps are, for Selbst et al. (2019), all misguided abstractions of the very concept of fairness:

> *We contend that by abstracting away the social context in which these systems will be deployed, fair-ML researchers miss the broader context, including information necessary to create fairer outcomes, or even to understand fairness as a concept… Fairness and justice are properties of social and legal systems like employment and criminal justice, not properties of the technical tools within. To treat fairness and justice as terms that have meaningful application to technology separate from a social context is therefore to make a category error, or as we posit here, an abstraction error*.
>
> (Selbst et al., 2019, p. 59)

For other critics of the fair-ML movement, it is not just a problem of abstraction, but of reductionism and reification:

> *While the field's practice of labelling these particular metrics [for measuring ML fairness] is benign, reifying fairness as constituted by satisfaction of the statistical constraints is mistaken. Reductive commitments to statistical parities of various types limit the realm of justice and fairness to one of merely adjudicating comparative*

> *claims of treatment and outcomes across groups, which represent only one relevant criterion for assuring fairness*.
>
> (Green & Hu, 2018, p. 2)

Even if we were to accept the limited utility of reductive mathematical measures of fairness, we would be confronted with the problem of incommensurability between the many formal definitions of fairness. As Kleinberg et al. (2016) have observed, many "fairness" problems – where, for example, base rates differ for different groups – will necessitate "inherent trade-offs" between otherwise incommensurable fairness conditions or definitions. In other words, many formal statistical definitions of fairness are contradictory and fundamentally fail to be instructive in the absence of applied domain expertise, or where ethical and political negotiations of "the good" are sidestepped (Green & Hu, 2018; Greene et al., 2019). Moreover, notions of "fairness" may not be fully reconcilable with notions of "equity" or "equality", which, in the Australian context, sometimes allow for "asymmetrical protection" – not just equality of treatment – for some groups, such as people with disabilities (Gaze & Smith, 2016, p. 99).

Thus far we have provided only an introduction to the criticisms of the FATML movement. There are, of course, other criticisms which deserve mention, but we will offer just two more from Green and Hu (2018). First, they argue that ML's "reliance on data and metrics can distort deliberative processes" (p. 3). While some have argued that ML, and algorithmic decision-making more generally, are well suited to the problem of removing "noise" and "bias" from human decision-making (Kahneman et al., 2021), Green and Hu (2018) counter that "algorithms have the potential to distort the values underlying laws and policies that, in principle, society has collectively determined to be fair, and to do so without proper democratic input" (p. 3). Secondly, fair-ML "narrows judgments about fairness and entrenches historical discrimination" (p.3) by assuming that unfair decisions are made by individuals rather than by unjust "laws and institutions that systematically benefit one group over another" (p. 4).

Some of the more thoughtful work in the FATML research literature has acknowledged these and other limitations. In one of the most important works within the field, Barocas et al. (2019) recognise that there are critical limitations to the practice of "fairness" testing – or fairness aware algorithms – and insist that "[d]espite these limitations and difficulties, empirically testing fairness is vital" (p. 122). Importantly, however, they add that

> *This does not mean that we can select and apply a fairness test based on convenience. Far from it: we need moral reasoning and domain-specific considerations to determine which test(s) are appropriate, how to apply them, determine whether the findings indicate wrongful discrimination, and whether an intervention is called for…Conversely, if a system passes a fairness test, we should not interpret it as a certificate the system is fair*.
>
> (Barocas et al., 2019, p. 121)

We believe it is critically important that equity advocates, institutional data analysts, and senior decision-makers in Australian universities work to familiarise themselves with the fairness, accountability, and transparency in ML movement and come to understand its critical limitations.

Moreover, we stress that senior university decision-makers, particularly those charged with technology and software service procurement, should be aware of the growing commercialisation of analytics "fairness" solutions. It is increasingly the case that commercial ML/AI platforms and consultancies are selling products and services that claim to address "bias" or "fairness" within a variety of advanced analytics processes. These products and services present an appealing means of outsourcing the ethical obligations associated

with a university's technology adoption to third parties, who may claim to offer mathematical proofs and guarantees of "fairness". We argue that senior administrators and procurement officers in Australia's universities should approach third-party "solutionism" with great skepticism. Local domain expertise – that is, professional human expertise – is fundamental to preserving fair outcomes throughout the highly complex sociotechnical and political environments of the modern university.

The more transparent of these products and services will, like the makers of Microsoft's "Fairlearn" toolkit, remain forthright concerning the tremendous sociotechnical challenge of achieving "fairness" in machine learning applications. They admit, rightly, that purely technical solutions, like those produced by FATML researchers, are not sufficient to guarantee that an ML/AI system is fully debiased (Bird et al., 2020, p. 1). Fairlearn's designers are very clear that their tool is not a "cure-all" for the issues of "fairness" in ML/AI. For instance, they openly state that

> *There are many aspects of fairness that are beyond the scope of Fairlearn. For example, Fairlearn cannot mitigate stereotyping harms, denigration harms, or over- or underrepresentation harms (except indirectly, when these harms arise as a result of allocation harms or quality-of-service harms). Also, Fairlearn does not focus on broader societal aspects of fairness, such as justice or due process*.
>
> (Bird et al., 2020, p. 3)

In sum, we argue that Australian universities, and equity advocates within these institutions, should advocate for a fully democratic, and decidedly human and "in-house", process of ML/AI adoption. A process that, like Barocas et al. (2019) strongly advise, incorporates fair-ML's "fairness testing" as only one necessary, but ultimately insufficient, part of the equity oversight process. To conclude this report, we now turn to a description of recommendations for the Australian university sector that we believe will better secure a culture of inclusive analytics into the future. That is, a culture of analytics practice that represents and benefits all students according to explicit and agreed notions of equity and fairness.

