

https://dspace.org/

# Does South Africa Offer Sufficient Safeguards To Address Algorithmic Discrimination In Recruitment? A Comparative Analysis

**Qaba, Kanyisa Palesa**

**2024**

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/45928

Qaba, Kanyisa Palesa. (2024). Does South Africa Offer Sufficient Safeguards To Address Algorithmic Discrimination In Recruitment? A Comparative Analysis [Master`s dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/45928

Downloaded from a DSpace repository. https://dspace.org/

# DOES SOUTH AFRICA OFFER SUFFICIENT SAFEGUARDS TO ADDRESS ALGORITHMIC DISCRIMINATION IN RECRUITMENT? A COMPARATIVE ANALYSIS

*by*

1405796

Submitted in partial fulfillment of the requirements for the degree of

Master of Laws by Coursework and Research Report

at the University of the Witwatersrand, Johannesburg

Date: 29 February 2024

DECLARATION

I, ____1405796_____(Student number, declare that this Research Report is my own unaided work. It is submitted in partial fulfillment of the requirements for the degree of Master of Laws (by Coursework and Research Report) at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

I have submitted my final Research Report through TurnItIn and have attached the report to my submission.

Word Count: _____10319_____[Incl. the body of your report, footnotes **BUT** excl. title page, declaration, abstract, table of contents, and bibliography]

## Table of Contents

**ABSTRACT** 4

**1. INTRODUCTION** 5

**2. DEFINING AI TECHNOLOGY** 8

2.1 DEFINITION OF AI 8

2.2 MACHINE LEARNING 10

2.3 PREDICTIVE ANALYTICS 11

**3. THE SOUTH AFRICAN LEGAL FRAMEWORK** 12

3.1 THE CONSTITUTION OF THE REPUBLIC OF SOUTH AFRICA,1996 (THE CONSTITUTION) 13

3.2 THE PROMOTION OF EQUALITY AND THE PREVENTION OF UNFAIR DISCRIMINATION ACT 4 OF 2000 (PEPUDA) 14

3.3 THE PROTECTION OF PERSONAL INFORMATION ACT (POPIA) 16

3.4 THE EMPLOYMENT EQUITY ACT (EEA) 21

**4. THE ARTIFICIAL INTELLIGENCE ACT OF THE EUROPEAN UNION (AI ACT)** 23

**5. LESSONS FOR SOUTH AFRICA** 29

**6. CONCLUSION** 30

**7. BIBLIOGRAPHY** 31

ABSTRACT

The era of the Fourth Industrial Revolution has seen a surge in the proliferation of Artificial Intelligence in virtually all industries, impacting employment, agriculture, and health care. This contribution traces the evolution of AI from its inception by Alan Turing to the current age of deep learning and digitalisation. I examine AI's increasing role in the recruitment process, highlighting both its potential to enhance efficiency and its danger of exacerbating discrimination, both directly and through proxies. I highlight that despite the possibility of algorithmic bias, scholars argue that identifying and rectifying it is more legally manageable than conventional human bias. The essay, therefore, advocates for the fair regulation of AI hiring tools that mitigate discrimination for all persons while still allowing space for the growth of the field of AI. First, focusing on South Africa's legislative framework, the essay investigates whether current laws offer adequate protection for victims of recruitment algorithmic discrimination. In the scope of the discussion, I offer a critique of existing legal protections, concluding that currently promulgated frameworks are insufficient. To account for these legal inadequacies, the paper turns to consider solutions at the international level, namely the AI Act of the European Union. Considering the AI Act's provisions circumventing AI bias, which echo a risk-based approach, I make suggestions to be considered by the legislature when it inevitably is tasked with regulating AI and its potentially harmful technologies. Ultimately, the paper underscores the importance of addressing AI discrimination in the recruitment process through robust legislative regulation.

## 1. INTRODUCTION

The emergence of the Fourth Industrial Revolution has led to a surge in the integration of Artificial Intelligence (AI) systems across various sectors, most notably in employment, entertainment, and healthcare.[1] The field of AI was first conceived in 1936 by Alan Turing, who invented the Turing Test.[2] The purpose of the test was to explore the feasibility of intelligent machines.[3] By 1955, scholar John McCarthy believed that most aspects of human learning and intelligence could be transferred to machines.[4] John McCarthy went on to coin the term Artificial Intelligence in 1956.[5] The obstacles faced during the 1960s and 1970s halted the further development of AI.[6] This era is referred to as AI's winter, which was caused by the lack of sophisticated computing power required to effectively deploy AI systems.[7] Governments and technological corporations resumed investment in AI research, rejuvenating the field of study during the 1980s and 1990s.[8] With the advent of deep learning (DL) in 2009 came the solidification of AI as a staple within the global technological discourse.[9] The decade after that saw the commonality of digitalisation, the use of the internet, and the emergence of big data, all of which contributed to the exponential rise of AI integration into mainstream society.[10]

AI has become part and parcel of conducting business in both the public and private sectors.[11] The workplace has seen an exponential rise in the use of AI, particularly to inform hiring decisions.[12] AI can remove mundane tasks that would have ordinarily required human input.[13] Yet, it also has the potential to exacerbate historically charged discrimination.[14] This fact has not deterred organisations from deploying AI recruitment solutions within their

[1] E.O Arakpogun et al 'Artificial Intelligence in Africa: challenges and opportunities' in Allam Hamdan et al '*The Fourth Industrial Revolution Implementation of Artificial Intelligence for Growing Business Success*' (2021) at 380.
[2] T.L Mamela Assessment of the impact of Artificial Intelligence on the performance of workforce at a South African Banking Institution (published LLM thesis, University of Johannesburg, 2021) at 2.
[3] D Mhlanga 'Artificial Intelligence in the industry of 4.0 and its impact on poverty, innovation, infrastructure development and the sustainable development goals: Lessons from emerging economies?' (2021) *UJ Content* at 4.
[4] Ibid.
[5] Ibid.
[6] Ibid.
[7] Ibid.
[8] Ibid.
[9] Ibid.
[10] Ibid.
[11] Ibid.
[12] A Kelly - Lyth 'Challenging biased hiring algorithms' (2021) *Oxford Journal of Legal Studies* at 899.
[13] E.O Arakpogun et al op cit note 1 at 379.
[14] Ibid. Hence, AI's categorisation as a technology which is simultaneously disruptive and developmental.

business models. Human recruiters who previously used keyword searches to scan curriculum vitae (CVs) for ideal candidates are currently using potentially discriminatory AI systems.[15] Organisations are now able to formulate unique hiring criteria using previously submitted applications.[16] Mounting evidence shows that these innovations tend to discriminate against existing marginal groups.[17] For example, interview assessment innovations that decipher human speech tend to associate traditionally female names with household work.[18] This example is not to say that human bias did not exist before AI proliferation in hiring, it did and is the reason why discrimination has crept into AI recruitment systems.[19] Scholars generally concur that identifying and rectifying algorithmic bias can be simpler and more effective compared to human bias, which is often difficult to detect.[20] Hence, well-regulated automated hiring tools may be the solution to mitigate discrimination in recruitment as a whole, allowing marginalised groups the highest level of legal protection from opaque discriminatory patterns in hiring.[21] This contribution will advocate for the fair regulation of AI hiring tools, which ensures a balance between the legitimate efficiency interests of organisations while still allowing applicants a level of protection against unlawful differentiation.

Globally, reports indicate that the present job market is characterised by significant numbers of job applications and elevated rates of staff resignations, primarily due to the increased accessibility of avenues for submitting applications.[22] Thus, companies have acknowledged the requirement for speed and proficiency in recruitment operations, ensuring that the best-suited candidates are timeously allocated employment.[23] Recruitment innovations are used in differing ways at each phase of hiring to ensure constant efficiency and productivity.[24] Once an application is submitted, predictive tools assist organisations in

[15] J Adams-Prassl 'What if your boss was an algorithm? Economic incentives, legal challenges and the rise of Artificial Intelligence at work' (2019) *Comparative Labour Law & Policy Journal* at 125.
[16] Ibid.
[17] A Kelly- Lyth op cit note 12 at 900.
[18] Ibid.
[19] Ibid.
[20] Ibid.
[21] Ibid.
[22] Z.F Borgesius 'Discrimination, Artificial Intelligence and Algorithmic Decision Making' (2018) *Council of Europe Directorate General of Democracy* at 24. This is not the case in South Africa, where high application volumes are attributed to the suffering employment rate and the scarcity of positions available. It has been reported that the COVID-19 pandemic exacerbated the deeply entrenched structural issues associated with the legacies of colonialism and apartheid. Further dividing the economic haves from the have-nots specifically on the grounds of race. See Kopano Gumbi 'South Africa's unemployment rate drops marginally in second quarter' available at *https://www.reuters.com/world/africa/south-africas-unemployment-rate-down-slightly-326-q2-2023-08-15/.*
[23] Ibid.
[24] M Bogen and A Rieke 'Help wanted: An examination of hiring algorithms,equity,and bias' (2018) *Upturn* at 3.

