DOI: 10.1111/jols.12535

ORIGINAL ARTICLE

Algorithm-facilitated discrimination: a socio-legal study of the use by employers of artificial intelligence hiring systems

NATALIE SHEARD

Melbourne Law School, University of Melbourne, Carlton, Australia

Correspondence
Melbourne Law School, University of Melbourne, 185 Pelham Street, Carlton, Victoria, 3053, Australia.
Email: n.sheard@unimelb.edu.au

Abstract

Artificial intelligence (AI) hiring systems (AHSs) are used by employers every day to screen and shortlist job candidates. Despite this, substantial gaps exist in our understanding of the real – as opposed to theoretical – risks of discrimination when these systems are deployed. This article reports on findings from qualitative empirical research investigating the use of AHSs by Australian employers. It demonstrates that the way in which these systems are operated in practice creates serious risks of algorithm-facilitated discrimination. This may arise from the data, the use of proxies, the system's implementation, new structural barriers, a failure to provide reasonable adjustments, or the facilitation of intentional discrimination. These findings are significant, make an original contribution to an emerging field of research, and are relevant in any jurisdiction where an AHS is in use. There is a lot at stake when such discrimination occurs; as one research participant acknowledged, a 'job application is literally a person's attempt to change their life with a new job'.

1 INTRODUCTION

It is estimated that around 30 per cent of Australian organizations and 42 per cent of global companies use predictive[1] artificial intelligence (AI) systems[2] in recruitment, with these figures set to grow significantly in the next five years.[3] These systems employ machine-learning algorithms[4] and techniques, such as natural language processing (NLP),[5] to classify, rank, and score job applicants. They promise time and cost savings for employers, improved quality of hires, and a superior candidate experience.[6] But AI hiring systems (AHSs) may also enable, reinforce, and amplify discrimination against historically marginalized groups. They have been found to discriminate against applicants who wear a headscarf[7] or have a Black-sounding name,[8] and when the system is unable to accommodate requests for reasonable adjustments to enable access by people with disability.[9] In the most well-known example, an AI system developed by Amazon learned to downgrade the applications of job seekers who used the word 'women's' in their curricula vitae (CVs).[10]

Distinct and novel risks of discrimination emanate from the use by employers of AHSs. A discriminatory AHS can cause harm at unprecedented speed and scale,[11] and has the capacity – as one research participant, P4, explained – to 'systematically lock … [disadvantaged groups] out of the workforce'.[12] Throughout Organization for Economic Co-Operation and Development (OECD) countries, people who experience labour market marginalization tend to be, in disproportionate numbers, Indigenous peoples, racial and ethnic minorities, people with disability or health conditions, of lower socio-economic backgrounds, transgender, older (over 55 years), or

[1] Predictive AI systems are those designed to make predictions about future events. They are distinguished from generative AI systems, which are designed to create novel content.

[2] AI is defined as 'a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments': OECD, 'OECD AI Principles Overview' *OECD.AI*, at <https://oecd.ai/en/ai-principles>.

[3] A. Kaabel et al., *Inclusive Artificial Intelligence in Recruitment: From Cautious to Converted* (2022); C. Lytton, 'AI Hiring Tools May Be Filtering Out the Best Job Candidates' *BBC Worklife*, 16 February 2024, at <https://www.bbc.com/worklife/article/20240214-ai-recruiting-hiring-software-bias-discrimination>.

[4] Machine-learning algorithms are used to 'train' a computer program to automatically recognize patterns in a set of data.

[5] 'Natural language processing' refers to machine-learning techniques used to analyse written text or speech.

[6] For example, HireVue asserts that the use of its system has resulted in a 90 per cent decrease in time to hire and a 50 per cent decrease in cost per interview for some of its clients: <https://www.hirevue.com/>.

[7] E. Harlan and O. Schnuck, 'Objective or Biased: On the Questionable Use of Artificial Intelligence for Job Applications' *BR24*, 16 February 2021, at <https://interaktiv.br.de/ki-bewerbung/en/>.

[8] R. Young and S. McMahon, 'Name Discrimination Study Finds Lakisha and Jamal Still Less Likely to Get Hired than Emily and Greg' *WBUR*, 18 August 2021, at <https://www.wbur.org/hereandnow/2021/08/18/name-discrimination-jobs>.

[9] Australian Royal Commission into Violence, Abuse, Neglect and Exploitation of People with Disability, *Final Report: Inclusive Education, Employment and Housing (Volume 7)* (2023) Part B, 388.

[10] J. Dastin, 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women' *Reuters*, 11 October 2018, at <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.

[11] For example, Sapia (formerly PredictiveHire) asserts that it can interview 'thousands of candidates every week', more candidates than a human recruiter 'could consider in a lifetime': PredictiveHire, 'Science Explained' *PredictiveHire*, at <https://www.predictivehire.com/science-explained>.

[12] See also C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (2016) 105–122.

younger (under 25 years).[13] AHSs have also increased the knowledge and power asymmetries between employers and job seekers. Moreover, the predictions and outputs of these systems are difficult to contest, given that their use tends to be invisible and their processes opaque.

The right to work is a fundamental human right recognized in a range of international human rights instruments.[14] When people are denied this right, it significantly affects their economic opportunities and ability to make a 'better life' for themselves and their families. They are also denied the personal and social benefits of independence, a sense of purpose, and participation in society.[15] The use by employers of AHSs may also infringe other human rights, including an individual's right to dignity, autonomy, and privacy.[16] In addition, these systems may create 'representation' harms when they perpetuate the subordination of historically disadvantaged groups through, for example, stereotyping and stigmatization.[17]

While AHSs are used by or on behalf of employers around the world every day, little is known about how these socio-technical systems – comprised of people, organizations, and algorithms – operate in practice. Recent empirical studies have shed light on the determinants of AI adoption within organizations,[18] and on some of the dynamics at play at the individual and organizational levels when AHSs are used in daily workflows.[19] However, that research has not examined how AHSs are used by organizations as a means of identifying potential sources of bias or violations of equality rights.[20] Significant gaps therefore remain in our understanding of the real-world risks of discrimination when AI systems are deployed by employers. This article aims to begin the task of addressing these lacunae in the research by reporting the findings from one of the first empirical studies in this area. Drawing on qualitative semi-structured interviews with Australian recruiters, career consultants, and AI experts and developers, and a qualitative content analysis of publicly available industry material, it seeks to answer the question: does the way in which AHSs are used in practice by Australian employers create real risks of discrimination against marginalized groups?

---

13 See for example E. T. Achiume, *Racial Discrimination and Emerging Digital Technologies: A Human Rights Analysis: Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance* (2020), at <https://digitallibrary.un.org/record/3879751?v=pdf>; R. D'Almada-Remedios et al., *Inclusive Recruitment: How to Tap into Australia's Overlooked and Underleveraged Talent* (2022) 13.

14 See for example the International Covenant on Economic, Social and Cultural Rights, opened for signature 16 December 1966, 993 UNTS 3 (entered into force 3 January 1976).

15 Australian Human Rights Commission, *Willing to Work: National Inquiry into Employment Discrimination against Older Australians and Australians with Disability* (2016) 5, at <https://humanrights.gov.au/our-work/disability-rights/publications/willing-work-national-inquiry-employment-discrimination>.

16 See for example the International Covenant on Civil and Political Rights, opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976), Preamble and Art. 17.

17 S. Barocas et al., 'The Problem with Bias: From Allocative to Representational Harms in Machine Learning' in *Special Interest Group for Computing, Information and Society* (2017).

18 See for example J. Ochmann and S. Laumer, 'Fairness as a Determinant of AI Adoption in Recruiting: An Interview-Based Study' (2019), at <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1015&context=digit2019>; P Horodyski, 'Recruiter's Perception of Artificial Intelligence (AI)-Based Tools in Recruitment' (2023) 10 *Computers in Human Behavior Reports* 100298.

19 See for example L. Li et al., 'Algorithmic Hiring in Practice: Recruiter and HR Professionals' Perspectives on AI Use in Hiring' in *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (2021) 166.

20 E. Albaroudi et al., 'A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring' (2024) 5 *AI* 383, at 400.

This article is in five parts. Part 2 outlines the terminology used. Part 3 provides an overview of the mechanics of algorithm-facilitated discrimination against job seekers by employers using AHSs. I then detail the research methodology and design in Part 4. In Part 5, I report my research findings that there are six ways in which employers create real risks of discrimination when they deploy AHSs. Discrimination may occur as a result of: (1) the training data that are used in the system, (2) the use of proxies, (3) the way in which the system is implemented, (4) the construction of new structural barriers, (5) a failure to provide reasonable adjustments, and (6) the facilitation of an intention to discriminate. In Part 6, I discuss the broader relevance of these findings beyond the Australian context and make some concluding observations regarding how discrimination may be prevented when employers use AHSs.