3.3. Recommendations: building a culture of *inclusive analytics* in Australian universities

In this report we have described both the promise and the peril of advanced data analytics for equity interests in Australia's universities. We have also included something of the scope of deployment of these advanced analytic tools, services, and methodologies throughout the student lifecycle. Our recommendations here are focused on universities themselves, though we acknowledge the need for broader reform at government level as well. We believe that at the governmental level, the most impactful change will come about via data protection laws that span well beyond the university sector alone. For example, the Australian Competition and Consumer Commission ACCC (2021) has recently made a number of important recommendations to government concerning Google's dominance of data aggregation and advertising services in Australia. We therefore conclude this report with a list of recommendations for universities and the university sector.

1. **That universities develop data analytics policies and procedures that protect equity interests throughout the full student lifecycle and across all business areas.**

While many universities have adopted policy or ethics statements on "learning analytics" in a narrow sense, institutions should also develop data analytics policies and procedures that protect equity interests and goals throughout the full student lifecycle, and include all business areas of the university: e.g., marketing, admissions, student services, etc. This would necessarily require that explicit equity protections be built into institutional data

governance policy and procedure. Fundamentally the “peril” posed to equity interests in universities is not limited to teaching and learning activities but extends out beyond the university through pre-enrolment marketing and post-graduation employability projects and digital tracking of alumni. A university’s declared commitments to equity and diversity should extend to these business functions as well.

2. **That universities broaden distribution of *analytic expertise*, particularly within the DVC (Academic) divisions.**

It is increasingly the case that student-facing (or focused) divisions within the university are at a distinct disadvantage when they lack personnel who possess a deep understanding of advanced analytics techniques, processes, and their potential pitfalls. As we have argued throughout this report, domain experts in non-ICT related fields – librarians, student advisers, recruitment officers, teachers/lecturers, etc. – are indispensable towards informing the ethical and equitable deployment of advanced data analytics throughout universities. Moreover, there is a particularly urgent need for domain experts in non-technical fields, to advance their understanding and engagement with advanced analytic techniques that are being deployed within their traditional domains of professional expertise.

3. **That universities broaden distribution of *equity and ethics expertise*, particularly including within data analytics (institutional research and performance), Information Services, and ICT divisions of the university.**

There is a clear need for universities to broaden engagement among institutional data experts, IS/ICT professionals, and information services specialist with these concerns regarding the equitable, ethical, and even legal, use of advanced data analytics. Just as areas outside of IS/ICT departments must deepen their expertise in advanced analytics, so too must ICT and information services divisions work to deepen their expertise regarding the possible unintended consequences of analytic projects on equity groups and equity interests more broadly. As domain experts in data analytics, ICT and information services, equity practitioners should work to partner with colleagues in these fields.

4. **That universities increase professional education of staff, including academics, engaged with analytics projects at each stage of the procurement, development and deployment process.**

In a systematic review of research relating to data literacies and the training of university teaching staff (faculty), Raffaghelli and Stewart (2020) found that where training was taking place it was largely concerned with mastery of accepted, and unproblematised, technical practices. This indicates that many university teachers are likely to be unprepared for work environments that increasingly include advanced analytic tools, services, and methodologies. We recommend that Graduate Certificates in Teaching and Learning begin to focus on advanced analytic literacies as part of the curriculum. Similarly, all university staff should have professional development offered to them as part of analytic project development and deployment.

5. **That universities establish in-house regulatory structures and professional expertise to ensure equity and fairness are protected through the deployment of advanced data analytics, e.g., standing committees to oversee analytics, similar to ethics committees.**

We argue for the creation of specialised institutional review boards that are of a similar composition to human research ethics committees, but that are of a decidedly interdisciplinary character. The boards could resemble, and be informed by, the Responsible Innovation Organisations (RIO) that have been called for in industry (The University of Melbourne, 2019). In a university context, a RIO would require broad representation from in-

house experts, members from equity cohorts, computer science/analytics experts, social scientists, teaching and learning experts, ethicists, legal professionals, and student representatives. In this way, universities may be able to better guarantee that the adoption of advance data analytics processes will undergo a full engagement with the politics of technology and inclusion. Greater oversight and regulation should not be an excuse for stifling institutional innovation. A program of inclusive analytics should instead provide the appropriate safeguards within which innovation can be leveraged to benefit all students in a spirit of inclusivity.

6. **That universities ensure that analytics-informed interventions are tailored, based on behavioural factors, and designed to reduce self-fulfilling prophecies based on immutable characteristics.**

As described throughout the report, there is a clear danger that advanced analytics, but particularly predictive analytics, may work to repeat historical discrimination or instigate self-fulfilling prophecies in "targeted" or "at-risk" cohorts. We argue that institutions must take care to implement appropriate safeguards in both the development of predictive modelling and in their deployment. The principle of "first do no harm" to the student is apt, but we should also ensure that our interventions anticipate inevitable error in the modelling process and ensure that any risks associated with the intervention are taken at the institution's expense and not that of students.

