Framing Gender Bias in the Design of AI Recruitment Technology

Clementine Isabel Collett

Exeter College

University of Oxford

Thesis submitted in partial fulfilment of the requirements for the degree of DPhil in Information, Communication, and the Social Sciences in the Oxford Internet Institute at the University of Oxford

Supervisors: Professor Victoria Nash and Professor Ekaterina Hertog

May 2024
Word count: 99,951



Abstract

The Internet has rapidly revolutionised recruitment practices. Over the last two decades or so, there has been one particularly prominent change: the rise of AI recruitment technology (AI-rec-tech). This includes technologies which source, screen, assess, and select candidates. Despite their widespread use, there has been a significant lack of research concerning what lies behind the design.

This research offers one of the first in-depth, empirical, qualitative explorations into how vendors understand 'gender bias'. Moreover, it studies how this understanding is mobilised in system design. This investigation takes place within eight AI-rec-tech vendors which produce assessment technology. It uses data from online qualitative interviews, walkthroughs and autoethnography of the assessments, alongside material from conferences, company websites, and reports.

An adapted version of Goffman's theory on 'frames' (1974) is used to illustrate vendors' understanding (their conceptualisation) of gender bias, and to unpack how this is transposed to system design (their operationalisation). This thesis argues that we can make sense of how vendors frame gender bias in hiring through three discrete, but not mutually exclusive, frames: (1) the intrapersonal frame, (2) the structural frame, and (3) the statistical frame. When these frames are held together, they display a socio-technical interpretation of gender bias in hiring.

This research is the first to apply theory on frames to study AI design. I argue this theory is valuable due to its translatable analytical provocations. Throughout, I employ elements of Goffman's theory to scrutinise the 'vulnerabilities' of these frames. The vulnerabilities don't offer certainty on the impacts of the systems but highlight complexities within or surrounding the frame itself. This enables recommendations concerning vendor claims, system design, regulation, auditing, and an agenda for future research on this topic. Overall, this thesis supplies an important contribution, offering rare insight into these companies, their rationale, and practices.

"I do not intend here to provide a lullaby but merely to sneak in and watch the way the people snore."

Erving Goffman, 'Frame Analysis', p.14.

Acknowledgements

One of the best parts of this doctorate has been the people I've met and worked alongside. I feel so lucky to have found such brilliant friends, colleagues, mentors, and collaborators.

I'd like to start by thanking my supervisors – Professor Victoria Nash and Professor Ekaterina Hertog. I'm so grateful for their support and encouragement throughout this research and for their meticulous and thoughtful feedback on this thesis. Vicki and Katya have taught me so much, including how to be rigorous while crafting research into a story which unfolds – always to keep in mind the big picture along with the intricate detail. It has been wonderful learning from them both and I am so appreciative of everything they've done for me.

My deepest gratitude to my mentor and previous supervisor, Professor Gina Neff. Thanks to Gina for her guidance, support, and teaching over the years, and for being a tireless champion for her students and young scholars generally. Witnessing Gina's extraordinary passion for ensuring responsible and fair technology in society is inspiring. I continue to learn so much from her work and resolution to make the world a better place. I hope this thesis might be a minute contribution to a mammoth task, which I could not have undertaken without Gina.

Speaking of mentors, thank you also to Professor Sarah Dillon, who I met in 2019 when we co-authored our report 'AI and Gender'. Sarah has taught me so much, including how to structure an argument (and a paragraph, and a sentence, for that matter!) and how to simultaneously bring creativity and precision into written work. I'm grateful for Sarah's words of encouragement and wisdom.

And to Professor Laura Sartori – grazie mille, Laura! The three months I spent working with Laura at the University of Bologna is a time I will always treasure. I am so grateful for Laura's time, generosity, and continued mentorship. Laura showed me the best sides of academia - I will always cherish our long afternoons in the office discussing books, articles, theory, exchanging ideas, and brainstorming. It was the experience I had always hoped for in doing a doctorate. I am also so grateful to all the extraordinary people who I met in Bologna at the Department for Political and Social Sciences and beyond

(especially those at the Pasticceria Sant'Isaia) for their warmth, care, coffee, and the world's best pistachio croissants.