2 | TERMINOLOGY

This article proposes a new term to capture the discrimination that is its subject. The term 'algorithm-facilitated discrimination' is used to describe the unfavourable treatment of, or disadvantageous effect on, members of groups because of their personal attributes such as disability,[21] race, ethnicity, national origin, sex and gender (including pregnancy, gender identity, intersex status, and sexual orientation), class,[22] and/or age when employers use AHSs. These attributes are collectively described as 'protected attributes', and individuals with those attributes as belonging to 'protected groups'. Algorithm-facilitated discrimination is not necessarily unlawful pursuant to discrimination legislation in Australia.[23]

The phenomenon of algorithm-facilitated discrimination is often described by regulators, researchers, and journalists as 'algorithmic discrimination' or 'algorithmic bias'. However, I contend that the term 'algorithm-facilitated discrimination' is to be preferred.[24] 'Algorithmic discrimination' is not able to encompass and, in fact, obscures the different modes of discrimination when an AHS is used by an employer. My research establishes that such discrimination may occur not only because it is embedded in the AI system (as is suggested by the term 'algorithmic discrimination'); it may also result from the interaction of social and organizational processes with AI technologies, such as the way in which the system is implemented or operated. 'Algorithm-facilitated discrimination' also positions any discrimination as emanating from a person rather

[21] Considering the importance of language, in this article I take a 'person-first' approach when discussing people and disability, using the phrase 'people with disability' to emphasize the person, not the disability, as recommended in People with Disability Australia, *PWDA Language Guide: A Guide to Language about Disability* (2021), at <https://pwd.org.au/wp-content/uploads/2021/12/PWDA-Language-Guide-v2-2021.pdf>. I acknowledge that many people with disability prefer 'identity-first' language that positions disability as an identity category.

[22] This personal attribute may also be described as 'social origin', 'social status', 'social condition', or 'social-economic status or disadvantage'.

[23] This is because there may be gaps in those laws. There is a need to review existing discrimination laws in Australia and other comparable jurisdictions to identify and address those gaps. This task has begun: see for example N. Sheard, 'Employment Discrimination by Algorithm: Can Anyone Be Held Liable?' (2022) 45 *University of New South Wales Law J.* 617; A. Kelly-Lyth, 'Algorithmic Discrimination at Work' (2023) 14 *European Labour Law J.* 152; S. Wachter, 'The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law' (2023) 97 *Tulane Law Rev.* 149.

[24] I have adapted this term from the terms 'technology-facilitated violence' and 'technology-facilitated abuse' used to describe the use of digital technologies to perpetrate violence and abuse: see for example J. Bailey et al. (eds), *The Emerald International Handbook of Technology-Facilitated Violence and Abuse* (2021).

than from an algorithm or AI system. The algorithm or AI system *facilitates* or is *the means employed by* a person to perpetrate discrimination (whether intentionally or otherwise). It makes clear that an AI system is not separate from, but rather under the control of, the person using that system, and it is that person who is responsible for it. Furthermore, by shifting the focus away from the technology, this new term situates algorithm-facilitated discrimination within broader patterns of social, political, and economic inequality.

AHSs are used at all stages of the hiring funnel.[25] My research focused on three AHSs used at the screening stage: CV parsing systems, candidate assessment systems ('assessment systems'), and video interview systems. These three systems were the focus of my investigation as I consider that they pose some the greatest risks of algorithm-facilitated discrimination in recruitment.

3 | HOW DOES DISCRIMINATION OCCUR IN AN AHS?

There is a rich corpus of legal and technical scholarship that has developed taxonomies to describe how predictive AI systems may lead to discrimination against protected groups.[26] Discrimination may occur at any point in the AI lifecycle and be a consequence of its data, social or technical bias, or the mode of implementation.

Data bias is perhaps the most common source of algorithm-facilitated discrimination.[27] Data are not neutral.[28] People decide how, when, and what data will be sourced and processed. When the training data are not representative of the population under consideration or they embed real-world discrimination, such discrimination may be transmitted to the AI system and disproportionately impact those already experiencing historical disadvantage.[29] This is a particular problem for AHSs. Training data of the kind required for these systems are expensive and difficult to obtain. AHS vendors usually purchase training datasets as a service when they acquire AI models from third parties,[30] and lack the ability and resources to thoroughly audit them.[31] AI systems trained on material from the internet, such as large language models that obtain datasets from Wikipedia, where, for example, less than 15 per cent of the authors are women or girls, risk 'perpetrating dominant viewpoints, increasing power imbalances and further reifying inequality'.[32]

[25] M. Bogen and A. Rieke, *Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias* (2018), at <https://www.upturn.org/work/help-wanted/>.

[26] See for example S. Barocas and A. D. Selbst, 'Big Data's Disparate Impact' (2016) 104 *California Law Rev*. 671; I. Bartoletti and R. Xenidis, *Study on the Impact of Artificial Intelligence Systems, Their Potential for Promoting Equality, Including Gender Equality, and the Risks They May Cause in Relation to Non-Discrimination* (2023), at <https://rm.coe.int/prems-112923-gbr-2530-etude-sur-l-impact-de-ai-web-a5-1-2788-3289-7544/1680ac7936>.

[27] International Organization for Standardization, *Information Technology – Artificial Intelligence (AI) – Bias in AI Systems and AI Aided Decision Making* (2022) 10–12, at <https://aistandardshub.org/ai-standards/information-technology-artificial-intelligence-ai-bias-in-ai-systems-and-ai-aided-decision-making/>.

[28] P. Kim, 'Data-Driven Discrimination at Work' (2016) 58 *William & Mary Law Rev*. 857, at 883.

[29] See for example V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (2017).

[30] For example, large language models may be purchased from Google, Meta, and OpenAI.

[31] M. Raghavan et al., 'Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices' in *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020) 469, at 475.

[32] E. M. Bender et al., 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' in *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021) 610, at 614.

Vendor-acquired datasets are usually aggregated and 'fine tuned' with an employer's own workplace data. However, this data may be limited, and reliance may therefore be placed on small and incomplete datasets. For example, employers typically have less data on employees with disability, and the data that they do have are likely to be incomplete, without comparability, and of low quality.[33] In addition, the training data for an AHS are likely to exclude people from disadvantaged populations who 'live on big data's margins … and whose lives are less "datafied" than the general population's'.[34] In Australia, that is likely to include First Nations peoples who have lower rates of digital inclusion than the rest of the population.[35]

Many of the AHSs on the market globally incorporate speech and language technologies (SLTs), such as NLP in large language models, designed to process and analyse text or speech.[36] However, there is a real risk that these technologies have been trained on data that are not representative.[37] As of 2022, only a small subset (0.28 per cent) of the over 7,000 languages in the world were found in the training data for SLTs.[38] English, Spanish, German, Japanese, and French were five of the seven languages best-represented in these datasets.[39] As a result, these systems have been found to perform poorly for people from other language groups.[40] By treating the 'expected' or 'standard' linguistic patterns in these languages as the norm, SLTs used in hiring processes are at risk of misunderstanding, penalizing, and unfairly screening out candidates from culturally and linguistically diverse groups as well as those with disability.

Discrimination against protected groups may also occur when the training data encode real-world discrimination. In this situation, there is often no statistical bias, as the training data may be accurate and representative. Algorithm-facilitated discrimination, however, may still occur as the algorithm models an unequal world, and this world is replicated in the outputs and predictions of that model.[41] There are many examples of how AI systems incorporating large language models trained on real-world data can quickly absorb social prejudices and stereotypes.[42] Human hiring discrimination may be replicated when the training data comprise organizational workplace data,

[33] N. Tilmes, 'Disability, Fairness, and Algorithmic Bias in AI Recruitment' (2022) 24 *Ethics and Information Technology* 28.

[34] Barocas and Selbst, op. cit., n. 26, p. 684.

[35] Australian Digital Inclusion Index, 'Digital Inclusion: The Australian Context in 2023' *Australian Digital Inclusion Index*, 2023, at <https://www.digitalinclusionindex.org.au/digital-inclusion-the-australian-context-in-2023/>.