evaluating and expeditiously rejecting or further assessing the candidate's suitability.[25] These tools rely on the diction in the applicant's CV, the background of the candidate, and details about the position in conjunction with trained data sets to determine a suitable fit for a role.[26] Data sets are built using former or current model employees ' CVs, cover letters, or performance reviews.[27] While AI enhances recruitment efficiencies, it could also potentially allow for the occurrence of discrimination, particularly when it filters applications, favouring members of elite groups over disenfranchised candidates.[28] To reduce the likelihood of discrimination from occurring, organisations require proper functioning software, representative data, and inclusive attribute selection in the data used to train algorithms.[29]

Defective software, unrepresentative data selection, and the absence of human oversight have been identified as the main perpetrators of discriminatory AI.[30] Borgesius suggests that AI applications are frequently black boxes, meaning their decisions can be perplexing, with rationales that are difficult to decipher.[31] This is attributed to the opaqueness and lack of transparency of AI tools.[32] Once the algorithm is trained using data inputs, human intervention is typically removed, allowing the AI tool to operate autonomously, causing a lack of transparency.[33] To illustrate, in 2018, Amazon discarded its human resources management tool, as it displayed patterns of sexism in its predictions and recommendations.[34] The system was first developed in 2014 to assist the human resources department in autonomously sifting through applications.[35] In this case, applications that included the word 'women's' were given a low rating for suitability and fit.[36] The system also allegedly demoted alumni educated at specific all-girls learning institutions.[37] The issue was that the system had been trained using data sourced from previous employees who mostly consisted of men, resulting in prejudicial

[25] Ibid.
[26] F.Z Borgesius op cit note 22 at 25.
[27] Ibid at 17.
[28] Ibid at 10.
[29]Ibid.
[30] Ibid. If the initial data set is itself incorrect, unrepresentative or prejudiced, the resulting predictions may mirror these limitations resulting in algorithmic discrimination.
[31] Ibid.
[32] Ibid.
[33] IBM Data and AI Team 'Blog AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What is the difference?' IBM Blog last accessed from *https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/* accessed on 24 September 2023.
[34] M Oppenheim 'Amazon scrapped 'sexist AI' recruitment tool' The Independent 11 October 2018 at last accessed from *https://www.independent.co.uk/tech/amazon-ai-sexist-recruitment-tool-algorithm-a8579161.html*, accessed on 15 September 2023.
[35] Ibid.
[36] Ibid.
[37] Ibid.

outcomes for female candidates.[38] This example illustrates the occurrence of direct discrimination based on gender. Scholars have also identified indirect or proxy discrimination[39] as a threat to equality in automated recruitment technology.

In light of this situation, this paper aims to answer whether South Africa's current legislative framework provides sufficient legal recourse for victims of recruitment algorithmic discrimination. First, I will discuss the difficulty researchers have in agreeing on a universal definition of AI. I will then provide a working definition of AI for my legal discussion. Secondly, I will analyse whether the current national and international legislative framework is equipped to provide legal protection to job seekers. In separate subsections, I will critically analyse the legal protection offered by the Promotion of Equality and Prevention of Unfair Discrimination Act, the Protection of Personal Information Act, and the Employment Equity Act. I will advance that the statutes are insufficient in combatting AI discrimination in the workplace. I will then critically analyse the drafts of the Artificial Intelligence Act of the European Union (EU), highlighting the innovations and shortcomings of their approaches. Lastly, I will emphasise the possible lessons South Africa's legislature could glean as it contemplates future AI regulation.

## 2. DEFINING AI TECHNOLOGY

### 2.1 Definition of AI

Despite the popularity of AI as a field of study, challenges persist in defining it precisely.[40] The dynamic nature of AI makes it difficult for scholars to find a consensus on an all-inclusive definition.[41] The term is often used interchangeably to describe tools employing machine learning (ML), algorithmic decision-making, and automated decision-making.[42] AI generally denotes a system that utilises the input of big data and computational procedures to provide predictions that either serve to assist or completely replace human decision-making.[43] AI-powered systems are constructed by evaluating big data inputs to extricate patterns, which are

[38] Ibid. In a differing study, it had been shown that having the name Jared, and the credential of playing high school lacrosse rendered an individual a more favourable candidate, despite these facts have no link to the ability to perform a specific job.
[39] Discussed in greater detail below in paragraph 2.
[40] E.O Arakpogun et al op cit note 1 at 359.
[41] Ibid.
[42] P.T Kim and M.T Bodie 'Artificial Intelligence and the challenges of workplace discrimination' (2021) *Journal of Labour & Employment* at 290.
[43] Ibid.

then used to forecast future outcomes.[44] An additional definitional complexity is the categorisation of AI into various typologies, each ascribing importance to its differing attributes.[45] While typologies do not directly change the definition, they delineate the various technological focal points, exacerbating the lack of definitional consensus.[46]

The first typology distinguishes AI systems according to the variety of intelligence they display, the second differentiates AI based on the technology it is embedded in, and the third distinguishes AI on the function executed by the system.[47] AI-based intelligence revolves around innovations capable of adapting and learning in a human-like manner.[48] This typology divides AI into three classes: artificial narrow intelligence, artificial general intelligence, and artificial superintelligence.[49] AI narrow intelligence is characterised by its ability to tackle one predetermined problem and is not able to transfer its skills to diversified domains.[50] General AI is geared at achieving human-like skills in a particular activity.[51] It is believed that once general AI is accomplished, it could progress to superintelligence, which will exceed human performance in a given task.[52] Superintelligence is premised on the possibility of versatile evolutionary system norms.[53] Minsky's 1968 definition closely aligns with this typology, conceiving AI 'as the science that deals with the development of machines capable of performing functions that a human can perform and that require human intelligence.'[54] AI-based on technology offers explanations of innovations based on the machinery in which it is embedded.[55] The technologies include DL, ML (its subcategories), natural language processing, and robots.[56] Nilsson's 1998 definition closely resonates with this typology, defining AI as a 'part of computer science that focuses on machine learning, making computers act intelligent, continuously learning and improving their function.'[57] Finally, AI based on functionality consists of four categories 'conversational, biometric, algorithmic, and robotic

[44] Ibid.
[45] H Benbya & T.H Davenport & S Pachidi 'Special issue editorial. Artificial Intelligence in organisations: Current and future opportunities' (2020) *MIS Quarterly Executive* Vol 19 at 3.
[46] Ibid.
[47] T.L Mamela op cit note 2 at 3.
[48] H Benbya & T.H Davenport & S Pachidi op cit note 45 at 3.
[49] Ibid.
[50] Ibid.
[51] Ibid at 4.
[52] Ibid.
[53] Ibid. Superintelligence does not exist yet and is currently being researched and discussed.
[54] PR Palos-Sánchez & P Baena-Luna et al 'Artificial Intelligence and Human Resources Management: A Bibliometric Analysis' (2022) *Applied Artificial Intelligence* at 3.
[55] Ibid at 4. AI General intelligence and Superintelligence are still subject to research and do not exist.
[56] Ibid.
[57] Ibid at 3.

[AI].'[58] Conversational AI denotes the ability of a system to understand and reply by way of natural human language.[59] Biometric AI depends on procedures that can estimate an individual's unique features, such as fingerprint or facial recognition.[60] Algorithmic AI centers around the use of ML to provide predictions based on input data.[61] On the other hand, robotic AI entails the fusion of physical robots with the capacity to perform sophisticated and complex duties.[62] Bolander's 2019 definition echoes this typology, explicating AI as the 'construction of machines – computers or robots that can perform tasks that otherwise only humans have done.'[63] For purposes of this discourse, the author concedes with the definition provided by Minsky. Minsky's definition is clear and technologically neutral, allowing a high level of flexibility for future developments and wide-ranging applications. Clarity and flexibility will allow for legal certainty as well as a wide scope of applicability.[64] The definition's clearness of what constitutes AI function will put society at ease, encouraging rather than deterring AI investment and the growth of the industry.[65] A wider definitional scope will allow lawmakers to increase legal protection.[66]

## 2.2 Machine Learning

AI comprises two major learning processes, namely ML and DL.[67] ML involves employing algorithms to analyse input data, extracting insights from it, and generating predictions across different domains.[68] A machine is trained through the input of big data sets that provide it with the capacity to learn how to complete a given task.[69] A simple example of ML is Google search.[70] When a user searches a particular question, Google will display the most relevant