I'm grateful to the examiners who have read my work throughout the DPhil milestones, from Transfer to Confirmation. Thanks to Professor Jeremias Adams-Prassl, Professor Rebecca Eynon, Professor Kathryn Eccles, and Professor Vili Lehdonvirta, for all their feedback which was crucial in shaping this thesis as it stands today.

Thank you to my thesis examiners, Professor Rebecca Eynon and Professor Julia Ticona – I am very grateful that you are the ones to read and evaluate this work.

If Erving Goffman were alive, or perhaps wherever he is, I hope he would enjoy this. I am so glad to have had (a small pile of) Goffman's books placed into my hands by my dad many years ago, and I am pleased to have had an excuse to explore his work even further. Reading Goffman's blunt, perceptive, witty writing feels like being in conversation with him. To me, this is truly masterful. Thanks to Erving for his work and legacy.

Thank you to all the participants who have contributed to this research. I am sincerely appreciative of your time, openness, and for fitting me into your busy schedules. I hope this thesis does you and your work justice.

Many thanks to the UK Research and Innovation Economic and Social Research Council Doctoral Training Partnership (UKRI ESRC DTP), grant number ES/P000649/1, and the Exeter College 'Exonian Graduate Scholarship' for funding this doctorate. Particularly, thanks to Alan McKechnie for answering all my questions (and always so promptly!) on funding and the intricacies of doing a doctorate. Similarly, thanks to all those at the OII who have made everything run smoothly, particularly Laura Maynard, Martin Krigjsman, Chrissy Bunyan, Lucy Hennings, Duncan Passey, and Maisy Gills.

Thanks to the Women's Forum for Economy and Society for welcoming me into the inspirational AI Daring Circle early on in my doctorate, and to colleagues at UNESCO, IADB, and OECD for being such magnificent collaborators on our report 'The Effects of

AI on the Working Lives of Women'. It was a pleasure working with everyone involved. I am so proud of what we produced and the impact it has had so far.

To all the exceptionally warm, bright, caring people I have been lucky enough to meet at the OII, thanks for the camaraderie, the drinks, the lunches, the dinners, the seminars, the conferences, the knowledge-sharing, and so much else. In particular, I'd like to thank Amanda Curtis – for being such a generous and thoughtful friend, and for encouragingly sitting with me one dark, rainy, November night while I re-structured my thesis; Joanna Rivera-Carlisle – for the constant warmth, laughter, library companionship, for the best baking imaginable, and for doing such beautiful drawings for my thesis; and Marta Ziosi - for your joyful friendship, constant cheer, and for letting me ask you endless questions about football and how to speak Italian. Thank you also to Zoe Hawkins, Prathm Juneja, Roland Chen, Paul Rivera-Carlisle, Charlie Smith, Laura Herman, Jakob Mökander, Nayana Prakash, Hannah Rose Kirk, Liam Bekirsky, and Felix Simon.

I would also like to thank the 'Gender and Technology' and 'Algorithms at Work' reading groups for all the fascinating discussions, papers, and presentations over the last few years. I have so enjoyed being a part of these spaces and learning from everyone involved.

To the whole group at Exeter College, including, of course, Nico and our beautiful college family (which now spans four generations!) - thank you for making Oxford so special. Speaking of which, many thanks to my friend and long-suffering housemate (and, really, by this point, brother) Elliott Thornley, for being such an incredible housemate and for raising a clan of wild hedgehogs with me. I will miss Barrett Street, our neighbours, and our little house so much.

Thank you to all my friends from home for keeping me rooted and for being a constant source of joy, love, laughter, and reassurance. Particularly Laura, Belinda, Jenny, and Fei Yen. To those from Jesus College in Cambridge – our Jesus bubble – thank you for becoming family during those years in Cambridge. I am so grateful for your joyful and uplifting friendship. I'd also like to thank my godmother, Steph, for her unwavering love, generosity, thoughtfulness, and encouragement.