[36] The phrase 'speech and language technologies' is taken from N. Markl, 'Language Variation and Algorithmic Bias: Understanding Algorithmic Bias in British English Automatic Speech Recognition' in *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (2022) 521, at 522.

[37] An example of how non-representative training data may lead to discrimination against protected groups is provided by the 'Gender Shades' study: see J. Buolamwini and T. Gebru, 'Gender Shades: Intersectional Accuracy in Commercial Gender Classification' in *Proceedings of Machine Learning Research: Conference on Fairness, Accountability and Transparency* (2018) 1.

[38] Markl, op. cit., n. 36, p. 2; Bender et al., op. cit., n. 32, p. 612.

[39] Markl, id.

[40] S. L. Blodgett et al., 'Language (Technology) Is Power: A Critical Survey of "Bias" in NLP' in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020) 5454.

[41] S. Mitchell et al., 'Algorithmic Fairness: Choices, Assumptions, and Definitions' (2021) 8 *Annual Rev. of Statistics and Its Application* 141, at 146.

[42] See for example M. Cheng et al., 'Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models' in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguists, Volume 1* (2023) 1504.

such as performance reviews[43] or the human grading of interviews.[44] Significantly, AI systems trained on data embedding present-day or historic discrimination and structural inequalities are at risk of creating unjust 'feedback loops'.[45] Feedback loops occur when 'correlations arising from historical discrimination' are used as 'quasi "causal" bases for decision making'.[46]

Discrimination may also be embedded in an AI system during its development through design and engineering decisions rooted in social and technical bias. Social bias results from the conscious or unconscious biases of individuals with significant input into the design and development of a system, or when broader societal or institutional values and norms are transmitted to the system.[47] For example, AHSs may encode 'proxy discrimination'[48] when features such as gaps in the employment history of job seekers are seen as relevant and included in the AI model. Gaps in employment history are, however, unlikely to have any saliency when it comes to job performance, but may operate as a proxy for sex or gender, as women are disproportionally the carers in our society, and for disability, as people with disability experience labour market marginalization and are more likely to need time out from work to obtain medical treatment for their conditions.[49]

Technical bias occurs through technological constraints, errors, inaccurate models, and bad design decisions. An illustration in the context of AHSs is provided in the O'Neil Risk Consulting & Algorithmic Auditing audit of the HireVue video interview system conducted in late 2020.[50] This audit revealed that some candidates' responses to video interview questions did not contain enough content for the algorithm to generate meaningful competency scores, and HireVue found in an analysis prompted by the audit that there were differences across ethnicities in the rate of videos falling into this category.[51] HireVue found that short answers, such as 'I don't know', were disproportionately given by minority candidates. By failing to incorporate a pre-assessment warning that short answers could not be scored, or to provide additional prompts to candidates to answer questions with more detail, the HireVue system exhibited a poor design and technical bias.

Finally, discrimination may result from the way in which AHSs are implemented by individuals and organizations. This cause of algorithm-facilitated discrimination is often underestimated, with the taxonomies developed in the literature largely focusing on the technical means by which discrimination may be embedded in an AI system. However, hiring systems that incorporate AI are 'an *assemblage* of human and non-human actors'.[52] The machine-learning algorithm is just

[43] See for example M. Cheong et al., *Ethical Implications of AI Bias as a Result of Workforce Gender Imbalance* (2020), at <https://www.tmbank.com.au/-/media/unibank/about-us/member-news/report-ai-bias-as-a-result-of-workforce-gender-imbalance.ashx>.

[44] See for example J. Yang et al., 'Professional Presentation and Projected Power: A Case Study of Implicit Gender Information in English CVs' in *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science* (2022) 140.

[45] See for example O'Neil, op. cit., n. 12, p. 87.

[46] Bartoletti and Xenidis, op. cit., n. 26, p. 29.

[47] B. D. Mittelstadt et al., 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 *Big Data and Society* 1, at 7.

[48] Proxy discrimination may occur when algorithmic models in AHSs use variables or features to assess job applicants. Many of these features appear to be neutral but may be statistically correlated with, and become proxies for, protected attributes.

[49] S. Nugent et al., *Recruitment AI Has a Disability Problem: Questions Employers Should Be Asking to Ensure Fairness in Recruitment* (2020) 10, at <https://osf.io/preprints/socarxiv/emwn5_v1>.

[50] O'Neil Risk Consulting & Algorithmic Auditing, *Description of Algorithmic Audit: Pre-Built Assessments* (2020).

[51] Id., p. 5.

[52] M. Ananny and K. Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 *New Media & Society* 973, at 984.

one part of this assemblage that is 'entangled' with individuals, organizations, cultures, and practices.[53] If we are to understand the risks of discrimination when employers use AHSs, we must explore the interaction of the social and the technical in the hiring process and how AHSs are 'shaped or reconfigured at the multiple points of consumption and use'.[54] We must also examine the economic, political, social, and organizational contexts in which they are deployed.[55]

4 | METHODOLOGY

Qualitative, semi-structured interviews were conducted with a convenience sample of participants (n = 23) over a one-year period between March 2022 and March 2023. Participants were recruited through snowball and 'opt-in' sampling techniques via my professional and personal networks, group posts or direct messages on the social media platform LinkedIn, and recommendations from other participants. Participants included in-house and agency talent acquisition, human resources, and diversity and inclusion professionals (collectively referred to in this article as 'recruiters') (n = 18). They had worked in, or provided recruitment services for, small, medium, and/or large organizations, both private and public, in a range of industries. In addition, to gain a broader perspective on the use of AHSs in Australia, I interviewed two careers coaches (to understand the impact of AHSs on job candidates), a leading Australian AI expert, and two employees of a large AI developer (the Director of AI Services and the AI Ethics Leader). All of the interviewees lived and worked in Australia. None of the participants were paid.

Of the 18 recruiters interviewed for this study, 13 were using and/or had previously used AHSs in their organizations. Assessment systems were the most common form of AHS used by participants (nine participants had used or were using these systems, though the mode of assessment varied), followed by CV parsing systems (seven participants had used or were using these systems) and video interviewing systems (five participants had used or were using these systems). There were a range of Australian and internationally developed systems in use. The five remaining recruiters were actively considering the use of an AHS (n = 1), worked for organizations that had made the conscious decision not to deploy an AHS (n = 2) or were too small to warrant the use of an AHS (n = 1), or had not considered using one (n = 1).

The interviewees were asked questions in two categories. First, all of the participants were asked to describe, in as much detail as possible, their understanding of the operation of the AHSs of which they had experience, including the features, target variables, and datasets. Second, the recruiters were asked to describe how they operationalized and used the AHS in their organization. The interview data were analysed using Virginia Braun and Victoria Clarke's 'reflexive thematic analysis' method[56] within the qualitative data analysis software NVivo.

The interview study was supplemented by a qualitative content analysis[57] of publicly available 'white papers' for two prominent AHS vendors in the global market, HireVue[58] and Sapia,[59] and

[53] Id., p. 981.

[54] J. Wajcman, 'Feminist Theories of Technology' (2010) 34 *Cambridge J. of Economics* 143, at 149.

[55] Li et al., op. cit., n. 19, p. 173.

[56] V. Braun and V. Clarke, *Thematic Analysis: A Practical Guide* (2022).

[57] In a content analysis study, the content of a particular text is analysed to reveal its messages and meaning: see W. L. Neuman, *Social Research Methods: Qualitative and Quantitative Approaches* (2014, 7th edn) 372.

[58] HireVue, *AI Explainability Statement* (2022).

[59] B. Jayatilleke, *Towards Establishing Fairness in AI-Based Candidate Screening* (2022), at <https://sapia.ai/resources/whitepaper/towards-establishing-fairness-in-ai-based-candidate-screening/>; Sapia, *Bias-Free Predictive Selection* (n.d.).

a research seminar given by Sapia's Chief Data Scientist, Buddhi Jayatilleke.[60] The white papers provide a level of transparency, not previously publicly available, about the technical operation of two market-leading AHSs. This content analysis was conducted to provide context and add technical detail to the interview data.

Three limitations of the study are highlighted. First, in this fast-moving area, human–computer interactions are constantly evolving with the introduction of new technologies to the global market, such as generative AI models including ChatGPT. It is therefore possible that the operationalization and use of AHSs by employers has changed since the interviews were conducted. Second, in this environment, there is likely to be an increased and increasing level of AI literacy among recruiters and more awareness around the risks of using AHSs. Third, it may be considered a limitation of the study that all of the participants were from a single jurisdiction. However, in Part 6, I argue that the findings of this study are relevant beyond the Australian context, as the majority of the AHSs used by research participants in their organizations are available on the global market.