[58] Ibid at 4.
[59] Ibid.
[60] Ibid.
[61] H Benbya & TH Davenport & S Pachidi op cit note 45 at 3. AI is also framed by two other categories, weak AI and strong AI. With weak (or narrow) AI, applications typically are geared at a specified task unable to transfer proficiencies to an array of domains. Conversely, strong AI is characterised by diversified intelligence and the ability to self-correct and learn based on historical findings and errors.
[62] Ibid at 5.
[63] PR Palos-Sánchez & P Baena-Luna et al op cit note 54 at 3.
[64] Ibid.
[65] Ibid.
[66] Ibid.
[67] W.H Gravett 'Is the dawn of the lawyer robot upon us? The Fourth Industrial Revolution and the future of lawyers' (2020) *PELJ* at 7.
[68] M.B Rapanyane & F. R Sethole 'The rise of AI and robots in the 4th Industrial Revolution: Implications for the future of South African job creation' (2020) *Contemporary Social Science* at 492.
[69] Ibid.
[70] N Adams South African Company Law in the Fourth Industrial Revolution: Does Artificial Intelligence create a need for legal reform? (published LLM thesis, University of Western Cape 2021) at 51.

results by assessing the user's search history.[71] Conversely, DL is a subset of ML that 'relies on what is referred to as neural networks, (which comprise of) an interconnected group of nodes designed to mimic the activity of neurons in the human brain to recognise complex patterns in data sets.'[72] The principal difference between ML and DL is the type of training data embedded in each system.[73] ML relies on data sets that involve a level of human intervention, enabling the application to forecast, learn, and predict accurately.[74] Programmers code a list of traits to understand the distinction in the input data powered by predominantly labeled information.[75] For instance, in the case of various fast foods, a burger, pizzas, and a taco.[76] A programmer developing an ML application might identify the unique characteristics of each food item and could conclude that bread is the defining feature.[77] They then label each image according to this distinction to streamline the learning of the ML system through supervised learning.[78] The labelling of information makes ML a powerful predictive analytical tool.[79] DL systems, on the other hand, do not automatically incorporate labelled input data.[80] DL applications can ingest unlabelled data in the form of images and text, which it then autonomously groups according to a designated feature. [81] Although DL has gained traction since the COVID-19 pandemic, ML is the predominantly researched and used system.[82] Thus, the focus of this contribution will be ML.

## 2.3 Predictive Analytics

Predictive Analytics (PA) refers to the use of input data and statistical and ML procedures that evaluate information enabling the possibility of determining future estimations.[83] The process involves advanced analytic techniques used to identify trends and relationships within the data, which then provide educated forecasts.[84] The principles of PA are transferable to algorithmic

[71] Ibid.
[72] IBM Data and AI Team op cit note 33.
[73] Ibid.
[74] Ibid.
[75] Ibid.
[76] Ibid
[77] Ibid
[78] Ibid.
[79] Ibid.
[80] Ibid.
[81] Ibid.
[82] F.Z Borgesius op cit note 22 at 24. AI, ML and predictive analytics will be used interchangeably throughout the discussion.
[83] V Bozic 'AI and Predictive Analytics' last accessed from *https://www.researchgate.net/publication/370074080_AI_and_Predictive_Analytics* on 1 December 2023
[84] Ibid.

recruitment.[85] To demonstrate, in 2014 Amazon patented a predictive shipping tool that evaluates a shopper's previous orders, wish lists, and searches, which then can predict an estimation of an upcoming order, pre-packaging it before it is even placed.[86] The same processes are applied to recruitment, wherein an organisation, based on previous employee data, identifies traits that it deems a successful candidate has and proceeds to filter prospective candidates based on the predictive tool.[87] If the data collected has elements of racial, gender, or religious exclusions, the recommendations provided will reflect these discriminatory patterns.[88] Exclusions occur directly (Amazon sexist AI) or through proxy discrimination. Proxy discrimination occurs when, on the face of it, a system makes neutral suggestions; however, on further investigation, the system has used secondary information, resulting in unfair discrimination.[89] Usually, this occurs when protected features are removed from a data training set to eliminate discrimination, yet the system still picks up on other features that indirectly perpetuate discrimination.[90] For example, a technological vendor can remove race categorisation from input data.[91] Yet, the AI still learns that individuals (who are of a specific race) who reside in a specified area are more favourable candidates.[92] Put differently, the system using a neutral feature such as residential address, makes correlations between race and employability. If the residential address correlates to race, the system will indirectly discriminate against applicants on a protected ground, as it will favour the race of individuals who are in favourable locations, which is usually based on proximity to the workplace. In South Africa (SA), where residential location is commonly racialised, proxy discrimination based on address is very likely to occur.[93]

## 3. THE SOUTH AFRICAN LEGAL FRAMEWORK

[85] Ibid.
[86] B.A Lozada 'The emerging technology of predictive analytics: Implications for homeland security' (2014) *Information Security Journal: A Global Perspective* at 119.
[87] V Bozic op cit note 83.
[88] Ibid.
[89] F.Z Borgesius op cit note 22 at 14.
[90] A Kelly – Lyth op cit note 12 at 906.
[91] Ibid.
[92] Ibid.
[93] See Group Areas Act of 1950.

SA offers individuals several legislative avenues to pursue remedies when faced with unlawful discrimination. This section will introduce each legislation and critically analyse its applicability to algorithmic discrimination in recruitment.

### 3.1 The Constitution of the Republic of South Africa,1996 (The Constitution)

The abolition of the Apartheid era ushered in democratic constitutionalism in SA, whereby the Constitution reigns as the supreme law.[94] The Constitution signifies the beginning of a new social and political era founded on the values of human dignity, the achievement of equality, and the advancement of human rights and freedoms.[95] Hence, any laws inconsistent with these values are deemed null and void.[96] Chapter 2 of the Constitution contains the Bill of Rights, which is viewed as the cornerstone of democracy in South Africa.[97] It upholds the rights of every individual within our nation and reinforces the democratic principles of human dignity, equality, and freedom.[98] In addition to their function as constitutional values, dignity, equality, and freedom are also framed as distinct rights that have a special function in the application and interpretation of the Constitution.[99] The Constitution further frames equality as both a right and a value that is deeply connected with the achievement of human dignity.[100] In *President of the Republic of South Africa v Hugo,* Goldstone J asserted that the prohibition of unlawful discrimination incorporates the recognition that 'all human beings will be accorded equal dignity and respect regardless of their membership of particular groups.'[101] Similarly, the court has formulated human dignity as a right and value that is closely interlinked with other fundamental rights, such as the right to privacy.[102] The right to privacy is viewed as reverence for identity and autonomy, which are traits encapsulated under individualistic dignity.[103]

S9 of the Constitution, more commonly known as the equality clause, provides that the State, its various agencies, and individuals are prohibited from the perpetuation of unfair

---

[94] A Chaskalson 'Dignity as a Constitutional Value: A South African Perspective' (2011) *American University International Law Review* at 1380.
[95] The Constitution of the Republic of South Africa, 1996 at s2.
[96] Ibid.
[97] A Chaskalson op cit note 94 at 1380.
[98] Ibid.
[99] Ibid.
[100] P De Vos 'Equality, human dignity and privacy rights' in D. Brand et al '*South African Constitutional Law in Context*' (2014) Oxford University Press at 426.
[101] *President of the Republic of South Africa v Hugo* 1997 (4) SA 1 at para 41.
[102] A Hughes 'Human Dignity and Fundamental Rights in South Africa and Ireland' (2014) *Pretoria University Law Press* at 263.
[103] Ibid.

discrimination based on a listed ground in s9(3).[104] The provision further obligates the legislature to enact statutes that preclude unfair discrimination.[105] The constitutional values and rights to equality, dignity, and privacy are central in discussing discriminatory AI, as it undermines the egalitarian values embedded in our law.