My deepest gratitude to the staff and the fellow readers at the Bodleian Library. I am grateful to them for always being a smiling face even on the dingiest of days. The Bodleian Library - specifically the Upper Reading Room and the Duke Humfrey's – is a place I love more than I can put into words. I've sat in that library almost every day for seven years - a quarter of my life - throughout my undergraduate degree and my doctorate. It's been my space of comfort and serenity, my home in Oxford. I cannot express how much I will miss this place.

To my parents and sister – Peter, Jill, and Katie – thank you for being my inspiration and for being there every step of the way. I never take for granted your love, support, cooking, laughter, care, and endless nuggets of wisdom. Thank you for giving the best pep talks and advice in the world (mum), for speaking about Goffman with me for hours (dad), for breaking down my timetable for thesis write-up in scrupulous detail (sis), and for so much more. I am so grateful to you for providing me with the support and opportunities which have got me to where I am today. This thesis is for you - I couldn't have done it without you three, and I couldn't feel luckier to be a part of our family.

Johnny - six years and two doctorates. Thank you for going first, for making me laugh when I don't feel it's possible, for giving me courage when I have none left, for demonstrating the meaning of hard work and determination, and for encouraging and supporting everything I do. I am so lucky to have met you.

Lastly, thanks to some particularly special people and places – Alpha Bar and Georgina's in the Covered Market; Jericho Coffee Traders; the Iffley Road swimming pool; the Thames towpath; Port Meadow; Christ Church meadows; the chapel choirs of Exeter, Worcester, and St John's colleges; Ella Fitzgerald; Van Morrison; Bob Dylan – for providing me with some of the fuel, energy, and tranquillity which has kept me going. And to the fiction writers whose work has kept me company over these years: thank you for showing me that imagination and life are always bigger.

CONTENTS

Abstract 2

Acknowledgements 4

List of Figures 14

List of Tables 14

CHAPTER 1: INTRODUCTION 15

1.1 The rise of AI recruitment technology (AI-rec-tech) 15
1.2 Framing gender bias in the design of AI-rec-tech 21
1.3 Key contributions 29
1.4 Definitions 30
1.5 Chapter outlines 35

CHAPTER 2: LITERATURE REVIEW AND THEORETICAL FRAMEWORK 41

2.1 Introduction 41
2.2 Uniting work on gender bias in relation to hiring and AI 42
2.2.1 Gender bias in hiring 43
2.2.2 Gender bias in AI 51
2.3 The gap in research on AI-rec-tech 55
2.3.1 Looking backwards: motivation for use, candidate perception, legal implications, and bias 56
2.3.2 Looking ahead: empirical work on gender bias and AI-rec-tech design 69
2.4 Theoretical framework 73
2.4.1 Laying foundations: the social shaping of technology 74
2.4.2 Structuring the research: Goffman's 'frame analysis' 76
2.4.3 Back to the building blocks: conceptualisation and operationalisation 82
2.5 Conclusion: research questions and literature contribution 85

CHAPTER 3: METHODS 88

3.1 Introduction 88
3.2 Data sample, sources, and methods 89
3.2.1 AI-rec-tech vendor companies 90
3.2.1.1 Company anonymity 92
3.2.1.2 Online interviews 97
3.2.1.3 Walkthroughs and autoethnography: demos and free assessments 104
3.2.1.4 Content research: conferences, websites, and reports 108
3.2.2 Law and regulation experts 110