5 | RESEARCH FINDINGS

The key findings from this research are that the socio-technical interaction between people, organizations, and AHSs creates real risks of algorithm-facilitated discrimination.[61] This part describes six different ways in which such risks arise.

5.1 | Data-driven discrimination

As discussed in Part 3, one of the primary means by which algorithm-facilitated discrimination may be embedded in an AHS is through the data used to train the system. All of the assessment systems used by participants in this study were purchased 'off the shelf'. The participants described the training data for these systems as coming from two sources: (1) their organization in the form of workplace data, and (2) AHS vendors. P3, an AI expert, expressed the view that, in practice, these two data sources are often combined.

5.1.1 | Organizational workplace data

Organizational workplace data are added to a vendor's own data to fine tune the system to an organization's needs. Employers in this study obtained organizational workplace training data by 'putting' current employees 'through the model'. These employees might just be the high performers (P5 reported putting 'our 50 top-performing graduates through the model') or, in other cases, all of the employees in a particular role such as first- or second- year graduates (P13). As P1 explained:

---

[60] B. Jayatilleke, 'Using NLP for Fair and Faster Hiring with Conversational AI' research seminar, La Trobe University, 4 May 2023.

[61] It is important to emphasize that all of the research participants in this study were acting in good faith in their use of AHSs.

> [W]e would get [high performers] … within the call centre to [do a behavioural assessment] … and then we'd build a profile off … what the ideal person looks like from all of those behaviours. Then we would send out the testing to people, and then we'd only bring people in for interview who matched the profile.

Training data comprising assessments of those employees by human graders were also used by employers (P17).

When organizational workplace data are used to train an AHS, there is a real risk that, as in the Amazon recruitment tool example, the training data are a product of historical and/or present-day hiring discrimination. Where such discrimination is or has been present, it will be encoded into the algorithmic model and replicated in the predictions and outputs of the AHS. For example, if the human graders exhibit bias – whether conscious or unconscious – in their assessment of candidates, this will be reproduced in the outputs of the AHS. In addition, employer-provided training data are often, by their nature, insufficiently large and incomplete and may, for those reasons, lead to inaccurate and unreliable results. Participants acknowledged that this is the case while the system is 'learning'. P17 described there being a 'chasm' in the AI's reliability for some questions until the system's learning is sufficiently advanced with additional data. If protected groups disproportionately experience these errors, algorithm-facilitated discrimination will ensue.

5.1.2 | Vendor-acquired data

Reliance by organizations on AHSs trained on the data provided by their AHS vendor also creates risks of algorithm-facilitated discrimination. As P19 stated, in this situation, organizations are heavily reliant on vendors to make sure their datasets are 'accurate'. However, the origin and representativeness of vendor-provided training data is not usually disclosed. For example, Sapia has built its system using two datasets.[62] As training data are difficult and costly to produce, it relies on 'big tech' to build the 'AI substrate' for its NLP models (Sapia uses Google's BERT model). Sapia then adds its 'specialised data' to fine tune the model. These specialized data consist of 'responses [of job applicants] to our assessments' when they have been required to do so by a potential employer who has licensed the system. By harvesting data in this way, Sapia had amassed, as of April 2022, in excess of one billion words in its interview dataset. HireVue's training data are obtained in the same way. Its NLP model is based on a 'state-of-the-art' language model called RoBERTa which, in turn, was also adapted from Google's BERT model.[63] This model is then fine-tuned with specialized data obtained from 'over 70 million interviews' that have been conducted using the HireVue system.[64]

The risks of algorithm-facilitated discrimination arising from vendor-provided datasets were alluded to by only one participant in this study. P19 questioned: 'Who are you really comparing [a candidate] against? Is it other grads? Is it other managers? Is it Australia? Does that matter? It might not.' It does matter. Potential privacy issues aside, there is limited transparency regarding the data used to train large language models, and the potential biases of these systems are well

[62] Jayatilleke, op. cit., n. 60.

[63] HireVue, op. cit., n. 58, p. 7.

[64] HireVue: <https://www.hirevue.com/>.

known.[65] In particular, Google's BERT model has been found to encode more bias against intersectional identities than it does on just the combination of bias across each individual axis,[66] and to associate phrases referencing persons with disability with negative sentiment words.[67]

There is also a risk that specialized training data, obtained from the vendor's clients, may contain a sampling bias. Such training data may not be representative of the population who will be subject to it, and, where this is the case, it will lead to errors and inaccuracies for those diversity cohorts not represented in the data. For AHSs that are new to the market or have a small development budget, there is a particular risk that specialized training data may have been obtained from only a narrow selection of employers and/or industries and may, therefore, disadvantage some demographic groups. For example, workplace data collected from employers in science, technology, engineering, and maths (STEM) industries are likely to privilege young, white men in the recruitment process.[68]

In its *AI Explainability Statement*, HireVue asserts that it uses 'thousands of expert human rater evaluations of standardised interviews' (the 'rater studies') to train its model to score candidate interview responses.[69] HireVue states that it has conducted expert ratings of 30,000 interviews with job applicants across 15 competencies over the past three and a half years.[70] HireVue provides a table summarizing the industries represented in the rater studies. It shows that 17 per cent of job applicants in that training data worked in 'Healthcare', 15 per cent in 'Retail', 10 per cent in 'Hospitality, Recreation, & Leisure', 10 per cent in 'Insurance', 8 per cent in 'RPO & Sourcing', 8 per cent in 'Transportation', 7 per cent in 'Banking & Finance', 6 per cent in 'Consulting Services', 3 per cent in 'Food & Beverage', and 3 per cent in 'Technology'.[71]

Some industries are either not represented (including education and professional services) or under-represented (for example, though AHSs are commonly used in the 'Food & Beverage' industry, data from that industry comprise only 3 per cent of the training data). While men and women are almost equally represented across all of the rater studies,[72] we know that some of these industries, such as transportation, are male dominated in Australia, whereas others, such as healthcare, are female dominated.[73]

There is also a risk that, if the AHS vendor is an international company, the specialized training data may not be representative of the unique demographics and relevant diversity groups in the country in which the AHS is deployed. In the HireVue system, for example, only 6 per cent of the job applicants in the rater studies training data are from Australia and New Zealand, with 9 per

---

[65] See Cheng et al., op. cit., n. 42.

[66] Bender et al., op. cit., n. 32, p. 614.

[67] B. Hutchinson et al., 'Social Biases in NLP Models as Barriers for Persons with Disabilities' in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguists* (2020) 5491.

[68] For example, 'women make up only 22% of AI workers globally': UN Women, 'In Focus: International Women's Day' *UN Women*, 1 March 2023, at <https://www.unwomen.org/en/news-stories/in-focus/2023/03/in-focus-international-womens-day>.

[69] HireVue, op.cit., n. 58, p. 8.

[70] Id., p. 12.

[71] Id., p. 29, Appendix C. The *AI Explainability Statement* does not set out what the abbreviation RPO stands for. In this context, it is assumed to mean recruitment processing outsourcing.

[72] Id. Men make up 48 per cent, and women 52 per cent, of the applicants in the HireVue rater studies sample.

[73] Workplace Gender Equality Agency (Australia), *Gender Segregation in Australia's Workforce* (2016) 5, Table 4, at <https://www.wgea.gov.au/publications/gender-segregation-in-australias-workforce>.

cent from Europe and 5 per cent from Asia, while 78 per cent are from 'Northern America'.[74] In terms of 'Race/Ethnicity', 36 per cent of job applicants in the training data are 'White', 33 per cent are 'Hispanic', 17 per cent are 'Black', and 14 per cent are 'Asian'. It is clear that this 'Race/Ethnicity' sample is not attuned to, for example, Australian demographics. The Australian population is not 33 per cent 'Hispanic' or 17 per cent 'Black', and the Australian community has a higher proportion of people born overseas compared to Northern America.[75] Australia's First Nations peoples are entirely absent from this dataset. Moreover, no intersectional groups are represented in this data. The risks of discrimination are therefore high when the HireVue system is used by employers in countries not adequately represented in the training data, as it is likely to be inaccurate at assessing many segments of those populations.