7. **That universities regularly monitor and evaluate the analytics project lifecycle for impact on equity and "fairness" interests.**

The lifecycle of advanced analytics projects should include equity monitoring, testing and reasonable audits of the process throughout the lifecycle. Ideally, these would be conducted by in-house teams with ongoing responsibility for ensuring ethical and equitable deployment of advanced analytics systems.

8. **That universities work towards benchmarking/collective agendas, potentially involving Universities Australia (UA) leadership.**

Given the speed of change and the ubiquity of analytics, sectoral and multi-institutional approaches would be helpful to ensuring ethical and equitable practices. Peak bodies such as Universities Australia could encourage communities of practice and prioritise discussion of analytics within conferences, symposia, and regular scheduled meetings, e.g. among DVCs (Academic). Institutional benchmarking of analytics practices, both within and beyond the academic portfolios, could be facilitated, and exemplars highlighted from within and beyond Australia.

9. **That universities conduct and facilitate further interdisciplinary research into the intersection of equity in higher education and advanced data analytics as an urgent priority.**

Research on the intersection between advanced data analytics and higher education equity efforts has been limited. Given that we believe advanced analytics is one of the more pressing issues for equity researchers and practitioners to engage with, we urge researchers to focus more of their efforts on this interdisciplinary field. Particularly, we see a need for:

- More refined theoretical/philosophical notions of equity and their integration with formal definitions of "fairness" now current in the field of fair-ML.
- More refined understanding of the intended and unintended consequences of various forms of "opting-out" of data collection for individuals and groups.
- Broader research into a program of equitable data collection and curation practices.

- Stronger evaluation and monitoring of analytics-driven interventions and publication of evaluation findings in peer-refereed journals and public fora.

References

Afuro, D., & Mutanga, B. (2021). Combating digital academic dishonesty: A scoping review of approaches. *International Journal of Engineering and Advanced Technology*, *9*, 82-88. https://doi.org/10.35940/ijeat.F1268.089620

Aggarwal, A., Lohia, P., Nagar, S., Dey, K., & Saha, D. (2019). *Black box fairness testing of machine learning models,* Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Tallinn, Estonia. https://doi.org/10.1145/3338906.3338937

Allen, L. (2021). Promoting representation through data: The case for more comprehensive ethnicity data in Australia. *Law in Context*, *37*(2), 1-8.

Anagnostopoulos, T., Kytagias, C., Xanthopoulos, T., Georgakopoulos, I., Salmon, I., & Psaromiligkos, Y. (2020). Intelligent predictive analytics for identifying students at risk of failure in Moodle courses. In V. Kumar & C. Troussas, *Intelligent Tutoring Systems* Cham.

Anderson, D. (1983). Access to privilege: Patterns of participation in Australian post-secondary education. Australian National University Press.

Andrews, J. (2018). Blaks and stats in Aboriginal Victoria: census resistance and participation. *Australian Aboriginal Studies*(1), 43-56.

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Askinadze, A., & Conrad, S. (2018, 27-29 June 2018). Respecting data privacy in educational data mining: An approach to the transparent handling of student data and dealing with the resulting missing value problem. 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE),

Australian Competition and Consumer Commission ACCC. (2021). *Digital advertising services inquiry: Final report*. Canberra Retrieved from https://www.accc.gov.au/system/files/Digital%20advertising%20services%20inquiry%20-%20final%20report.pdf

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching & Learning*, *4*(2), 152-162.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 253-272). Cambridge University Press.

Baker, R. S. (2021). Artificial intelligence in education: Bringing it all together. In *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots* (pp. 43-51). OECD Publishing. https://doi.org/doi:https://doi.org/10.1787/589b283f-en

Barabas, C., Virza, M., Dinakar, K., Ito, J., & Zittrain, J. (2018). Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. Conference on Fairness, Accountability and Transparency,

Barnard, B. (Mar 22, 2020). Redesigning college admission: COVID-19, access and equity. *Forbes*. https://www.forbes.com/sites/brennanbarnard/2020/03/22/redesigning-college-admission-covid-19-access-and-equity/#108c671d2ee6

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning*. fairmlbook.org. http://www.fairmlbook.org

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Legal Review*, *104*, 671-732.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115. https://doi.org/https://doi.org/10.1016/j.inffus.2019.12.012

Beer, A. (2018). The closure of the Australian car manufacturing industry: Redundancy, policy and community impacts. *Australian Geographer*, *49*(3), 419-438.

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J. T., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2018). AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv:1810.01943v1* http://arxiv.org/abs/1810.01943

Berry, H. L., Hogan, A., Owen, J., Rickwood, D., & Fragar, L. (2011). Climate change and farmers' mental health: risks and responses. *Asia Pacific Journal of Public Health*, *23*(2_suppl), 119S-132S.

Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress, and recommendations (Research Report),* EDUCAUSE Center for Applied Research. Louisville, CO. http://net.educause.edu/ir/library/pdf/ers1207/ers1207.pdf

Binns, R. (2020). *On the apparent conflict between individual and group fairness,* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain. https://doi.org/10.1145/3351095.3372864

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., & Walker, K. (2020). Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft, Tech. Rep. MSR-TR-2020-32*.

Bonchi, F., Hajian, S., Mishra, B., & Ramazzotti, D. (2017). Exposing the probabilistic causal structure of discrimination. *International Journal of Data Science and Analytics*, *3*(1), 1-21.

Boser, U., Wilhelm, M., & Hanna, R. (2014). The power of the Pygmalion Effect: Teachers' expectations strongly predict college completion. *Center for American Progress*.