### 3.2 The Promotion of Equality and the Prevention of Unfair Discrimination Act 4 of 2000 (PEPUDA)

The most legally compatible law addressing claims of algorithmic discrimination is PEPUDA, a statute specifically intended to counteract discriminatory laws, policies, and practices.[106]

S6 of PEPUDA can be utilised to substantiate a claim of algorithmic discrimination perpetuated by a private entity. The provision states that 'neither the State nor any person may unfairly discriminate against any person'.[107] PEPUDA further defines discrimination as both commissions and omissions 'including a policy, law, rule, practise, condition or situation which directly or indirectly - (a) imposes burdens, obligations or disadvantage on; or (b) withholds benefits, opportunities or advantages from any person on one or more of the prohibited grounds'.[108] Algorithmic recruitment tools informing hiring outcomes fall within the ambit of a workplace 'practice'.[109] To prove direct discrimination, for example, Amazon's sexist AI, a female candidate could allege that there is a direct differentiation of applicants based on gender, which favours male candidates thus disadvantaging women.[110] Essentially, a *prima facie* case of direct discrimination can be established if an applicant can prove that bias in a hiring practice occurred on a protected ground.[111] The unlawfulness of the system is proved by evidence presented by the candidate.[112] The practical limitation in such cases would be the difficulty of gathering evidence proving that the AI system generated a discriminatory outcome.[113] An individual's challenge in bringing such a claim is that AI systems generally operate covertly, rendering their decisions unexplainable, making it difficult to ascertain if the decision was

[104]The Constitution at s9(3) and (4).
[105] Ibid at s9(4).
[106] F.Z Borgesius op cit note 22 at 18.
[107] PEPUDA s6. Includes juristic and natural persons.
[108] Ibid at s1 (viii) The prohibited grounds include 'race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth.'
[109] A Kelly – Lyth op cit note 12 at 906.
[110] M Oppenheim op cit note 34 at 1.
[111] A Kelly – Lyth op cit note 12 at 906.
[112] Ibid at 1. PEPUDA places the burden of proof on the claimant, who must present evidence alleging unlawful discrimination.
[113] A Kelly – Lyth op cit note 12 at 906.

indeed discriminatory and if so on what grounds.[114] S1 of PEPUDA extends the prohibited grounds to also include

> 'any other ground where discrimination based on that ground - (i) causes or perpetuates systemic disadvantage; (ii) undermines human dignity or (iii) adversely affects the equal enjoyment of a person's rights and freedoms in a serious manner that is comparable to discrimination on a ground in paragraph (a)[listed grounds].'[115]

PEPUDA's extended provision accommodates a claimant's allegation of proxy discrimination. For instance, an argument can be made that even in the absence of race markers from training data sets, systems still have the capacity to yield unfavourable predictions using alternative markers that correlate to the protected trait; for example, postal addresses that are usually racialised.[116] This correlation could perpetuate systematic disadvantage on the aspirations of disenfranchised candidates.[117] It could further be argued that proxy postal address discrimination would adversely affect the right of the applicant to secure employment,[118] causing the deterioration of equal enjoyment of the rights to human dignity and equality in a similar manner to how deprivation based on race would affect the candidate.[119] One can see how an applicant would have difficulty presenting evidence that their rejection was based on proxy postal address discrimination.[120]

In both cases of direct and proxy discrimination, ascertaining the logic utilised by a system would be difficult without access to advanced and costly computer specialists.[121] Indirect discrimination adds a layer of complexity to this onus, as it is an offense that is often hidden or concealed.[122] Although the scope of PEPUDA is wide, expanding on the prohibited grounds, the onerous evidentiary burden renders PEPUDA an unsuitable legal avenue for combatting algorithmic discrimination.

---

[114] F.Z Borgesius op cit note 22 at 10.
[115] PEPUDA 73 at s1. (see footnote 102 for the prohibited grounds)
[116] P.T Kim and M.T Bodie op cit note 42 at 290.
[117] PEPUDA s1(xxii)(b)(i) and (ii).
[118] M Bogen and A Rieke op cit note 24 at 3. The author contends that predictive tools can be a power factor gatekeeper when recruiters decide who will be eligible for employment.
[119] PEPUDA s1(xxii)(b)(iii).
[120] F.Z Borgesius op cit note 22 at 19.
[121] C Abungu 'Algorithmic decision-making and discrimination in developing countries' (2022) *Journal of Law, Technology and the Internet* at 49.
[122] F.Z Borgesius op cit note 22 at 18.

### 3.3 The Protection of Personal Information Act (POPIA)

The legislature enacted POPIA to provide context and dimension to the right to privacy contained in s14 of the Constitution.[123] The provisions of POPIA outline principles for the lawful processing of personal information.[124] These principles include 'accountability, processing limitation, purpose specification, further processing limitation, information quality, openness, security safeguards and data subject participation.'[125] POPIA is relevant to algorithmic discrimination owing to the processing of personal data during digital recruitment.[126]

An organisation processing personal data must do so on a lawful basis.[127] There are three potential justifications for the partial processing of data in hiring.[128] An employee can rely on legitimate interest, contractual obligations or consent to the processing.[129] A legitimate interest can be used where the data controller has informed the candidate of the reasons for processing their data, and such reason is related to the functions of the controller.[130] Moreover, the data subject must be provided with the right to object to the processing.[131] Candidates are permitted to object to processing their data based on personal concerns if, for example, they feel processing could lead to discriminatory outcomes.[132] Once the applicant objects, processing will cease unless an organisation can establish a legislative provision which empowers it to continue.[133] Provided an organisation stipulates a right to object, legitimate interest is a viable justification for partially automated hiring.[134] Legitimate interests can include efficiency, high application volumes and budgetary constraints.[135]

Processing could also be permissible if it is 'necessary to carry out the actions for the conclusion or performance of a contract to which the data subject is a party.[136] The General

[123] POPIA (Preamble)
[124] Papadopolous & Snail ka Mtuze (eds) 'Cyber Law @ SA: The Law of the Internet in South Africa' 4e (Van Schaik 2022) at 504.
[125] POPIA s8.
[126] F.Z Borgesius op cit note 22 at 18.
[127] A Kelly – Lyth op cit note 12 at 909. See POPIA at s13(1).
[128] Ibid at 909.
[129] Protection of Personal Information Act at s11(1)(f) A legitimate interest could be that automated processing is the only means of considering all applications submitted.
[130] Ibid at s11(1)(f).
[131] A Kelly – Lyth op cit note 12 at 910. See POPIA s11(3)(a).
[132] Ibid at 910. See POPIA s11(3)(a).
[133] Ibid at s11(3)(a).
[134] A Kelly – Lyth op cit note 12 at 910.
[135] Ibid.
[136] Ibid.

Data Protection Regulation (GDPR) contains a similar provision that the European Data Protection Board has clarified. According to the board, for a data controller to rely on this provision. A data controller would have to demonstrate that processing occurs within the confines of a valid contract and that processing is necessary for the performance of obligations contained therein.[137] In SA, the relationship between applicant and employer does not give rise to any contractual obligations, therefore an employer would not be able to rely on this ground.[138]

The final justification for processing is consent obtained from the applicant.[139] The imbalance in the applicant-potential employer relationship renders this avenue unviable, as consent would unlikely be given voluntarily.[140] The inequity between the data subject and controller, in this case, the potential employer, does not lead to voluntary consent.[141] As the data subjects' withheld consent may result in unfavourable consequences, such as the inability to secure the potential job. [142]

S26 of POPIA prohibits the processing of special category information.[143] This information concerns 'the religious or philosophical beliefs, race or ethnic origin, trade union membership, political persuasion, health or sex life or biometric information'[144] or certain criminal history of the data subject.[145] Most employers using automated hiring tend to avoid processing sensitive information, fearing it will perpetuate bias or infringe on data protection laws.[146] Algorithmic recruitment presents two issues. First, recruiters may need to include sensitive information in training data, as this would likely mitigate and substantially reduce bias ingrained in previous hiring decisions.[147] This universal prohibition generates challenges for evaluating and vindicating algorithmic discrimination.[148] Researchers have found that

[137] European Data Protection Board 'Guidelines 2/2019 on the processing of personal data under Art 6 (1)(b) GDPR in the context of the provision of online services to data subjects' available at https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf accessed on 13th January 2024.
[138] There is no offer and acceptance.
[139] Ibid.
[140] Ibid.
[141] A Kelly – Lyth op cit note 12 at 911.
[142] Ibid.
[143] POPIA at s26(a) or Art 9 of GDPR.
[144] Ibid at s26(b)
[145] Ibid.
[146] Ibid. Scholars have identified a contradiction between conformity with data protection law and the collection of sensitive data to mitigate algorithmic discrimination.
[147] Ibid.
[148] F.Z Borgesius op cit note 22 at 24.

'blinding' an AI tool to sensitive information may preclude it from identifying and mitigating previously perpetuated bias.[149] Consequently, a recruiter wishing to include special information in their automated hiring procedure would find it difficult to trace a lawful basis for doing so.[150] As previously explained, the consent justification would be difficult to deploy given the asymmetries in the applicant-potential employer relationship.[151] Similarly, recruiters would not be able to claim the necessity of processing for 'the establishment, exercise or defense of a right or obligation in law' just because it would likely reduce discrimination.[152] The second issue pertains to the lack of clarity on whether data is considered sensitive if the AI identifies several proxies for special characters.[153] An AI tool could conceivably be able to deduce racial identity (in cases of video interviews) from a combination of traits *inter alia*, the pronunciation of words, speech patterns, and postal addresses of schools attended.[154] Whether this is considered sensitive information will entirely be contingent on how accurate the inference is and whether it is being sought intentionally.[155] Having considered the general rules applicable to the partially automated processing of personal information, I will now consider a provision that applies exclusively to fully automated decisions.