5.1.3 | AI error

As discussed in Part 3, the majority of assessment systems on the global market employ NLP techniques to analyse candidates' interview responses.[76] A lack of representativeness in the data used to train these systems may introduce AI error. Following legal action and the discrediting of the use of facial analysis in video interviewing systems,[77] many of these systems[78] now use NLP to assess a transcription of the audio component of a candidate's video interview. The HireVue system operates in this way using a third-party provider to convert the audio of a candidate's video interview to text. HireVue asserts that this service is able to recognize 'the sound of words based on its experience and learning from over 50,000+ hours of human-transcribed content across a range of topics, industries, accents and inflections'.[79] Exactly which 'accents' and 'inflections' are represented in this data, however, is not disclosed. It is also not clear what variety of English the system has been trained on.[80] As the system is built in the United States (US), is it correct to assume that it is trained on standard or mainstream American English?

HireVue states that its transcription service has a 'word error rate (WER)[81] for English-language speakers in the US of less than 10 per cent on average.[82] However, it discloses that, when tested on non-native English speakers with accents from other countries of origin, this WER increases to between 12 and 22 per cent depending on the candidate's country of origin. The countries of origin tested by HireVue are not disclosed, but it does state that the WER for non-native English speakers

[74] HireVue, op.cit., n. 58, p. 29, Appendix C.

[75] For example, 29.8 per cent of the Australian population were born overseas, whereas this figure is only 15.3 per cent for the US: Australian Bureau of Statistics, 'Australia's Population by Country of Birth' *Australian Bureau of Statistics*, 26 April 2022, at <https://www.abs.gov.au/statistics/people/population/australias-population-country-birth/2021>.

[76] For example, the HireVue and Sapia systems evaluate candidates using NLP techniques.

[77] L. F. Barrett et al., 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements' (2019) 20 *Psychological Science in the Public Interest* 1.

[78] We do not know the precise number, as there is no research regarding this aspect of the video interviewing systems on the global market.

[79] HireVue, op. cit., n. 58, p. 6.

[80] In Nina Markl's study, the SLT performed 'significantly worse' on speakers of regional varieties of British English: Markl, op. cit., n. 36, p. 1.

[81] WER is an established metric for assessing transcription accuracy. It is the 'number of errors (deletions, insertions, substitutions) in an automatic transcript relative to the number of correct words in the reference transcript': id., p. 5.

[82] HireVue, op. cit., n. 58, p. 11.

from China is 22 per cent. HireVue justifies the high levels of WER for job applicants with non-native English skills on the basis that its speech-to-text system is more accurate than alternative transcription services (including a service offered by Amazon), which, it asserts, can have a WER as high as 45 per cent for non-native English speakers from China. While HireVue estimates the WER for human transcription to be 5–10 per cent, it states that this means of transcription is neither 'economically feasible nor time-efficient … when processing the millions of interview responses so they can be auto-scored with our AI algorithms'.

Participants in this study were able to conceive of a range of circumstances in which WERs by SLTs may lead to discrimination against protected groups. They hypothesized that the assessment of a transcription of the audio of a person's interview responses may not be accurate for candidates who are non-native English speakers or have a disability that affects their speech, such as a stutter. If the words that a person speaks are not accurately transcribed, then that candidate's interview cannot achieve a high score by the algorithm. As P20 observed, this score will not be an assessment of a person's communication skills but will instead become an assessment of their English-language ability.

An awareness of this issue led two participants, P17 and P19, to seek reassurance from their vendor that the AHS that they deployed did not disadvantage candidates with accents. Both participants were given assurances by the vendor that the AHS was 'really good at understanding accents' (P17). No evidence was produced by the vendor to support this contention. For this statement to be true, the training data would need to include examples of the range of accents and language variations commonly found in Australia.

Other participants in the study expressed concern that AHSs may not be able to account for the fact that language and behaviours are learned and culturally contingent. P21 worried that an assessment system may not be able to accurately interpret cultural and linguistic nuances and variations in a candidate's written or spoken answers to interview questions. For example, she described how candidates from one cultural and linguistic background

> might use a lot of superlative language, like really enthusiastic adjectives which could be considered really good by an AI – in other words, 'Awesome', and 'Fantastic', and 'Oh, it was brilliant when this happened' – whereas others again would use really basic [language] like 'Good', 'I liked it', 'Found enjoyment there'. But … that level of vibrancy is not captured in those kinds of words.

How a person behaves and their emotional affect, vocabulary, and linguistic features vary between different cultures, regions within countries, and different countries, and this may disadvantage candidates from diverse backgrounds, as an AHS cannot understand these contextual factors.

5.2 | Proxy discrimination

Algorithm-facilitated discrimination is likely to include proxy discrimination when AHSs are used by employers. P9, who works for an AI developer, was of the view that the use of proxy variables poses the greatest risk of algorithm-facilitated discrimination. He stated that

> when talking about artificial intelligence, it turns out the most risky points are … the implicit features, or implicit … discrimination, because by simply removing protected attributes, it is not enough and artificial intelligence is particularly good in identifying … these proxies.

The interviews with recruiters revealed two ways in which proxy discrimination may occur: (1) it may be embedded in the algorithmic model in an AHS during its development, or (2) it may occur in the use of such a system by those deploying it.

The algorithmic model in the AHS may be designed to assess candidates on a feature that is highly correlated with, and therefore becomes a proxy for, a protected attribute. Participants in this study identified, for example, that features such as 'work experience in the last five to ten years' or an unbroken work history may be proxies for sex, age, and/or disability. As P2 acknowledged, women who have been out of the workforce raising a family may have a gap in their work history and lack recent work experience. P22 stated that people with disability may also have gaps in their employment history or a lack of prior experience because of the requirements of managing their disability.

Participants also identified other features that may become proxies for protected attributes when they are assessed by an AHS. They hypothesized that a requirement to make eye contact in a video interview may be a proxy for neurotypicality and result in discrimination against neurodivergent job seekers. Similarly, the need to speak and write English like a native speaker to score well on an assessment may be a proxy for, and therefore discriminate on the basis of, race or national or ethnic origin. Furthermore, assessing and scoring a candidate on their speed or accuracy in completing a written test may be a proxy for physical ability and may discriminate against people with disability.

As suggested in the above quote, the primary computational method for removing proxy discrimination is through 'anti-classification' whereby features found to be highly correlated with protected attributes or proxies are removed or downweighted.[83] Sapia states that it '[does] not use any attributes such as gender, age, race in our model building process'.[84] In its *AI Explainability Statement*, HireVue asserts that it 'permanently remove[s] features in data that are consistently associated with specific protected groups … [and] the model is then re-trained without the identified features'.[85] HireVue gives the pronunciation of particular words as an example of such a feature as it could be correlated with 'ethnicity'.

However, the effectiveness of the removal of features associated with protected attributes from the algorithmic model to avoid their use as datapoints in candidate assessments was questioned by research participants. P16, a recruiter and software engineer, doubted that discrimination could ever be removed from a CV parsing system. He was of the view that the 'signals' of protected attributes cannot be removed by simply taking out names and addresses. He gave the example of an applicant's race or nationality being encoded in other parts of their CV, such as the name of the university that they attended or the name of an employer (which may be unfamiliar or an international company).

The second means by which proxy discrimination may occur is through the use of filters in the AHS. Some employers used their CV parsing system as a tool to filter, and thereby, screen out candidates. Recruiters P6 and P11 described using the search functionality in the CV parsing system to screen CVs 'based on key words'. This could be used to search for applicants with certain qualifications or skills or to reduce the pool of applicants for a shortlist. P1 explained that the filtering functionality in a CV parsing system does 'come in handy' when you have

[83] J. Sánchez-Monedero et al., 'What Does It Mean to "Solve" the Problem of Discrimination in Hiring? Social, Technical and Legal Perspectives from the UK on Automated Hiring Systems' in *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability and Transparency* (2020) 458, at 461; Raghavan et al., op. cit., n. 31, p. 9.

[84] Sapia, op. cit., n. 59.

[85] HireVue, op. cit., n. 58, p. 14.

hiring managers who impose hard requirements, such as, 'I only want to see people who … [have a grade point average (GPA) of] 95 and above'. This can lead to candidates being 'screened out' without further human consideration, as the system locates candidates with that GPA and 'they're the only people then that we'd look at'. GPA is a known proxy for social class.[86]

LinkedIn's CV parsing system also enables recruiters to filter resumés. P10 explained:

> [W]hen you're looking in 'LinkedIn Recruiter', there's about ten different fields and then 20 extra ones you can search through. So, it can be anything from the university they went to, to the industry that they take themselves in on LinkedIn, the skills that they have put on LinkedIn. But the main ones recruiters are going to use are job title, probably location if it's an office-based role, and then the keyword searching.