Brennan, B. (2019). *Opting Out of Digital Media*. Routledge.

Brett, M. (2018). *Equity performance and accountability*. Perth, WA: National Centre for Student Equity in Higher Education (NCSEHE), Curtin University, and La Trobe University

Campbell, S., Macmillan, L., & Wyness, G. (2019). *Mismatch in higher education: prevalence, drivers and outcomes,* UCL Institute of Education and Nuffield Foundation. London. https://www.nuffieldfoundation.org/project/undermatch-in-higher-education-prevalence-drivers-and-outcomes

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, *538*(7623), 20-23.

Casuat, C. D., Isira, A. S. M., Festijo, E. D., Alon, A. S., Mindoro, J. N., & Susa, J. A. B. (2020). A development of fuzzy logic expert-based recommender system for improving

students' employability. 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC),

Chaudhry, M. A., & Kazim, E. (2021). Artificial intelligence in education (AIEd): A high-level academic and industry note 2021. *AI and Ethics*, 1-9. https://doi.org/10.1007/s43681-021-00074-z

Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2020). Bias and debias in recommender system: A survey and future directions. *arXiv preprint arXiv:2010.03240*.

Christensen, C. M., & Eyring, H. J. (2011). *The Innovative University*. Jossey-Bass.

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Wash. L. Rev.*, *89*, 1.

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, *31*(5), 498-512.

Clow, D., Ferguson, R., Macfadyen, L., Prinsloo, P., & Slade, S. (2016). *LAK failathon,* Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh, United Kingdom. https://doi.org/10.1145/2883851.2883918

Coates, H., Kelly, P., Naylor, R., & Borden, V. (2017). *Innovative Approaches for Enhancing the 21st Century Student Experience,* Australain Government, Department of Education and Training. Canberra. https://ltr.edu.au/resources/SP14_4618_Coates_Report_2017_0.pdf

Collins, P. H., & Bilge, S. (2020). *Intersectionality* (2nd ed.). Wiley. https://books.google.com.au/books?id=fyrfDwAAQBAJ

Commonwealth of Australia. (1990). A Fair Chance for All: National and Institutional Planning for Equity in Higher Education. Canberra: Australian Government Publishing Service

Corbett-Davies, S., & Goel, S. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*, 1-25.

Corrin, L., Kennedy, G., French, S., Shum, S. B., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). *The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper*. https://melbournecshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics

Crozier, R. (2020, May 6 2020). La Trobe Uni builds AI 'explorer' to help arts students visualise a career. *iTnews*. https://www.itnews.com.au/news/la-trobe-uni-builds-ai-explorer-to-help-arts-students-visualise-a-career-547781

Dawson, D., Schleiger, E., Horton, J., McGlaughlin, J., Robinson, C., Quezada, G., Scowcroft, J., & Hajkowicz, S. (2019). *Artificial Intelligence: Australia's Ethics Framework,* Australia. Data61 CSIRO.

Dawson, S., Jovanovic, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. Proceedings of the seventh international learning analytics & knowledge conference,

Degli Esposti, S. (2014). When big data meets dataveillance: The hidden side of analytics. *Surveillance & society*, *12*(2), 209-225.

DESE. (2020). *Factors affecting higher education completions,* Department of Education, Skills and Employment (DESE). Canberra. https://docs.education.gov.au/system/files/doc/other/transitions_predicting_completion_rates.pdf

Donovan, J. M., Caplan, R., Matthews, J. N., & Hanson, L. (2018). *Algorithmic accountability: a primer,* Data & Society. https://datasociety.net/library/algorithmic-accountability-a-primer/

Drew, N., Wilks, J., Wilson, K., & Kennedy, G. (2016). Standing up to be counted: Data quality challenges in Aboriginal and Torres Strait Islander higher education statistics. *Australian Aboriginal Studies*(2), 104-120.

Edizel, B., Bonchi, F., Hajian, S., Panisson, A., & Tassa, T. (2020). FaiRecSys: Mitigating algorithmic bias in recommender systems. *International Journal of Data Science and Analytics*, *9*(2), 197-213. https://doi.org/10.1007/s41060-019-00181-5

Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). Artificial intelligence: A clarification of misconceptions, myths and desired status. *Front Artif Intell*, *3*, 524339-524339. https://doi.org/10.3389/frai.2020.524339

Engina, G., Aksoyerb, B., Avdagicb, M., Bozanlıb, D., Hanayb, U., Madenb, D., & Ertek, G. (2014). Rule-based expert systems for supporting university students. *Procedia Computer Science*, *31*, 22-31. https://doi.org/doi:10.1016/j.procs.2014.05.241

Esteban, A., Zafra, A., & Romero, C. (2020). Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization. *Knowledge-Based Systems*, *194*, 1-14. https://doi.org/https://doi.org/10.1016/j.knosys.2019.105385

Eubanks, V. (2018). Automating Inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130-160. https://doi.org/10.3102/0091732x20903304

Foster, E., & Siddle, R. (2020). The effectiveness of learning analytics for identifying at-risk students in higher education. *Assessment & Evaluation in Higher Education*, *45*(6), 842-854. https://doi.org/10.1080/02602938.2019.1682118

Fritze, J. G., Blashki, G. A., Burke, S., & Wiseman, J. (2008). Hope, despair and transformation: climate change and the promotion of mental health and wellbeing. *International journal of mental health systems*, *2*(1), 1-10.