POPIA applies to automated and non-automated processing of personal data by a responsible party who is domiciled in SA or abroad, provided the processing takes place locally.[156] POPIA does not provide an exact definition of automated decision-making, though it is widely accepted that it would be inclusive of decisions taken by predictive hiring tools.[157] POPIA prohibits entirely automated decisions, stating that a data subject:

> 'may not be subject to a decision which results in legal consequences for him, her or it or which affects him, her or it to a substantial degree, which is solely based on automated processing of personal information intended to provide a profile for such person, including his or her

[149] A Kelly – Lyth op cit note 12 at 912.
[150] Ibid.
[151] Ibid. See s27(1)(a).
[152] POPIA s27(11)(b).
[153] A Kelly – Lyth op cit note 12 at 912.
[154]Ibid.
[155] Ibid at 913.
[156] POPIA supra note 91 at s3(1)(a) Processing means 'any operation or activity or any set of operation, whether or not automatic means, concerning personal information, including – (a) the collection, receipt, recording, organization, collation, storage, updating or modification, retrieval, alteration, consultation or use; (b) dissemination by means of transmission, distribution or making available in any other form; or (c) merging linking, as well as restriction, degradation, erasure or destruction of information.'
[157]V Bronstein 'Prioritising command and control over collaborative governance: The role of The Information Regulator under the Protection of Personal Information Act' (2022) PER/PELJ at 18.

performance at work, or his or her or its creditworthiness, reliability, location, health, personal preferences or conduct.'[158]

S71(1) of POPIA prohibits entirely automated processing of personal data by algorithms if such processing has legal ramifications or will affect the data subject in the listed circumstances.[159] However, there is uncertainty regarding the applicability of this general provision to automated decisions in the recruitment of employees. A compelling argument can be made that AI discrimination does tend to have legal consequences on candidates by undermining their fundamental rights either directly or through proxies.[160] Similarly, candidates could postulate that algorithmic discrimination would substantially affect their ability to secure employment, as the algorithms now act as gatekeepers to employment.[161] The word 'including' in s71(1) indicates that the legislature intended an open list of protected grounds.[162] However, POPIA's failure to overtly include recruitment as a prohibited circumstance may indicate its unsuitability to incidents of AI discrimination in recruitment.

S71(2)(a) of POPIA contains exceptions to the general prohibition. The provision states that solely automated decisions are permissible if procured concerning the finalisation or performance of terms of an agreement, and the supplications of the applicant have been fulfilled as per the terms of an agreement or if appropriate measures have been taken to guard the applicant's recognised interests.[163] Due to the absence of any contractual obligations between an applicant and a recruiter, this provision would not be useful to an employer attempting to escape liability for discrimination caused by a fully automated selection process in hiring. In the absence of judicial scrutiny, the proper application of POPIA remains uncertain. POPIA was enacted with provisions inspired by the GDPR. S71(2) of POPIA is similar to Art 22(1) of the GDPR. Parviainen contends that the provision can be interpreted as either an entitlement not to be subjected to solely automated decisions or as a prohibition of wholly automated decisions.[164] If Art 22 (1) confers a right, it permits the use of systems which generate automated decisions that inform recruitment provided the data subject elects to exercise their

[158] POPIA supra note 91 at s71(1) and Art 22 of GDPR (*kafka* provision).
[159] V Naidoo 'POPI and its possible relation to AI or machine learning systems' (2021) *Without Prejudice* at 68.
[160] B Kammerer 'Hired by a robot: The legal implications of AI video interviews and advocating for greater protection of job applicants' (2022) *IOWA Law Review* at 826.
[161] Ibid.
[162] POPIA at s71(2).
[163] POPIA supra note 91 at s71(2)(a)(i) and (ii).
[164] H Parviainen 'Can algorithmic recruitment systems lawfully utilise automated decision-making in the EU?' (2022) *European Labour Law Journal* at 229. General provision in GDPR is contained in Article (Art) 22(1).

right.[165] If an applicant refuses wholly automated processing, the firm would have to consider whether the processing falls within the ambit of any exceptions which are contained in Art 22(2).[166] If processing would not qualify as an exception, then the firm would be obligated to limit automated processing to candidates who chose to exercise their entitlement.[167] Understood as a prohibition, Art 22(1) would outlaw automated decision-making entirely, unless the data controller can prove that the processing is considered an exception and they have provided the necessary safeguards to protect the data subject's interests, freedoms and rights.[168] If the decision does not satisfy an exception or the data controller does not provide safeguards, then wholly automated processing would not be permissible, hence HR would have to manually review the application.[169]

A point of contention regarding the appropriateness of POPIA as a legal remedy to AI discrimination is the burden of proof.[170] Plaintiffs alleging discrimination perpetrated purely via automated processes would have to explain that discrimination had taken place.[171] This would require the individual to explicate the logic embedded in the system to reach a discriminatory outcome.[172] As previously mentioned, attempting to explain the logic of automated decisions can be difficult, especially concerning fully automated systems that can completely deviate from the original intended outcomes.[173] Although POPIA provides for the right to contend a decision and even have the decision explained,[174] these provisions only come into operation contingent upon proving that the processing was rendered in terms of an exception in s71(2).

Internationally, it has been noted that there is a compliance and enforcement deficit regarding the provisions of data protection laws.[175] Scholars note that the Information Regulator is likely not equipped to investigate and circumvent the evident risks associated with

[165] Ibid.
[166] Ibid. See s71(2)(a)(i-ii) of POPIA.
[167] Ibid.
[168] Ibid at 230.
[169] Ibid.
[170] F.Z Borgesius op cit note 22 at 24.
[171] Ibid.
[172] Ibid.
[173] M Bogen and A Rieke op cit note 24 at 8.
[174] POPIA s71(3)(b).
[175] F.Z Borgesius op cit note 22 at 24.

AI deployment.[176] The exceedingly technical nature of AI is likely to expose knowledge gaps in similar institutions, which could make enforcing the laws difficult.[177] The use of open-ended and abstract diction in data protection law has been identified as a hurdle to its application and thus the regulation of predictive AI.[178] For instance, the wording employed in the general prohibition of automated decisions in s71. The provision is extremely broad and uncertain. While the open-ended nature of the list of included prohibited acts allows for wider application, it is also likely to create legal ambiguity for firms wishing to invest and use AI hiring tools. These limitations point to POPIA's inefficiencies in providing the necessary protection to applicants.

### 3.4 The Employment Equity Act (EEA)

The EEA was promulgated to ensure that employers are allowed equal opportunities by mitigating unfair differentiation in workplace policies and practices.[179] S6(1) of the EEA contains a general prohibition on all unfair discrimination, stating that

> 'No person may unfairly discriminate, directly or indirectly, against an employee, in an employment policy or practice, on one or more grounds including race, gender, sex, pregnancy, marital status, family responsibility, ethnic or social origin, colour, sexual orientation, age, disability, religion, HIV status, conscience, belief, political opinion, culture, language [and] birth or any other arbitrary ground.'[180]

s9 of the EEA further stipulates that when considering s6, the term 'employee' is inclusive of applicants.[181] Thus, s6 read in conjunction with s9 prohibits unfair discrimination against employees and applicants.[182] The EEA permits an applicant legal recourse provided they can prove AI bias in a hiring practice. Despite the utility in the wide definitional ambit of the EEA, which includes applicants as employees, there are limitations in the protection it can offer ordinary South African jobseekers. As with the previously discussed legal frameworks,

[176] V Bronstein op cit note 157 at 11. The Information Regulator is an independent institution which was established by POPIA. It is the body that oversees the compliance and enforcement of the provisions of POPIA and the Promotion of Access to Information Act (PAIA).
[177] Ibid.
[178] Ibid.
[179] EEAA at s5.
[180] Ibid at s6(1) The statute further states that affirmative action programs which align with the purpose of the act do not constitute unfair discrimination.
[181] Ibid at s9.
[182] Ibid at s9 and s6.

the burden of proof set by the EEA would likely render it difficult for a claim of AI bias to be successful.[183] S11 of the EEA dictates that

> 'if unfair discrimination is alleged on a ground listed in 6(1), the employer against whom the allegation is made must prove on a balance of probabilities that such discrimination – (a) did not take place or (b) is rational and not unfair or is otherwise justifiable.'[184]

Contentions based on s6(1) require that the candidate first make arguments establishing the alleged discrimination, which the firm must then disprove by showing that it did not take place or that it was justifiable.[185] The EEA further states that if the discrimination is based on an arbitrary ground, the applicant, on a balance of probabilities, must convey that the action complained of is not rational, amounts to discrimination and that it is unfair.[186] Relying on either provision would require the applicant to argue that discrimination had occurred. The latter provision extends the burden by requiring the applicant to prove that the discrimination was not rational and amounted to unfair discrimination. The opaque nature of decisions taken by an AI predictive tool renders it difficult to establish the occurrence of algorithmic discrimination in the first place, let alone proving the absence of a legitimate rationale behind it.[187] Without expensive tools and experts, deciphering the AI data building blocks and its decisions would be difficult and costly.[188] Considering the destitute economic reality many South Africans face, job seekers are not in a fiscal position to obtain technical experts to assist in proving a case of algorithmic discrimination.[189] The onerous burden of proof thus limits the applicability of the EEA to algorithmic discrimination in hiring.[190]

As illustrated above, the recourse provided in South African statutes is restricted, and more nuanced legal protections are required to combat the discriminatory side effects of AI's use in recruitment. The following sections will consider the provisions of the Artificial Intelligence Act, specifically those which speak to the circumvention of algorithmic discrimination in recruitment. The aim is to expose the availability of a legal framework for AI regulation and how SA can learn from it to provide candidates with sufficient legal protection while continuing to allow space for innovation.