For example, recruiters can search for, and filter, applicants based on 'year of graduation', 'years of experience', and 'seniority', all of which are proxies for age. As discussed above, 'experience' in LinkedIn, or any other CV parsing system, may also be a proxy for sex and/or disability. Furthermore, the filter and search option 'geographic location' in LinkedIn, or any other CV parsing system, could be used as a proxy for social class. This criterion and that of 'schools attended' may also discriminate on the basis of race or national or ethnic origin. LinkedIn was recently exposed for filtering out candidates automatically, and without the knowledge of recruiters, if the country selected on their LinkedIn profile did not match the country in which the role was located.[87]

P21, another participant from a small recruitment agency, explained how her applicant tracking system (ATS)[88] is integrated with SEEK[89] and LinkedIn and enables her to input screening questions. She explained that she developed the screening questions and could also 'tell the system' what is 'a bad answer or a good answer'. She surmised, as she had never tried it, that there is nothing to prevent those with discriminatory intent, or unconscious bias, from programming the system to exclude particular groups, such as women or older candidates. As she stated, 'you can just submit whatever [screening questions] you like'.

5.3 | Implementation discrimination

5.3.1 | Model development

While some recruiters in this study had opted for algorithmic models that had been pre-built by the vendor, others described working with the vendor to build a customized model in the AHS for particular roles, profiles, or competencies. These recruiters described selecting target variables

[86] See for example S. Lamb et al., *Educational Opportunity in Australia 2020: Who Succeeds and Who Misses Out* (2020), at <http://vuir.vu.edu.au/42362/1/educational-opportunity-in-australia-2020.pdf>.

[87] N. Kayser-Bril, 'LinkedIn Automatically Rates "Out-Of-Country" Candidates as "Not Fit" in Job Applications' *Algorithm Watch*, 31 August 2021, at <https://algorithmwatch.org/en/linkedin-recruitment-feature-discrimination/>.

[88] An ATS maintains a database of all applicant and job information. It is used by recruiters to streamline, manage, and track candidates through each step in the hiring process: see E. St-Jean and P. Thibodeau, 'What Is an Applicant Tracking System?' *Tech Target*, June 2024, at <https://www.techtarget.com/searchhrsoftware/definition/applicant-tracking-system-ATS>.

[89] Founded and headquartered in Australia, SEEK provides an online employment marketplace connecting job seekers and employers: see SEEK, 'About' *SEEK*, at <https://www.seek.com.au./about>.

for their systems ranging from 'a good fit' (P19), the 'right fit' (P2), and 'fit to role' (P19), to 'ideal person' (P1), 'success in the role' (P5), and '[someone who is a] great add to your business' (P8).

Concepts of 'fit', 'success', and the 'ideal' candidate are difficult to measure and convert into mathematical models in an AHS. Characteristics, personality traits, and behaviours considered desirable to employers are 'neither performed nor interpreted universally and objectively'.[90] It is argued that when an algorithmic model is developed for an AHS, there is a risk that it may privilege traits and norms associated with those who possess social, economic, and political power in society. Critical scholars have long challenged the 'ideal candidate norm', as, though it has the appearance of objectivity, the ideal candidate is in practice usually 'clothed with the attributes of the dominant gender, culture, religion, ethnicity or sexuality'.[91] Feminists view this as a 'male norm', as it positions men as the benchmark for what it means to be an ideal worker.[92] Similarly, disability advocates assert that this norm constructs people with disability as deficient, 'atypical', and 'abnormal'.[93] It also ignores the 'accumulating systemic disadvantage that many marginalised job seekers experience'.[94]

It is argued that the algorithmic models developed by employers in conjunction with AHS vendors are at risk of (re)producing and encoding ideal candidate norms. Research participants in organizations that had made the decision not to use an AHS in recruitment recognized this risk. P15 and P16 viewed criteria such as 'high performing' to be value laden and subjective. P15 also recognized the risk that AHSs may reinforce existing patterns of disadvantage, stating:

> high performing/not high performing – there's an element of subjectivity to that. Who's assessing that? … [I]f someone's made a decision early on in someone's career that they are not super talented, … was that a biased decision and then did that person not get offered opportunities? And then did all of that accumulate so that five years down the track, actually they're not performing well, but actually it was the biased decision early on that then was built upon?

This participant is describing a feedback loop. The job seeker is unable to progress in their career, as real-world discrimination has been built into the AHS.

It is further argued that when ideal candidate norms are built into an algorithmic model in an AHS, this creates a particularly potent form of discrimination against historically disadvantaged groups. Not only does the consistency of an AHS turn it into an unpassable barrier for job seekers, but it also reconfigures and enhances this norm in particularly harmful ways.

Human hiring discrimination can be inconsistent and unpredictable. As P4, an AI developer, asserted, the result of such discrimination is that members of protected groups may have to 'work harder' and some may be 'blocked by people's bias', but some may also 'get through' and achieve employment. It is asserted in this article that AHSs radically change this position. As HireVue explains, 'unlike interviews conducted by humans, our AI models are completely consistent across

[90] E. Drage and K. Mackereth, 'Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Difference"' (2022) 35 *Philosophy & Technology* 89, at 96.

[91] S. Fredman, 'Substantive Equality Revisited' (2016) 14 *International J. of Constitutional Law* 712, at 719.

[92] C. MacKinnon, *Feminism Unmodified: Discourses on Life and Law* (1987) 34.

[93] M. Whittaker et al., *Disability, Bias and AI* (2019) 12, at <https://ainowinstitute.org/publication/disabilitybiasai-2019>.

[94] D'Almada-Remedios et al., op. cit., n. 13, p. 22. See also H. van Dijk et al., 'Meritocracy a Myth? A Multilevel Perspective of How Social Inequality Accumulates through Work' (2020) 10 *Organizational Psychology Rev*. 240.

candidate pools'.[95] Therefore, when you take the ideal candidate norm and (quoting P4 again) 'put it into code, the risk is that no one from a particular group can ever get through'.

As a result of unrepresentative training data and the process of feature selection in the development of the algorithmic model, it is likely that the AHSs used by employers not only reinforce traditional ideal candidate norms but also create a new ideal candidate 'on steroids'. This ideal candidate, in a video interviewing system, becomes someone who displays the facial movements of the white-skinned men who are represented in the training data, and speaks using the 'pace, volume and diction'[96] of standard American English. In a text-based assessment system, the ideal candidate becomes someone who demonstrates the typing patterns or the speed and accuracy of non-disabled workers. Candidates who fall outside this codified model of the ideal candidate may be penalized by receiving a low score or ranking, or they may be automatically screened out of the hiring process. In this way, the use of an AHS 'remap[s] … and calcif[ies] the boundaries of inclusion and marginalisation' in employment opportunities.[97]

P22, a careers coach for people with disability, provided an example of algorithm-facilitated discrimination resulting from new codified models of the ideal candidate. One of P22's clients, a top student on his university course, had been unsuccessful at obtaining graduate employment. It was clear to this careers coach that the client's atypical answers to the questions asked in these assessments had been marked down and that this had resulted in him not being able to get 'through' the AHSs. P22 explained:

> [T]his particular student has achieved amazing grades – the same, if not better than the rest of their cohort – and they get to a psychometric test and, due to the barriers that face them because of their ASD [Autism Spectrum Disorder], cannot get past [it] … I think it was 23 [tests] they said that they've gone through and can't get past them. If they do get past it, then they get a self-recorded video interview and can't get past that.

This new codified model of the ideal candidate is particularly pernicious, as, with the lack of transparency inherent in AHSs, these new features are largely invisible to job seekers. They will, in all likelihood, not even know the basis for their assessment. Furthermore, social norms encoded and concealed in an AHS may appear to be neutral and unchallengeable, as they are based on 'science'.

5.3.2 | Customization

The majority of recruiter participants in this study explained that they are responsible for customizing and setting up many of the parameters of their AHSs. For example, P5 described how she made the decision to implement a pop-up screen in her organization's assessment system that provides a brief explanation of that system. One of the most important parameters that recruiters may establish for their assessment systems or video interviewing systems are the time limits for

[95] HireVue, op. cit., n. 58, p. 4.

[96] Id., p. 27, Appendix B.

[97] S. West et al., *Discriminating Systems: Gender, Race and Power in AI* (2019) 16, at <https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2>.

candidates to answer questions and/or complete an assessment or interview. P11 described how she set up time limits for candidates to provide their answers in a video interviewing system. She allowed 30 seconds for preliminary 'warm-up' questions and two minutes for 'behavioural-based' questions.