Gagliardi, J. S., Parnell, A., & Carpenter-Hubin, J. (Eds.). (2018). The Analytics Revolution in Higher Education: Big Data, Organizational Learning, and Student Success. Stylus Publishing.

Gale, T., & Tranter, D. (2011). Social justice in Australian higher education policy: An historical and conceptual account of student participation. *Critical Studies in Education*, *52*(1), 29-46.

Gasevic, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gasevic, D., Shum, S. B., Nelson, K., Alexander, S., Lockyer, L., Kennedy, G., Rogers, T., Corrin, L., Fisher, J., Colvin, C., & Wade, A. (2016). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement,* Canberra, ACT. Australian Government Office for Learning and Teaching. http://www.olt.gov.au/project-student-retention-and-learning-analytics-snapshot-currentaustralian-practices-and-framework

Gaze, B., & Smith, B. (2016). Equality and Discrimination Law in Australia: An Introduction. Cambridge University Press.

Ge, S., & Chen, X. (2020). The application of deep learning in automated essay evaluation. In E. Popescu, T. Hao, T.-C. Hsu, H. Xie, M. Temperini, & W. Chen (Eds.), *Emerging Technologies for Education. SETE 2019. Lecture Notes in Computer Science* (Vol. 11984, pp. 310–318). Springer. https://doi.org/https://doi.org/10.1007/978-3-030-38778-5_34

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA),

Graesser, A. C., Conley, M. W., & Olney, A. (2018). Intelligent Tutoring Systems. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences*. Routledge.

Green, B. (2020). Data science as political action: grounding data science in a politics of justice. *Available at SSRN 3658431*.

Green, B., & Hu, L. (2018). The myth in the methodology: Towards a recontextualization of fairness in machine learning. The Debates workshop at the 35 th International Conference on Machine Learning, Stockholm, Sweden.

Green, D., King, U., & Morrison, J. (2009). Disproportionate burdens: the multidimensional impacts of climate change on the health of Indigenous Australians. *Medical Journal of Australia*, *190*(1), 4.

Greene, D., Hoffman, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. Proceedings of the 52nd Hawaii International Conference on System Sciences,

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining,

Harvey, A., Andrewartha, L., & MaNamara, P. (2015). A forgotten cohort? Including people from out-of-home care in Australian higher education policy. *Australian Journal of Education*, *59*(2), 182-195.

Harvey, A., Andrewartha, L., Sharp, M., Wyatt-Smith, M., Jones, S., Shore, S., & Simons, M. (2020). *From the military to the academy: Supporting younger military veterans in Australian higher education,* Department of Veterans' Affairs Supporting Younger Veterans Grants Program. https://www.semanticscholar.org/paper/From-the-military-to-the-academy%3A-supporting-in-Harvey-Andrewartha/792318f085ce8ca54d3c77dcbec9eabcf1c46e77

Harvey, A., Burnheim, C., & Brett, M. (2016). Towards a fairer chance for all: Revising the Australian student equity framework. In A. Harvey, C. Burnheim, & M. Brett (Eds.), *Student Equity in Australian Higher Education: Twenty-five years of a Fair Chance for All* (pp. 3-20). Springer.

Harvey, A., Cakitaki, B., & Brett, M. (2018). *Principles for equity in higher education performance funding,* Centre for Higher Education Equity and Diversity Research, La Trobe University. Melbourne. (Report for the National Centre for Student Equity in Higher Education Research., Issue. https://www.ncsehe.edu.au/publications/principles-for-equity-in-higher-education-performance-funding/

Harvey, A., & Leask, B. (2020). At the policy margins: People from refugee backgrounds in Australian higher education. In *Refugees and Higher Education* (pp. 193-205). Brill. https://doi.org/https://doi.org/10.1163/9789004435841_014

Harvey, A., & Luckman, M. (2014). Beyond demographics: Predicting student attrition within the Bachelor of Arts degree. *The International Journal of the First Year in Higher Education*, *5*(1). https://doi.org/10.5204/intjfyhe.v5i1.187

Herodotou, C., Naydenova, G., Boroowa, A., Gilmour, A., & Rienties, B. (2020). How can predictive learning analytics and motivational interventions increase student retention and enhance administrative support in distance education? *Journal of Learning Analytics*, *7*(2), 72-83. https://doi.org/10.18608/jla.2020.72.4

HESP. (2018). *Final Report - Improving retention, completion and success in higher education,* Deaprtment of Education and Training, Australian Government. Canberra. https://docs.education.gov.au/documents/higher-education-standards-panel-final-report-improving-retention-completion-and-success

Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, *22*(7), 900-915. https://doi.org/10.1080/1369118X.2019.1573912

IBM. *The four v's of big data*. http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Institute for Social Science Research. (2018). *Review of the Identified Equity Groups,* The University of Queensland. Unpublished report prepared for the Australian Department of Education and Training.

Jaramillo-Morillo, D., Ruipérez-Valiente, J., Sarasty, M. F., & Ramírez-Gonzalez, G. (2020). Identifying and characterizing students suspected of academic dishonesty in SPOCs for credit through learning analytics. *International Journal of Educational Technology in Higher Education*, *17*(1), 45. https://doi.org/10.1186/s41239-020-00221-2

Jaschik, S. (2016). Are at-risk students bunnies to be drowned? *Inside Higher Ed*. https://www.insidehighered.com/news/2016/01/20/furor-mount-st-marys-over-presidents-alleged-plan-cull-students

Johnson, J. A. (2018). Toward Information Justice: Technology, Politics, and Policy for Data in Higher Education Administration. Springer.