[183] F.Z Borgesius op cit note 22 at 24.
[184] EEAA at 11(1)(a-b).
[185] Ibid at s11.
[186] Ibid at s11(2)(a-c).
[187] F.Z Borgesius op cit note 22 at 10.
[188] Ibid 33 at 24.
[189] F.Z Borgesius op cit note 22 at 24.
[190] Ibid.

## 4. THE ARTIFICIAL INTELLIGENCE ACT OF THE EUROPEAN UNION (AI Act)

In April 2021, the EU Commission introduced a proposal statute to regulate AI.[191] The Bill is praised as the first comprehensive legal instrument that tackles AI regulation through a risk-based approach (RBA).[192] The Commission's initial proposal was considered by the Council and Parliament of the EU, the legislative branches responsible for enacting laws through inter-organisational negotiations.[193] The Council and Parliament published their approaches in response to the initial proposal. At the time of writing this contribution, a prefinal AI Act agreed to by the Council and Commission was leaked online. Below, I will provide summaries of all four drafts of the proposal. The paragraph outlining the Commission's proposal will elaborate on certain criticisms offered by academics, allowing the reader to understand the context of its later revisions. Finally, I will delineate the prefinal politically agreed Bill, specifically concentrating on elements relevant to recruitment tools that are classified as high-risk. The suggested recommendations for South Africa will be crafted based on a holistic view of the EU's law-making processes, particularly the contentious hurdles each body faced in reaching the ultimate political agreement.

The general objective of the Commission's proposal was to 'ensure the proper functioning of the single market by creating the conditions for the development and use of trustworthy AI systems.'[194] To this end, the Act adopted a technologically neutral definition of AI and an RBA.[195] The Act defines AI as software developed with techniques that employ DL and varieties of ML, including supervised, unsupervised and reinforcement learning.[196] The neutrality of the definition came under scrutiny due to its incorporation of the term 'software', which is perceived as indicating a particular technology, thus undermining the assertion of neutrality and potential adaptability in the future.[197] Due to the broad and flexible nature of the

[191] T Madiega 'Artificial Intelligence Act' 2021 *The European Parliamentary Research Service* at 3 last accessed from https://s899a9742c3d83292.jimcontent.com/download/version/1643294959/module/8356404363/name/EUanalysis-AIA-%28Nov2021%29.pdf on 7 June 2023.

[192] L Edwards 'Regulating AI in Europe: four problems and four solutions' (2022) *Ada Lovelace Institute* at 9.

[193] I McGowan, A Allen et al 'CDT Europe's AI Bulletin June 2023' available at https://cdt.org/insights/cdt-europes-ai-bulletin-june-2023/ accessed on 23rd of January 2024.

[194] T Madiega op cit note 191 at 4. Reinforcement learning refers to software that learns by interacting with its environment and providing responses for the actions it takes.

[195] Ibid at 4.

[196] AI Act Annex 1(a) The Act further includes techniques which involve 'logic and knowledge-based approaches including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems. Included are also statical approaches, Bayesian estimation, search and optimization methods.'

[197] T Madiega op cit note 191at 9.

definitional provisions, there were concerns regarding legal clarity regarding what constitutes AI.[198] This ambiguity may have dissuaded potential investments in AI or inadvertently sustained adverse effects on fundamental rights.[199] The Act categorises AI systems according to the risk they pose to the fundamental rights and the health and safety of natural persons.[200] There are four risk categories namely, unacceptable, high, limited, and minimal risk.[201] AI systems that pose an unacceptable risk are prohibited, while those posing a high risk are subject to legislative regulation.[202] Limited-risk systems must adhere to stipulated transparency obligations, while low or minimal-risk systems are not subject to any obligations.[203] The delineation of systems in this manner hinges on the use as envisaged by the technological innovators.[204]

The proposal automatically classifies predictive hiring systems as high-risk, including those used to select suitable applicants.[205] This categorisation as high-risk is justified by the acknowledgement that recruitment systems of this nature can potentially perpetuate discrimination.[206] The AI Act requires all high-risk AI systems to adhere to the requirements laid out in Chapter 2 of the Act and are subject to *ex-ante* conformity assessments.[207] It is also recommended that the intended purpose of the AI system and a risk management system referred to in Art 9 be considered when assessing compliance with the provisions.[208] Art 9 of the Act stipulates that a risk management protocol be established and maintained throughout the life cycle of all high-risk systems.[209] The generic application of the Commission's RBA has been branded as 'illusory and arbitrary'.[210] It has been claimed that the approach failed to acknowledge the true nature of AI applications as products whose impact is continuously

[198] Ibid.
[199] Ibid at 4. There is still a reconciliation between effective regulation and the possibility of continued innovation.
[200] T Madiega op cit note 191 at 5. See AI Act at 5.2.3.
[201] Ibid.
[202] Ibid at 5. Title III, Art 6 defines high-risk systems as those which create adverse consequences on individual safety or affect the fundamental rights of users. High-risk systems would include biometric identification and categorisation of individuals, employment AI systems which govern labour management and access to self-employment.
[203] Ibid at 5.
[204] N Helberger & N Diakopoulos 'ChatGPT and the AI Act' (2023) *Internet Policy Review* at 3.
[205] O Stockham and C Fourcans 'Discrimination in Hiring: The Case for Alignment of the EU Act with EU Equality Laws' (2023) last accessed from *https://cdt.org/insights/discrimination-in-hiring-the-case-for-alignment-of-the-eu-ai-act-with-eu-equality-laws/* See Annex III 4(a) on 7 July 2023.
op cite note. See Art 6(2) and Annex III.
[206] Ibid.
[207] AI Act at Art 8(1). Also see AI Act at 5.2.3.
[208] Ibid at Art 8(2).
[209] AI Act at Art 9(1).
[210] L Edwards op cit note 192 at 11.

evolving based on various groups and individuals who use it throughout their lifecycles.[211] The RBA categorisation of AI in this manner has also been critiqued as being static, as it did not account for the flexibility of a system's impact in diverse circumstances.[212] Its significant emphasis on legal technicalities rather than the potential impact on fundamental rights was also highlighted.[213] Scholars were weary that the proposal made no provisions for user participation, human rights impact assessments, and the right to legal remedies in the event of non-compliance.[214] The identified pitfalls, among others, prompted a response from the Council on 6 June 2022.[215]

The EU Council proposal defines AI as systems that function autonomously, utilising both machine and/or human-supplied data and inputs to deduce methods for accomplishing specified objectives through machine learning and/or logic and knowledge-based methods.[216] It generates outputs, such as content (in generative AI systems), predictions, recommendations, or decisions, thereby impacting the environments in which the AI system operates.[217] The distinction of this definition from that of the Commission lies in its omission of reference to specific technologies. This omission strengthens assertions of technological neutrality and potential adaptability for future advancements in AI.

The Council's proposal made major amendments to the classification of high-risk applications. The Commission's proposal automatically designates any AI application in the areas listed in Annex III as high-risk.[218] The Council asserts that Annex III applications will only be categorised as high-risk if their output or recommendations are given substantial weight in coming to a final decision.[219] For instance, an employment system whose hiring recommendations are directional rather than decisive would not be classified as a high-risk, despite being listed in Annex III.[220] The issue with this addendum is its evidentiary restriction. A victim who may want to allege discrimination would have no way of telling if the system was decisive and subject to the provisions governing high-risk applications or if it was

[211] G Finocchiaro 'The regulation of AI' (2023) *AI & Society* at 5.
[212] Ibid.
[213] Ibid at 11. This is further deduced from the lack of individual participation provided by the Act.
[214] Ibid.
[215] I McGowan, Asha Allen et al 193.
[216] Council AI Act Art 3(1).
[217] Ibid.
[218] AI Act (Commissions) Art 6(2).
[219] I McGowan, Asha Allen et al 193.
[220] Ibid.

employed as an accessory to inform the final decision taken by a human recruiter, which negates categorisation as high-risk and thus the need to comply with the requirements of the proposal.