Participants in this study readily imagined how such time limits may disadvantage members of protected groups. They envisaged that they may disadvantage people with disability, including those with anxiety disorders or speech impediments, and/or people who are non-native English speakers. Such candidates, they believed, might need longer to provide their responses if they are to be assessed in a non-discriminatory way by the system. P2, who coaches candidates, complained that, her clients are often not told in advance what the time frame is for responding to questions, and this had resulted in them being 'pretty much cut off halfway through' their responses.

Participants described how candidates are normally given between one and three days to complete an assessment. P22, a careers coach for people with disability, said that for some people with disability this may be insufficient as they may have pressing medical issues arising from their disability, such as needing to change medication, which cannot be resolved within this time frame. Another recruiter, P21, was of the view that such time limits present 'real barriers' for women with caring responsibilities who may not have the time or the environment to do their best on the assessment in the time allowed.

5.4 | Structural barriers

Participants in this study identified two significant structural barriers that may prevent equality of access to job opportunities. The first barrier is created by the need for job seekers to possess 'digital resources' if they are to submit to an assessment by an AHS. Digital resources in this context include the ability to afford, access, and effectively use AHSs.[98] Participants explained that older candidates may be disadvantaged if they do not have sufficient digital literacy. Applicants experiencing poverty or living in non-urban areas will be disadvantaged if they do not have access to a reliable internet connection and/or a computer or smartphone.[99]

Many protected groups within OECD countries lack the digital resources needed to access employment opportunities through AHSs. In a 2020 report, the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance highlighted that countries in the Global North have concerning digital divides along the axis of race and ethnicity.[100] In the United Kingdom, research conducted in 2024 found that approximately 8 per cent of the population lack basic digital skills and 25 per cent of people with disability or health conditions feel left behind by technology.[101] In Australia, the 2023 Australian Digital Inclusion Index found that nearly a quarter (23.6 per cent) of the population were digitally

[98] J. McCosker et al., *Measuring Australia's Digital Divide: Australian Digital Inclusion Index 2023* (2023) 3, at <https://www.digitalinclusionindex.org.au/wp-content/uploads/2023/07/ADII-2023-Summary_Report_Final-1.pdf>.

[99] C. Wilson et al., 'Australia's Digital Divide Is Not Going Away' *The Conversation*, 29 March 2018, at <https://theconversation.com/australias-digital-divide-is-not-going-away-91834>.

[100] Achiume, op. cit., n. 13, p. 7.

[101] Good Things Foundation, 'Our Digital Nation' *Good Things Foundation*, 25 April 2024, at <https://www.goodthingsfoundation.org/policy-and-research/research-and-evidence/research-2024/digital-nation>.

'excluded' or 'highly excluded'.[102] Highly excluded Australians were more likely to have disability, live in public housing, or have not completed secondary school.[103]

Participants recognized that when AHSs are used, a lack of digital resources may lead job applicants with protected attributes to decide not to put themselves forward for roles or to drop out of a hiring process.[104] For example, P1 asserted that applicants with disability may decide having to submit to an AHS is 'too hard' and drop out of the process rather than ask for reasonable adjustments. In the view of P9, non-citizens in Australia with different languages and cultures are less likely to use LinkedIn or SEEK as they do not provide an 'effective' way for them to get a job. This cohort prefers to use social networks and informal ways of obtaining employment.

P7, P17, and P18 also reported that candidates from particular ethnic groups have high drop-out rates when AHSs are used. P17 and P18 had found that the drop-out rate for candidates from non-English speaking backgrounds was very high on their systems. P17 reported that the rate was 50 per cent for his assessment system, while it was 90 per cent for P18's video interviewing system. While neither suggested that this rate was higher for members of protected groups, both P17 and P18 recruit international candidates primarily from south-east Asia and India for allied health roles. P17 attributed the rate to a lack of motivation on the part of candidates, and not wanting to 'take the time out' to complete the assessment. The rate had caused P18 to cancel the use of the video interviewing system.

The second structural barrier identified in this research is the standardization of recruitment processes across organizations. This study found that AHSs enable hiring processes to be standardized *within* an organization. For example, the use of an assessment system enabled a large national retail chain to implement a consistent hiring process for all of the roles in its retail stores Australia-wide. The study also found some evidence of standardization *across* organizations. While only one participant had knowledge of such standardization, his experiences are troubling. P22, a careers coach for people with disability, explained that, for graduate roles, many employers use the same vendor to provide personality assessments. When this occurs, a job seeker's score on an assessment for one employer becomes the score that is used by all other employers who have licensed the same product. In his experience, if a job seeker, for example, with disability does not perform well on an assessment, they are bound by the results. The job seeker is only able to redo the assessment if they apply for a new graduate role and the previous assessment is over 12 months' old.

5.5 | Failure to provide reasonable adjustments

Algorithm-facilitated discrimination may occur where an employer using an AHS does not provide an adjustment that is reasonable and necessary for a job applicant to be evaluated by the system in a non-discriminatory way. In recent research, the Diversity Council of Australia found that job seekers with disability face 'critical accessibility problems' when applying for jobs online, including complex navigation, timeout restrictions, lack of video captioning or image alt-text,

[102] Australian Digital Inclusion Index, op. cit., n. 35.

[103] McCosker et al., op. cit., n. 98, p. 5.

[104] Some recruiters in Lan Li et al.'s study also identified poor assessment completion rates when AHSs are adopted. They expressed concern that candidates with protected attributes such as age, gender, and socio-economic status may be disadvantaged by these AI technologies: Li et al., op. cit., n. 19, p. 172.

poor screen contrast, inaccessible form fields, and mouse-only input options.[105] In the US, the Equal Employment Opportunity Commission (EEOC)[106] recommends that reasonable adjustments might include 'an alternative version of the test, including one that is compatible with accessible technology (like a screen-reader) if the applicant or employee uses such technology'.[107] P22, a careers coach for people with disability, explained that common adjustments sought by his clients are extended time to complete an assessment and/or the opportunity to take breaks during it. He stated that, for example, a person with a mental illness may experience additional anxiety, brain fog, or fatigue that may affect their ability to perform in the assessment if such adjustments are not provided.

All of the recruiters interviewed who had used or were using AHSs in the hiring process stated that applicants with disability are able to seek reasonable adjustments. In addition, some recruiters suggested that they would consider a request from job seekers for reasonable adjustments on the basis of any protected attribute. For example, P11 reported that they had permitted some candidates to do a phone interview, rather than one by video, if they felt more comfortable using that format. However, with the normalization of the use of assessment and video interviewing systems, particularly for graduate recruitment, participants acknowledged that it will become increasingly difficult for participants to opt out of this mode of assessment outside of requests for reasonable adjustments on the grounds of disability.

Recruiters indicated that job seekers are advised that they may seek reasonable adjustments either at the commencement of the hiring process or at the point when an assessment is offered to an applicant. Despite this, participants explained that, in practice, such adjustments are rarely requested. Suggested reasons included a lack of trust and a desire on the part of people with disability to establish that they 'can do it' and will not be 'a burden' (P22). This is consistent with other research that has found that disclosure of a disability in employment requires a secure work environment built on trust with an employer.[108]

This research found an additional reason why a reasonable adjustment may not be sought when AHSs are used in recruitment: job seekers may not know that they require one. There is a lack of transparency when AHSs are used by employers, and candidates are often not told that an algorithm will evaluate them. In addition, they are usually provided with little, if any, information about what is involved in completing the assessment, the features to be assessed, or how those features are weighted. As a result, candidates with disability often may not become aware of the need for reasonable adjustments until it is too late. P22 explained:

> Another really big thing is 'Well, they [job applicants with disability] could request reasonable adjustments …' Great, but you put that question on the application page, which is good, but you don't tell them what's going to come. So there's no transparency a lot of the time about what the recruitment process is going to be, so how

[105] Kaabel et al., op. cit., n. 3, p. 10.

[106] The EEOC is the regulator responsible for enforcing federal discrimination laws in the US.

[107] EEOC, *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees* (2022) para. 4, at <https://data.aclum.org/wp-content/uploads/2025/01/EOCC_www_eeoc_gov_laws_guidance_americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.pdf>.