3.2.2.1 Online interviews 111
3.2.2.2 Content research: regulatory and legal documents 114
3.3 Data analysis **115**
3.3.1 Frame analysis and thematic analysis 115
3.3.2 NVivo: seeing and coding 118
3.4 'The hustle': the struggle to access companies **119**
3.5 Data collection and analysis considerations **128**
3.5.1 Reflexivity and positionality in collection and analysis 128
3.5.2 Me, myself, and my interview persona 131
3.5.3 Data sufficiency: is rich data enough data? 133
3.5.4 Data configuration 136
3.6 Dynamic considerations **137**
3.6.1 Power 137
3.6.2 Rapport and fondness 140
3.6.3 Truth and self-censorship 143
3.7 Ethical considerations **145**
3.8 Conclusion **148**

CHAPTER 4: THE INTRAPERSONAL FRAME **149**

4.1 Introduction: framing gender bias as intrapersonal **149**
4.1.1 Chapter outline 151
4.2 Re-stating ideas on gender bias as human perception **152**
4.2.1 Pattern recognition (stereotypes) 155
4.2.2 Homophily 158
4.2.3 Subjectivity 160
4.3 Breaks in the intrapersonal frame: 'designing out' human perceptual biases? **163**
4.3.1 The remover? 166
4.3.2 The improver? 175
4.3.3 A note on the presentation of gender 187
4.4 Discussion **191**
4.5 Conclusion **199**

CHAPTER 5: THE STRUCTURAL FRAME **201**

5.1 Introduction: framing biases as structural **201**
5.1.1 Chapter outline 205
5.2 Structural injustice and the American dream **207**
5.3 Accounting for structural injustice in design **215**
5.3.1 Neutralising questions 216
5.3.2 Assessing skills and behaviours 220
5.4 Two 'out-of-frame' vulnerabilities **228**
5.4.1 Permeability: out-of-frame influences 232
5.4.2 Invisibility: exclusions from the frame 242
5.5 Discussion **254**

5.6 Conclusion **261**

CHAPTER 6: THE STATISTICAL FRAME **262**

6.1 Introduction: framing gender bias as statistical **262**
6.1.1 Chapter outline 269
6.2 Bias as adverse impact: taking root in Title VII of the Civil Rights Act 1964 ..... **271**
6.3 Measuring adverse impact **276**
6.4 Different 'keys': the dissonant tunes of vendors and regulators **284**
6.4.1 The compliance keys: the de-contextualisation vulnerability 286
6.4.2 The sufficiency keys: the loophole vulnerability 295
6.5 Discussion **305**
6.6 Conclusion **313**

CHAPTER 7: DISCUSSION **315**

7.1 Introduction **315**
7.2 Thesis summary: the three frames **315**
7.3 Notes on the nature and characteristics of the three frames **321**
7.4 Common 'vulnerabilities' **325**
7.5 'Laminating' frames: separate, sequential, and socio-technical **329**
7.6 Frame consistency **332**
7.7 Action recommendations **336**
7.8 Conclusion **342**

CHAPTER 8: CONCLUSION **344**

8.1 Thesis contributions **344**
8.2 Limitations **355**
8.3 Agenda for future research **361**
8.3.1 Client implementation: research within sociology & science and technology studies 362
8.3.2 Impacts of background, education, and opportunity: research within education 363
8.3.4 Audits and impact assessments: research collaboration between social data science, statistics, and sociology 365
8.3.5 Law and regulation: research within law 366
8.3.6 Candidates: research within psychology 367
8.3.7 Impact on demographics: research within social data science 368
8.4 Conclusion **368**

REFERENCES **370**

APPENDICES **393**

APPENDIX 1 – INTERVIEW GUIDE (VENDORS) 393
APPENDIX 2 – FIELDWORK JOURNAL FOR FREE ASSESSMENT 395
APPENDIX 3 – INTERVIEW GUIDE (REGULATORS) 396
APPENDIX 4 – SCREENSHOTS OF INFORMATION SHEET AND CONSENT FORMS.....397