Jones, K. M. (2019). Advising the whole student: eAdvising analytics and the contextual suppression of advisor values. *Education and Information Technologies*, *24*(1), 437-458.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgement*. William Collins.

Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2015). Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. The MIT Press.

Kilbertus, N., Gascón, A., Kusner, M., Veale, M., Gummadi, K., & Weller, A. (2018). Blind justice: Fairness with encrypted sensitive attributes. *Proceedings of the 35th International Conference on Machine Learning*, 2630-2639.

Kilbertus, N., Rojas-Carulla, M., Parascandolo, G., Hardt, M., Janzing, D., & Schölkopf, B. (2017). Avoiding discrimination through causal reasoning. *arXiv:1706.02744*.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv:1609.05807*.

Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous Data Sovereignty: Toward an Agenda*. ANU press.

Kukutai, T., & Walter, M. (2021). Indigenous data sovereignty: Implications for data journalism. In L. Bounegru & J. Gray (Eds.), *Towards a Critical Data Practice* (pp. 65). Amsterdam University Press.

Kurniadi, D., Abdurachman, E., Warnars, H. L. H. S., & Suparta, W. (2018). The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm. *IOP Conference Series: Materials Science and Engineering*, *434*, 012039. https://doi.org/10.1088/1757-899x/434/1/012039

La Nauze, J. (1940). Some aspects of educational opportunity in South Australia. *Australian Educational Studies*, 101-124.

Lacity, M., Scheepers, R., Willcocks, L., & Craig, A. (2017). *Reimagining the University at Deakin: An IBM Watson Automation Journey,* London School of Economics and Political Science. (The Outsourcing Unit Working Research Paper Series, Issue 17/04). http://www.umsl.edu/~lacitym/LSEOUWP1704.pdf

Laney, D. (2001). *3D Data management: Controlling data volume, velocity, and variety,* META Group Inc. Application Delivery Strategies. Stamford CT. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Li, I. W., & Carroll, D. R. (2017). Factors Influencing University Student Satisfaction, Dropout and Academic Performance: An Australian Higher Education Equity Perspective, National Centre for Student Equity in Higher Education (NSEHE), Curtin University. Perth. https://www.ncsehe.edu.au/publications/success-failure-higher-education-uneven-playing-fields/

Lin, J., Pu, H., Li, Y., & Lian, J. (2018). Intelligent recommendation system for course selection in smart education. *Procedia Computer Science*, *129*, 449-453.

Liz Thomas Associates Ltd. (2019). *Commuter students in London: Pilot project, Qualitative perceptions of students about commuting and studying in London*. https://www.londonhigher.ac.uk/wp-content/uploads/2019/08/CSIL_Perceptions_Aug2019.pdf

London Higher. (2019). *Commuter students in London: Results of a pilot project on factors affecting continuation,* London Higher. London, UK. https://www.londonhigher.ac.uk/wp-content/uploads/2019/08/CSIL_Continuation_Aug2019.pdf

MacMillan, D., & Anderson, N. (October 14, 2019). Student tracking, secret scores: How college admissions offices rank prospects before they apply. *The Washington Post*. https://www.washingtonpost.com/business/2019/10/14/colleges-quietly-rank-prospective-students-based-their-personal-data/

Makhlouf, K., Zhioua, S., & Palamidessi, C. (2021). On the applicability of machine learning fairness notions. *SIGKDD Explor. Newsl.*, *23*(1), 14–23. https://doi.org/10.1145/3468507.3468511

Marcinkowski, F., Kieslich, K., Starke, C., & Lünich, M. (2020). Implications of AI (un-) fairness in higher education admissions: the effects of perceived AI (un-) fairness on exit, voice and organizational reputation. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,

Martin, L. (1994). *Equity and general performance indicators in higher education*. Australian Government Publishing Service.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1-35.

Microsoft Australia Education. (2018). The University of Sydney: Exploring AI and implementing Chatbots to achieve operational efficiencies. https://educationblog.microsoft.com/en-us/2018/03/the-university-of-sydney-exploring-ai-and-implementing-chatbots-to-achieve-operational-efficiencies/

Miller, B. (2021). Is technology value-neutral? *Science, Technology, & Human Values*, *46*(1), 53-80. https://doi.org/10.1177/0162243919900965

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. The MIT Press.

Narayanan, A. (2018). Translation tutorial: 21 fairness definitions and their politics. Procceedongs of the Conference on Fairness. Accountability and Transparency, New York.

Naylor, R., Coates, H., & Kelly, P. (2016). From equity to excellence: Reforming Australia's National Framework to create new forms of success. In A. Harvey, C. Burnheim, & M. Brett (Eds.), *Student Equity in Australian Higher Education: Twenty-five years of A Fair Chance for All* (pp. 257-274). Springer.

Naylor, R., & James, R. (2016). Systemic equity challenges: An overview of the role of Australian universities in student equity and social inclusion. In M. Shah, A. Bennett, & E. Southgate (Eds.), *Widening Higher Education Participation* (pp. 1-13). Chandos Publishing. https://doi.org/10.1016/B978-0-08-100213-1.00001-9

Neef, D. (2015). Digital Exhaust: what everyone should know about big data, digitization and digitally driven innovation. Pearson Education.