[108] M. Gignac et al., 'Does It Matter What Your Reasons Are When Deciding to Disclose (or Not Disclose) a Disability at Work? The Association of Workers' Approach and Avoidance Goals with Perceived Positive and Negative Workplace Outcomes' (2021) 31 *J. of Occupational Rehabilitation* 638.

can they advocate for themselves to be able to go 'Oh, this is what I'll need to go into recruitment process'? … [T]hey don't know what's coming.

Furthermore, the EEOC guidelines state that algorithm-facilitated discrimination may occur when an AHS is not able to assess whether a job seeker is able to perform the genuine and reasonable requirements of the role if a reasonable adjustment is made.[109] At best, assessment and video interviewing systems evaluate whether a candidate is able to perform a role under typical working conditions. For example, P17 described a 'quick-fire' assessment process for an administrative role where candidates undertake a work sample test that assesses their attention to detail and typing accuracy. Candidates are screened out of the hiring process if they are not able to perform in those conditions. However, this assessment system does not evaluate whether a candidate with disability is able to perform work of the kind in the sample test if provided with an adjustment, such as improved screen contrast or non-mouse-only input options.

5.6 | Intentional discrimination

Today, we are unlikely to see a job advertisement indicating 'Women need not apply', and it is the accepted wisdom that discrimination has become more subtle and increasingly less overt.[110] As Belinda Smith asserts, the 'battle line has at least moved forward – it is no longer drawn over blatant and intentional exclusion, but has moved to a more indirect and structural form of discrimination'.[111] However, this research project suggests that this optimism may not be warranted. AHSs are able to 'breathe new life into traditional forms of intentional discrimination'.[112] These systems enable employers to knowingly and intentionally reject job applicants with protected attributes, while at the same time hiding and masking those motivations and actions.[113]

As discussed above, this research study has found that AHSs may be used by recruiters to intentionally filter out job applicants from protected groups. As P6 explained, with a CV parsing system, 'you could add some additional filters after it's come up with the search but then that bias would come from the recruiter, not the system'. P21 also described how LinkedIn and SEEK, through integration with ATSs, facilitate the inputting by recruiters of whatever screening questions they like, even discriminatory ones. To use a blatant example, a recruiter could add a screening question asking the gender of an applicant and then program the system to reject applications from women. According to P21, SEEK and LinkedIn have no in-built checks to prevent these types of questions.

This is not a fanciful example. One of the first cases worldwide of algorithm-facilitated discrimination by employers using an AHS involved intentional discrimination. In that case, iTutorGroup, an English-language tutoring service for students in China, sought to hire tutors based in the US to

[109] EEOC, op. cit., n. 107.

[110] S. Strum, 'Second Generation Employment Discrimination: A Structural Approach' (2001) 101 *Columbia Law Rev*. 458, at 459–460.

[111] B. Smith, quoted in Senate Standing Committee on Legal and Constitutional Affairs, Parliament of Australia, *Effectiveness of the Sex Discrimination Act 1984 in Eliminating Discrimination and Promoting Gender Equality* (2008) para. 5.12, at <https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Legal_and_Constitutional_Affairs/Completed_inquiries/2008-10/sex_discrim/report/index>.

[112] Barocas and Selbst, op. cit., n. 26, p. 692.

[113] Id., p. 713.

provide remote tutoring.[114] It was alleged that iTutorGroup programmed its AHS to automatically reject female applicants aged over 55 years and male applicants aged over 60 years. iTutorGroup rejected more than 200 qualified applicants based on their age through the use of this system. The company agreed to settle the case for $365,000 to be distributed among those applicants.

AHSs can mask discriminatory intentions, such as those of iTutorGroup, as, at least in Australia, there is no legal requirement that transparency be provided to job seekers by way of an explanation of how the system works.[115] AHSs also furnish employers with 'plausible deniability'. As P22, a careers coach for people with disability, stated, AHSs provide 'a hidden way of discriminating because it's AI doing it'. In a large recruitment exercise, such as for graduate recruitment, this may exclude members of protected groups at considerable scale and create a significant barrier at the critical career entry stage.

6 | CONCLUSION

This research study has evidenced how AHSs are used in practice by employers, and cast light on the various socio-technical mechanisms by which discrimination against marginalized groups may occur. It has found that the use of AHSs by Australian employers may solidify traditional forms of hiring discrimination, play an active role in creating new forms of structural discrimination, and pave the way for intentional discrimination.

The training data are at risk of embedding present-day and historical discrimination and may not be representative of the diversity of the population in the country in which the AHS is deployed. Many of the features built into the algorithmic models contain proxies for protected attributes, which may prevent members of protected groups from being shortlisted for jobs. Discrimination may be a consequence of the way in which the system is customized and set up by recruiters, or of a failure to provide reasonable adjustments to people with disability. A significant and novel contribution of this article is to identify and describe how AHSs create a new form of systemic discrimination against protected groups through the imposition of hard barriers to employment, and a (re)configuration, enhancement, and enforcement of ideal candidate norms. Finally, AHSs offer fresh opportunities for employers to engage in intentional discrimination.

The findings from this study are relevant beyond the Australian context, having an application in any country where AHSs are deployed by employers. The majority of AHSs used by research participants, such as the HireVue system, were designed and developed outside Australia and are available worldwide. LinkedIn also has a global reach. Therefore, the risks of discrimination emanating from the AHS itself, via the training data or the use of proxies in the AI model, as discussed in this article, will be present in any jurisdiction in which such a system is used. The risks of discrimination arising from the implementation of the system by people and organizations, including intentional discrimination, will also be present. AHS vendors provide a range of standard features and instructions to clients regarding how a system should be operationalized.[116] It is reasonable to

[114] EEOC, 'iTutorGroup to Pay $365,000 to Settle EEOC Discriminatory Hiring Suit' *US Equal Employment Opportunity Commission*, 11 September 2023, at <https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit>.

[115] N. Sheard, 'No Notice and No Explanation: The Incontestability of Hiring Discrimination by Algorithm' (2022) 35 *Aus. J. of Labour Law* 119. By contrast, see *Regulation (EU) No 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC* [2016] OJ L 119, Arts 13(2)(f) and 14(2)(g) (the General Data Protection Regulation).

[116] See for example HireVue, op. cit., n. 58, p. 2.

assume that employers in other jurisdictions have adopted the same features and complied with some or all of these instructions in similar ways to those identified in this study. Finally, AHSs may form structural barriers to equality of access to job opportunities in any country where social stratifications result in the digital exclusion of protected groups.

If we do not want disadvantaged groups to be subject to algorithm-facilitated discrimination, we need to take urgent action. It is essential that governments undertake review and reform of their discrimination laws to address gaps in the protection of job seekers from this form of discrimination.[117] These research findings provide the evidence base to do so. They are intended to inform legislators considering how best to tackle the problem of algorithm-facilitated discrimination.

In addition, we need greater transparency from the providers and deployers of AI systems, including large language models, regarding the operation of these systems. The training data must be curated and documented. We need increased levels of understanding by employers of the AHSs implemented within their organizations and their potential to cause harm at scale. It is essential that employers provide comprehensive training to those responsible for customizing, operating, and overseeing these systems. In the European Union, legislative reforms such as the General Data Protection Regulation[118] and the Artificial Intelligence Act[119] aim to address these issues. How effective they will be remains to be seen. In Australia, we are inching closer to AI regulation, with the government's recent proposal for mandatory guardrails for 'high-risk' AI applications,[120] but legislation is not yet in sight.

Finally, and more fundamentally, the identification in this research of significant risks to equality rights when employers use AHSs raises the question: should these systems be used at all? Should AHSs be deployed before necessary legal protections are in place and before we have developed a deeper understanding not only of the systems themselves and our interaction with them, but also of their impacts on historical, structural, and intersectional disadvantage in the global labour market?

ACKNOWLEDGEMENTS

I acknowledge the valuable and constructive comments of the anonymous reviewers. I also thank Robin Banks for her feedback and careful proofreading of an earlier draft.

Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

How to cite this article: Sheard, N. Algorithm-facilitated discrimination: a socio-legal study of the use by employers of artificial intelligence hiring systems. *Journal of Law and Society*. 2025;52:269–291. https://doi.org/10.1111/jols.12535

[117] See n. 23.

[118] General Data Protection Regulation, op. cit., n. 115.

[119] *Regulation (EU) 2024/1689 of the European Parliament and the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828* [2024] OJ L.

[120] Department of Industry, Science and Resources, Australian Government, *Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings* (2024), at <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2f6f02ebfe6a8190c7bdc/page/proposals_paper_for_introducing_mandatory_guardrails_for_ai_in_high_risk_settings.pdf>.