In June of 2023, the EU Parliament published its final proposal.[221] To contrast the definitions above, Parliament defines AI as

> 'A machine-based system that is designed to operate with varying levels of autonomy that can for explicit or implicit objectives, generate outputs such as predictions, recommendations or decisions, that influence physical or virtual environments.'

The parliamentary definition also excludes the word software and concisely summarises what AI systems generally do.[222] Scholars posit that framing AI in this manner is advantageous as it distinguishes AI from simpler software models, thereby avoiding excessive regulation that could potentially stifle future innovation.[223] The exclusion of technologically specific diction ensures that the definition will be compatible with the plethora of industries it intends to regulate.[224] This new definition provides innovators with clarity on what types of systems are considered AI and their functions.[225] This will ensure that innovators can pre-empt the need to comply with the AI Act, easing the anxiety they may have around future innovation.

The draft adds a determining layer for high-risk classification, which refines the RBA, circumventing the critiques that the original RBA was illusionary and arbitrary.[226] According to the parliamentary draft, a high-risk system will only be high-risk if it falls within the industries in Annex III and it poses a 'significant risk of harm under the high-risk categorisation rules in Art 6(2)'.[227] The significant risk test considers 'severity, intensity, probability of occurrence and duration of its effects and its ability to affect an individual, a plurality of persons or to affect a particular group of persons.'[228] Parliament also created a provision for

[221] Ibid.
[222] K Sheriff, K.C Ham and J.D Seiver 'European Parliament Approves Amendments to Expand the Scope of the EU AI Act' last accessed from https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/06/european-union-ai-amendments-approved on 16 February 2024.
[223] Ibid.
[224] Ibid.
[225] Ibid.
[226] Ibid.
[227] Ibid.
[228] Parliament AI Act at Art 3(1b).

fundamental impact assessments of systems, which was notably absent from the Commission's proposal. The impact assessments include

> 'identifying the category of persons likely to be affected by it [the system]; assessing the system's impact on fundamental rights as they are defined in the EU Charter, particularly potential harms to marginalised or vulnerable communities; and drafting a plan for mitigating risk posed by the system.'[229]

The proposal further provides impacted individuals the right to an explanation of the decisions that impact socio-economic welfare, fundamental rights, health or safety.[230] Parliament also integrated a right to legal recourse for victims of discriminatory or unsafe high-risk AI systems.[231] Recital 84(a) states that natural and juristic persons or communities of individuals should be able to lodge complaints of noncompliance with the Bill by a deployer to 'their national supervisory authority.'[232]

On 9 December 2023, a political agreement on the AI Act was successfully negotiated between the Council and Parliament,[233] and the pre-final AI Act was leaked to the public on 22 January 2024.[234] Although there remains uncertainty around the technical provisions of the Act, consensus has been reached regarding the regulatory framework.[235] The pre-final draft's definition of AI is similar to the one contained in the parliamentary draft.[236] The statute retains the RBA and its four categories with additions that regulate general AI systems such as ChatGPT.[237] For high-risk applications, classification as high-risk does not only hinge on appearance in Annex III, but the system will also have to pass the significant harm test. The harm test mentioned above excludes AI systems, which 'perform narrow procedural tasks, improving the results of previously completed human activities, detecting decision-making

[229] K Sheriff, K.C Ham and J.D Seiver op cit note 222.
[230] Ibid. Art 68c of Parliament Draft.
[231] Parliament AI Act Art 84a The proposal first implores that deployers provide for an internal procedure for groups of natural person and legal persons to submit complaints, if this is an option.
[232] Ibid.
[233] I McGowan, A Allen and R Ceraulo 'CDT Europe's AI Bulletin December 2023' last accessed at *https://cdt.org/insights/cdt-europes-ai-bulletin-december-2023/* on 5 January 2024.
[234] C Hainsdorf el at 'The pre-final text of the EU's AI Act leaked online' last accessed at https://www.whitecase.com/insight-alert/pre-final-text-eus-ai-act-leaked-online#:~:text=On%20January%2022%2C%202024%2C%20the,followed%20by%20a%20258%2Dpage on 4 February 2024.
[235] I McGowan, A Allen et al op cit note 233.
[236] See Amendment 165 of Parliament Draft & Art 3(1) of pre-final text.
[237] Ibid.

patterns without replacing human assessment or mere preparatory tasks to risk.'[238] In the context of employment, the categorisation of AI tools as high-risk would vary according to their function and influence in a final decision. For instance, if a system which factors in protected characteristics is the sole decider of whether to offer a candidate employment, the system would be considered high-risk.[239] Such a system could potentially discriminate both directly and indirectly. Alternatively, if the recruitment system is used only to verify the educational authenticity of the candidate, then the system would not be considered high-risk.

According to this draft, high-risk systems would still have to comply with Chapter 2, which has largely remained unchanged. However, in the original statute, Art 19 required that providers of AI tools should ensure that their systems undergo a self-conformity assessment before entering the market.[240] Yet, in the prefinal proposal, Art 19 has been omitted. I believe the omission was based on the quarrelsomeness of *ex-ante* self-conformity assessments. Scholars argued that the previous Act unduly relied on the self-conformity processes that are vulnerable to abuse or manipulation by deployers.[241] For instance, the self-conformity process allows producers to misrepresent high-risk systems as low or minimal risk, that are subject to little or no regulatory standards.[242] At the time of writing this contribution, it was unclear if the original provision or an amended version would be included in the final draft.

The issue of the burden of proof is averted by the proposal's right to receive an explanation of the decision. The text provides that subjects of high-risk systems (recruitment AI) have the right to have the decision explained if such decision has legal ramifications or 'similarly significantly affects him or her in a way that they consider to adversely impact their health, safety or fundamental rights.'[243] These explanations could potentially provide evidence of the occurrence of discrimination. Additionally, the new text grants users the right to file a complaint with a market surveillance authority.[244] The complaints can be lodged by either a natural or juristic person with evidence of an infringement of the provisions.[245] The text's recommitment to its original human rights agenda now includes provisions for fundamental

[238] C Hainsdorf el at op cit not 234.
[239] Art 8 (1) Pre- Final draft.
[240] Art 19(1) Commission proposal.
[241] T Madiega op cit note 191 at 8.
[242] Ibid.
[243] Ibid at Art 68(c).
[244] Ibid at Art 68(a).
[245] F.Z Borgesius op cit note 22 at 10.

rights assessments for high-risk systems before they are introduced to the market.[246] This provision is, however, only limited to deployers who are governed by public law or private operators providing services for public bodies using systems mentioned in Annex III.

The prefinal AI Act offers a high level of protection against algorithmic discrimination. The inclusion of fundamental human rights impact assessments, the right to an explanation and the right to judicial submissions exhibit the text's pledge to the substantive protection of human rights and the deterrence of discrimination.

## 5. LESSONS FOR SOUTH AFRICA

The first hurdle SA faces in regulating AI is deciding on a horizontal or vertical approach to regulation. The EU opted for a horizontal approach, which is essentially a one-size-fits-all.[247] The issue with an undifferentiated approach is the inflexibility of the law, which could translate to a disconnect between the statute's ability to regulate certain industries.[248] Academics have argued that AI, being such a dynamic technology, requires laws of a similar nature, which may be industry or sector-specific.[249] This fragmented approach would entail differing legislations for each industry.[250] These instruments would consist of specified definitions of AI according to the differing uses in each sector, easing the difficulty of defining AI.[251] On the contrary, it has also been observed that a horizontal approach to regulation would be cost and time-effective.[252] A general approach would avoid fragmentation of enforcement, enabling consistent regulation throughout a country or region.[253] Chapman further points out that a centralised or general approach would encourage public and political participation through effective management.[254]

Despite critiques of the centralised approach, in the context of SA, where AI literacy is still relatively nascent, a centralised approach appears to be the only feasible option.[255] It will enable the pooling of scarce AI experts who can be tasked to draft a unified legal instrument on the regulation of AI. Once the country has been upskilled, a fragmented approach may

[246] Ibid at Art 29(a).
[247] G Finocchiaro op cit note 211 at 4.
[248] Ibid.
[249] Ibid.
[250] Ibid.
[251] Ibid.
[252] D Chapman 'The ideal approach to Artificial Intelligence legislation: A combination of the United States and European Union' (2023) *University of Miami Law Review* at 290.
[253] Ibid.
[254] Ibid.
[255] V Bronstein op cit note 157 at 11.

become a viable reinforcement to established general regulation. South Africa's strong civil society organisations will be able to participate effectively throughout the drafting of a general approach, ensuring that the voices of users are amplified and ultimately inform the provisions of a uniquely South African AI regulation.[256]

SA will have to grapple with constructing its own RBA. The addition of the significant harm test in the EU AI Act allows for a fair and nuanced approach to the categorisation of high-risk applications and should be retained by SA. The right to receive explanations for decisions which affect rights should also be included to allow transparency and resolve the issue of the burden of proof. However, the provision on self-assessments can be altered to include third-party, state-funded audits. State-funded audits will ensure that producers of AI tools consistently maintain genuine compliance with the provisions at minimal or no direct extra cost.[257] The digital divide in SA has resulted in a situation whereby the 'have nots' are vulnerable to digital exclusion and the occurrence of discrimination online.[258] To avoid these undesirable consequences, I suggest that fundamental rights assessment should not be limited to deployers governed by public law but include all high-risk AI applications regardless of the deploying entity or person. Conducting rights assessments as a prerequisite will ensure the mitigation of effects caused by discriminatory AI tools that exacerbate the pre-existing economic disparities as SA journeys into the Fourth Industrial Revolution.