Norton, A., Cherastidtham, I., & Mackey, W. (2018). *Dropping out: The benefits and costs of trying university,* Grattan Institute. https://grattan.edu.au/report/dropping-out/

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., & Krasanakis, E. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1356.

O'Niel, C. (2016). Weapons of Math Destruction: How big data increases inequality and threatens democracy. Penguin.

O'Shea, S. (2015). “I generally say I am a mum first... But I’m studying at uni”: The narratives of first-in-family, female caregivers transitioning into an Australian university. *Journal of Diversity in Higher Education*, *8*(4), 243.

O’Sullivan, D., Rahamathulla, M., & Pawar, M. (2020). The impact and implications of COVID-19: An Australian perspective. *The International Journal of Community and Social Development*, *2*(2), 134-151.

Obermeyer, Z., Nissan, R., Stern, M., Eaneff, S., Bembeneck, E. J., & Mullainathan, S. (2021). *Algorithmic Bias Playbook,* The University of Chicago Booth School of Business, The Centre for Applied Artificial Intelligence. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries [Review]. *Frontiers in Big Data*, *2*(13). https://doi.org/10.3389/fdata.2019.00013

Pangburn, D. (2019). Schools are using software to help pick who gets in. What could go wrong? *Fast Company*. https://www.fastcompany.com/90342596/schools-are-quietly-turning-to-ai-to-help-pick-who-gets-in-what-could-go-wrong

Pardo, A., Mirriahi, N., Martinez-Maldonado, R., Jovanovic, J., Dawson, S., & Gašević, D. (2016). *Generating actionable predictive models of academic performance,* Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh, United Kingdom. https://doi.org/10.1145/2883851.2883870

Pearl, J. (2019). The Book of Why: The New Science of Cause and Effect. Penguin Press.

Phan, T., Goldenfein, J., Mann, M., & Kuch, D. (2021). Economies of virtue: The circulation of ‘ethics’ in big tech. *Science as Culture*, 1-15. https://doi.org/10.1080/09505431.2021.1990875

Piano, S. L. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*(1), 1-7.

Qureshi, B., Kamiran, F., Karim, A., Ruggieri, S., & Pedreschi, D. (2020). Causal inference for social discrimination reasoning. *Journal of Intelligent Information Systems*, *54*(2), 425-437. https://doi.org/10.1007/s10844-019-00580-x

Raffaghelli, J. E., & Stewart, B. (2020). Centering complexity in 'educators' data literacy' to support future practices in faculty development: a systematic review of the literature. *Teaching in Higher Education*, *25*(4), 435-455. https://doi.org/10.1080/13562517.2019.1696301

Rainie, S. C., Kukutai, T., Walter, M., Figueroa-Rodríguez, O. L., Walker, J., & Axelsson, P. (2019). Indigenous data sovereignty. In T. Davies, Stephen B. Walker, Mor Rubinstein, & F. Perini (Eds.), *The State of Open Data: Histories and Horizons* (pp. 300-319). African Minds and International Development Research Centre,.

Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2020). A survey on predictive models of learning analytics. *Procedia Computer Science*, *167*, 37-46. https://doi.org/https://doi.org/10.1016/j.procs.2020.03.180

Razak, S. F. A., Mashhod, F., Zaidan, Z. N. B., & Yogarayan, S. (2021, 3-5 Aug. 2021). RPA-based bots for managing online learning materials. 2021 9th International Conference on Information and Communication Technology (ICoICT),

Ruggieri, S., Pedreschi, D., & Turini, F. (2010). Data mining for discrimination discovery. *ACM Trans. Knowl. Discov. Data*, *4*(2), Article 9. https://doi.org/10.1145/1754428.1754432

Saha, D., Schumann, C., McElfresh, D. C., Dickerson, J. P., Mazurek, M. L., & Tschantz, M. C. (2020). Human comprehension of fairness in machine learning. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,

Seidel, E., & Kutieleh, S. (2017). Using predictive analytics to target and improve first year student attrition. *Australian Journal of Education*, *61*(2), 200-218.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the conference on fairness, accountability, and transparency,

Selwyn, N., & Gašević, D. (2020). The datafication of higher education: discussing the promises and problems. *Teaching in Higher Education*, *25*(4), 527-540. https://doi.org/10.1080/13562517.2019.1689388

Shah, H. (2018). Algorithmic accountability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128), 20170362.

Siemens, G., Dawson, S., & Lynch, G. (2013). Improving the Quality and Productivity of the Higher Education Sector: Policy and strategy for systems-level deployment of learning analytics. Society for Learning Analytics Research. https://ltr.edu.au/resources/SoLAR_Report_2014.pdf

Slade, S., & Prinsloo, P. (2013). Learning Analytics:Ethical Issues and Dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529. https://doi.org/10.1177/0002764213479366

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F. N., Arvanitakis, G., Benevenuto, F., Gummadi, K. P., Loiseau, P., & Mislove, A. (2018). Potential for Discrimination in Online Targeted Advertising. *Proceedings of Machine Learning Research*, *81*, 1-15.

Stephens, D. (2001). Use of computer assisted assessment: Benefits to students and staff. *Education for Information*, *19*, 265-275. https://doi.org/10.3233/EFI-2001-19401

Stephenson, B., Cakitaki, B., & Luckman, M. (2018). Ghosts in the machine: Towards solving the mystery of non-participating enrolments (NPE) and understanding their importance for institutional analytics, Australasian Association for Institutional Research (AAIR) Forum, Melbourne.