Abbreviations

AAA	Algorithmic Accountability Act 2023 (US)
ADA	Americans with Disabilities Act of 1990 (US)
ADEA	Age Discrimination in Employment Act of 1967 (US)
ADHD	Attention Deficit Hyperactivity Disorder
ADS	Automated Decision System
AI	Artificial Intelligence
AI-rec-tech	Artificial Intelligence Recruitment Technology
ATS	Applicant Tracking System
BIPA	Biometric Information Privacy Act (US)
CDEI	Centre for Data Ethics and Innovation (UK)
CEO	Chief Executive Officer
CJEU	Court of Justice of the European Union
COO	Chief Operating Officer
COVID-19	Coronavirus Disease 2019
CPO	Chief Product Officer
CSM	Customer Success Manager
CSO	Chief Science Officer
CTBH	Cultural Test Bias Hypothesis
CTO	Chief Technology Officer
CUREC	The Central University Research Ethics Committee
CV	*Curriculum Vitae*
DEI	Diversity, Equality, and Inclusion
DNA	Deoxyribonucleic Acid
DREC	Departmental Research Ethics Committee
DSIT	Department for Science, Innovation and Technology (UK)
EEOC	Equal Employment Opportunity Commission (US)
EHRC	Equality and Human Rights Commission (UK)
EPIC	Electronic Privacy Information Center (US)
ESRC	Economic Social Research Council (UK)
EU	European Union
FTC	Federal Trade Commission (US)
GDPR	General Data Protection Regulation (EU)
HCI	Human-Computer Interaction
HOS	Head of Sales
HR	Human Resources
IADB	Inter-American Development Bank
ID	Identification
IFOW	Institute for the Future of Work (UK)
IHR	In-House Recruitment (UK-based Company)
IO Psychologists or IOP	Industrial and Organisational Psychologists
IP	Intellectual Property
MEPs	Members of European Parliament
ML	Machine Learning
MOOCs	Massive Open Online Courses

NDA	Non-Disclosure Agreement
NLP	Natural Language Processing
NYC	New York City
OECD	Organisation for Economic Co-operation and Development
OFCCP	Office of Federal Contract Compliance Programs (US)
OII	Oxford Internet Institute
RQ	Research Question
STEM	Science, Technology, Engineering, and Mathematics
STS	Science and Technology Studies
TV	Television
UNESCO	United Nations Educational, Scientific and Cultural Organization
Uniform Guidelines	Uniform Guidelines on Employee Selection Procedures 1978 (US)
UK	United Kingdom
US	United States of America
VPA	Virtual Personal Assistant
VP	Vice President
…	In dialogue = a pause
[…]	In dialogue = missing dialogue, not included
[ ]	In dialogue = added word for clarity, not spoken by interviewee

List of Figures

FIGURE 1: REPRODUCED REJECTION EMAIL - EXAMPLE 1 111

FIGURE 2: REPRODUCED REJECTION EMAIL - EXAMPLE 2 111

FIGURE 3: REPRODUCED FROM BRAUN AND CLARKE, 2006, P. 87 117

FIGURE 4: REPRODUCED REJECTION EMAIL - EXAMPLE 3 122

FIGURE 5: REPLICATION OF DEMO AVATAR, CREDIT: JOANNA RIVERA-CARLISLE 190

FIGURE 6: REPLICATION OF ASSESSMENT RESULT IMAGE 1, CREDIT: JOANNA RIVERA-CARLISLE 248

FIGURE 7: REPLICATION OF ASSESSMENT RESULT IMAGE 2, CREDIT: JOANNA RIVERA-CARLISLE 249

List of Tables

TABLE 1: VENDOR COMPANY AND INTERVIEWEE CODES AND ROLES 97

TABLE 2: VENDOR COMPANY AND INTERVIEWEE CODES AND ROLES 100

TABLE 3: VENDOR COMPANY DEMOS AND FREE ASSESSMENTS 105

TABLE 4: RECRUITMENT TECHNOLOGY CONFERENCES ATTENDED 110

TABLE 5: LAW AND REGULATION EXPERTS: PSEUDONYM/NAME AND INSTITUTION 113

TABLE 6: LAW AND REGULATION EXPERTS: ROLE AND EXPERTISE 269

CHAPTER 1: INTRODUCTION

1.1 The rise of AI recruitment technology (AI-rec-tech)

"Why am I un-hireable? […] Is there something wrong with me? Am I broken?"