## 6. CONCLUSION

In part one of this discussion, I highlighted the perforation of AI into the mainstream, noting that it has found its way into most industries. I then considered the role of AI in recruitment and how, at times, its operation could lead to discriminatory outcomes. Part two grappled with the difficulty of assigning a precise definition of AI because of the numerous typologies elaborated by scholars. Part three ventured into the limited legislative protection against AI bias offered by South African law. Due to the opaque nature of the operation of AI tools, the overall limitation was identified as the burden of proving the AI bias had taken place and on what grounds. Part four expanded on the solutions at the international level, specifically in the four drafts of the EU AI Act. The recommendations provided were based on an overview of

[256] Ibid.
[257] G Finocchiaro op cit note 211 at 4. This suggestion is based on the claim the EU AI Act imposes considerable administrative costs in its implementation.
[258] OO Olaitan et al 'A framework to test South Africa's readiness for the fourth industrial revolution' 2021 *South African Journal of Information Management* at 3.

the three drafts of the Act, with a specific focus on the pre-final draft which could be subject to amendment later this year. Overall, I advanced that SA AI regulation should be based on an understanding of its unique digital context, which generally requires greater protection of users.

## 7. BIBLIOGRAPHY

### Books

Papadopolous & Snail ka Mtuze (eds) 'Cyber Law @ SA: *The Law of the Internet in South Africa'* 4e (Van Schaik 2022).

**Case Law**

*Bernstein and Others v Bester (CCT23/95) [1996] ZACC 2; 1996 (4) BCLR 449; 1996 (2) SA 751.*

*President of the Republic of South Africa v Hugo* 1997 (4) SA 1.

**Chapters in Books**

Arakpogun EO et al 'Artificial Intelligence in Africa: Challenges and opportunities' in Allam Hamdan et al '*The Fourth Industrial Revolution Implementation of Artificial Intelligence for Growing Business Success.*' (2021).

De Vos P 'Equality, human dignity and privacy rights' in D. Brand et al '*South African Constitutional Law in Context*' (2014) Oxford University Press.

**Journal Articles**

Abungu C 'Algorithmic decision-making and discrimination in developing countries' (2022) *Journal of Law, Technology and the Internet.*

Adams-Prassl J 'What if your boss was an algorithm? Economic incentives, legal challenges and the Rise of Artificial Intelligence at work' (2019) *Comparative Labour Law & Policy Journal.*

Benbya H & Davenport T.H & Pachidi S 'Special Issue Editorial. Artificial Intelligence in organisations: Current and future opportunities' (2020) *MIS Quarterly Executive* Vol 19.

Borgesius F.Z 'Discrimination, Artificial Intelligence and Algorithmic Decision Making' (2018) *Council of Europe Directorate General of Democracy.*

Bronstein V 'Prioritising command and control over collaborative governance: The role of The Information Regulator under the Protection of Personal Information Act' (2022) *PER/PELJ.*

Chapman D 'The ideal approach to Artificial Intelligence legislation: A combination of the United States and European Union' (2023) *University of Miami Law Review*.

Chaskalson A 'Dignity as a constitutional value: A South African perspective'(2011) *American University International Law Review*.

Finoccchiaro G 'The regulation of AI' (2023) *AI & Society*.

Gravett W.H 'Is the dawn of the lawyer robot upon us? The Fourth Industrial Revolution and the future of lawyers' (2020) *PELJ*.

Helberger N & Diakopoulos N 'ChatGPT and the AI Act' (2023) *Internet Policy Review*.

Kammerer B 'Hired by a robot: The legal implications of AI Video interviews and advocating for greater protection of job applicants' (2022) *IOWA Law Review*.

Kelly - Lyth A 'Challenging biased hiring algorithms' (2021) *Oxford Journal of Legal Studies*.

Kim P.T and Bodie M.T 'Artificial Intelligence and the challenges of workplace discrimination' (2021) *Journal of Labour & Employment*.

Lozada B.A 'The emerging technology of predictive analytics: Implications for homeland security' (2014) *Information Security Journal: A Global Perspective*.

Mhlanga D 'Artificial Intelligence in the industry of 4.0 and its impact on poverty, innovation, infrastructure development and the sustainable development goals: Lessons from emerging economies?' (2021) *UJ Content*.

Naidoo V 'POPI and its possible relation to AI or machine learning systems' (2021) *Without Prejudice*.

Olaitan O.O et al 'A framework to test South Africa's readiness for the fourth industrial revolution' (2021) *South African Journal of Information Management.*

Parviainen H 'Can algorithmic recruitment systems lawfully utilise automated decision-making in the EU?' (2022) *European Labour Law Journal*.

PR Palos-Sánchez & P Baena-Luna et al 'Artificial Intelligence and Human Resources Management: A Bibliometric Analysis' (2022) *Applied Artificial Intelligence.*

Rapanyane M.B & Sethole F.R 'The rise of AI and robots in the 4th Industrial Revolution: Implications for the future of South African job creation' (2020) *Contemporary Social Science*.

**Legislation**

AI Act of the European Union (Commission, Council, Parliamentary Drafts and Prefinal Draft).

Employment Equity Act 55 of 1998.

Promotion of Equality and Prevention of Unfair Discrimination Act 4 of 2000.

Protection of Personal Information Act 4 of 2013.

The Constitution of the Republic of South Africa,1996

**Legislative Guidelines**

European Data Protection Board's Guidelines 2/2019 on the processing of personal data under Art 6 (1)(b) GDPR in the context of the provision of online services to data subjects' https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf.

Edwards L 'Regulating AI in Europe: four problems and four solutions' (2022) *Ada Lovelace Institute*.

Madiega T 'Artificial Intelligence Act' (2021) *The European Parliamentary Research Service https://s899a9742c3d83292.jimcontent.com/download/version/1643294959/module/8356404363/name/EUanalysis-AIA-%28Nov2021%29.pdf.*

**Theses**

Mamela TL Assessment of the Impact of Artificial Intelligence on the Performance of Workforce at a South African Banking Institution (published LLM thesis, University of Johannesburg, 2021).

Adams N South African Company Law in the Fourth Industrial Revolution: Does Artificial Intelligence create a need for legal reform? (published LLM thesis, University of Western Cape 2021).

**Websites**

Bozic V 'AI and Predictive Analytics' (2023) available at *https://www.researchgate.net/publication/370074080_AI_and_Predictive_Analytics*.

Gumbi K 'South Africa’s unemployment rate drops marginally in second quarter' available at *https://www.reuters.com/world/africa/south-africas-unemployment-rate-down-slightly-326-q2-2023-08-15/.*

Hainsdorf C el at 'The pre-final text of the EU's AI Act leaked online' available at *https://www.whitecase.com/insight-alert/pre-final-text-eus-ai-act-leaked-online#:~:text=On%20January%2022%2C%202024%2C%20the,followed%20by%20a%20258%2Dpage.*

IBM Data and AI Team *'Blog AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What is the difference?' IBM Blog* available at *https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/.*

McGowan I, Allen A et al 'CDT Europe's AI Bulletin June 2023' available at *https://cdt.org/insights/cdt-europes-ai-bulletin-june-2023/.*

McGowan I, Allen A et al 'CDT Europe's AI Bulletin December 2023' available at *https://cdt.org/insights/cdt-europes-ai-bulletin-december-2023/.*

Oppenheim M *'Amazon scrapped 'sexist AI’ recruitment tool' The Independent 11 October 2018* available at *https://www.independent.co.uk/tech/amazon-ai-sexist-recruitment-tool-algorithm-a8579161.html.*

Sheriff K, Ham K.C and Seiver J.D 'European Parliament Approves Amendments to Expand the Scope of the EU AI Act' available at https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/06/european-union-ai-amendments-approved.

Stockham O and Fourcans C 'Discrimination in Hiring: The Case for Alignment of the EU Act with EU Equality Laws' (2023) available at https://cdt.org/insights/discrimination-in-hiring-the-case-for-alignment-of-the-eu-ai-act-with-eu-equality-laws/.