Stephenson, B., Cakitaki, B., & Luckman, M. (2021). *"Ghost student" failure among equity cohorts: Towards understanding Non-Participating Enrolments (NPE),* National Centre for Student Equity in Higher Education (NSEHE), Curtin University. Perth. https://www.ncsehe.edu.au/publications/success-failure-higher-education-uneven-playing-fields/

Suresh, H., & Guttag, J. V. (2020). A framework for understanding unintended consequences of machine learning. *ArXiv*. https://arxiv.org/pdf/1901.10002.pdf

The University of Melbourne. (2019). *Artificial Intelligence: Governance and Leadership White Paper 2019, Response from the University of Melbourne*. https://about.unimelb.edu.au/__data/assets/pdf_file/0024/82662/UoM_submission_AI-Governance-White-Paper_2019.pdf

Turcu, C., & Turcu, C. (2019). *On robotic process automation and its integration in higher education,* International Conference of The Future of Higher Education (ICT4777), https://conference.pixel-online.net/FOE/files/foe/ed0010/FP/6243-ICT4777-FP-FOE10.pdf

van der Aalst, W. M. P., Bichler, M., & Heinzl, A. (2018). Robotic process automation. *Business & Information Systems Engineering*, *60*(4), 269-272. https://doi.org/10.1007/s12599-018-0542-4

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & society*, *12*(2), 197-208.

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2), 1-17. https://doi.org/10.1177/2053951717743530

Veale, M., Van Kleek, M., & Binns, R. (2018). *Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making,* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, https://arxiv.org/abs/1802.01029

Verma, S., & Rubin, J. (2018). Fairness definitions explained. 2018 IEEE/ACM International Workshop on Software Fairness (Fairware),

Walker-Gibbs, B., Ajjawi, R., Rowe, E., Skourdoumbis, A., Krehl, M., Thomas, E., O'Shea, S., Bennett, S., Fox, B., & Alsen, P. (2019). *Success and failure in higher education on uneven playing fields,* National Centre for Student Equity in Higher Education (NSEHE), Curtin University. Perth. https://www.ncsehe.edu.au/publications/success-failure-higher-education-uneven-playing-fields/

Walter, M. (2016). Data politcs and Indigenous representation in Australian statistics. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 79-97). ANU press.

Walter, M., Lovett, R., Maher, B., Williamson, B., Prehn, J., Bodkin-Andrews, G., & Lee, V. (2021). Indigenous Data Sovereignty in the era of big data and open data. *Australian Journal of Social Issues*, *56*(2), 143-156.

Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, *22*(3), 233-243.

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1-37.

Wang, S., Beheshti, A., Wang, Y., Lu, J., Sheng, Q. Z., Elbourn, S., Alinejad-Rokny, H., & Galanis, E. (2021). Assessment2Vec: learning distributed representations of assessments to reduce marking workload. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova, *Artificial Intelligence in Education* Cham.

Webber, K. L., & Zheng, H. Y. (Eds.). (2020). *Big Data on Campus: Data Analytics and Decision Making in Higher Education*. Johns Hopkins University Press.

Weller, M. (2015). MOOCs and the Silicon Valley narrative. *Journal of interactive media in education*, *1*(5), 1-7. https://doi.org/http://dx.doi.org/10.5334/jime.am

Willems, J. (2010). The equity raw-score matrix – a multi-dimensional indicator of potential disadvantage in higher education. *Higher Education Research & Development*, *29*(6), 603-621. https://doi.org/10.1080/07294361003592058

Willems, J., Farley, H., & Garner, J. (2018). Digital equity in Australian higher education: how prisoners are missing out. 41st HERDSA Annual International Conference, Adelaide, SA.

Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology*, *45*(2), 107-114. https://doi.org/10.1080/17439884.2020.1761641

Williamson, B., & Hogan, A. (2020). *Commercialisation and privatisation in/of education in the context of Covid-19,* (929510997X). Education International. Brussels, Belgium.

Wise, P., & Mathews, R. (2011). *Socio-Economic Indexes for Areas: Getting a Handle on Individual Diversity Within Areas,* Australian Bureau of Statistics, Analytical Services Branch. Canberra. https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/C523F80A0B938ACBCA25790600138037/$File/1351055036_sep%202011.pdf

Wolff, A., Zdrahal, Z., Herrmannova, D., & Knoth, P. (2014). Predicting student performance from combinded data sources. In A. Peña-Ayala (Ed.), *Educational data mining*. Springer International Publishing. https://doi.org/10.1007/978-3-319-02738-8_7

Yu, R., Lee, H., & Kizilcec, R. F. (2021). *Should college dropout prediction models Include protected attributes?,* Proceedings of the Eighth ACM Conference on Learning @ Scale, Virtual Event, Germany. https://doi.org/10.1145/3430895.3460139

Zacharias, N., & Brett, M. (2019). *Student Equity 2030: A long-term strategic vision for student equity in higher education,* National Centre for Student Equity in Higher Education (NCSEHE), Curtin University. Perth, WA.

Zarsky, T. (2013). Transparent predictions. *University of Illinois Law Review*, *2013*, 1503-1569.

Zhang, L., Wu, Y., & Wu, X. (2016). A causal framework for discovering and removing direct and indirect discrimination. *arXiv:1611.07509*.

Žliobaitė, I., & Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. *Artificial Intelligence and Law*, *24*(2), 183-201.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power (First edition. ed.). PublicAffairs.