These are the words of Kyle Behm, spoken in the documentary *Persona: The Dark Truth Behind Personality Tests* which aired in 2021 (Hawkins, 2021). At this point in the documentary, we've heard some of Kyle's story. In 2012, Kyle applied for a part-time job at his local supermarket in Nashville, Tennessee. He was recommencing his studies at Vanderbilt University after a leave of absence, due to his suffering with bipolar disorder. The job application required Kyle to take a personality test, developed by Kronos, a workforce management company based outside of Boston. But after completing the test, Kyle wasn't called for interview. His friend, who had worked at the supermarket, made some enquiries and told Kyle he'd apparently been 'red-lighted' by the assessment (O'Neil, 2016). The test had indicated that he was likely to ignore customers if they were upset or making him upset. In the documentary, Behm says, "it's not fair that by answering honestly about things that were related to my mental health, I was excluded from work" (Macabasco, 2021). This happened recurringly to Kyle – he couldn't seem to get a job interview; all the companies were using similar tests, and he was constantly being 'red-lighted' (O'Neil, 2016).

Tragically, Kyle passed away in 2019 while the documentary was still being filmed. His story is an important one, and he leaves a powerful legacy in bravely communicating his

experience: there must be more investigation into the rapidly developing hiring mechanisms which dictate the course of millions of people's lives and livelihoods.

Over the past two decades or so, there has been a particularly prominent change in terms of how hiring takes place: the rise of AI recruitment technology (henceforth AI-rec-tech). Up until the late 1990s, 'analog recruiting' (paper CV and cover letter followed by a face-to-face interview) often sufficed. Following this there was the rise of 'digital recruiting 1.0' (online job boards), succeeded by 'digital recruiting 2.0' (aggregated job boards e.g. Indeed, and digital professional and social network platforms, e.g. LinkedIn and Facebook). Now, we have entered 'digital recruiting 3.0', which has introduced AI into the hiring process (Black & van Esch, 2020).

According to much of the literature, there are four different types of hiring technology which echo the progressive stages of a job application (Ajunwa & Schlund, 2020; Black & van Esch, 2020; Bogen & Rieke, 2018; Bursell & Roumbanis, 2024; IFOW, 2020; Rab-Kettler & Lehnervp, 2019; Raghavan et al., 2020). AI can be used to make recommendations or predictions concerning the most suitable candidate by utilising human-based inputs to abstract such perceptions into models manually or automatically (OECD, 2019). Algorithmic techniques such as neural networks, machine learning (ML), and speech/image recognition can be used at different stages of the recruitment pipeline in the following ways, for example:

- **Sourcing.** This includes systems which scrape data from platforms like LinkedIn or Facebook and match or recommend candidates to jobs, or jobs to candidates. This might also take the form of systems which use neural networks to write job

descriptions, some of which try to make language as neutral as possible to promote a more diverse applicant pool.

- **Screening.** This includes systems which screen CVs using machine learning.[1]
- **Assessment/Interview.** Assessments might take the form of games, exercises, or tests which utilise behavioural science, psychometrics, or neuroscience. Algorithmic techniques such as decision trees or machine learning can be used to profile a candidate and compare this to a desired profile, or the client's top performers. This can be used to recommend which candidates might be the most suitable. Some systems might entail analysis of a pre-recorded asynchronous video interview through image and speech recognition, where an algorithm will discern skills and/or personality traits through vocal tone, facial expression, and transcribed answers.
- **Selection.** This might include algorithms which examine a candidate's criminal history or social media utilising image recognition, for example, to aid with the final decision.

The use of AI in hiring is becoming increasingly commonplace (O'Neil, 2016); there is a growing global market for this technology (Köchling & Wehner, 2020; Sloane, 2023). In 2022, the AI recruitment market size was valued at $540.4mn, and is projected to grow to around $590.5mn by 2030, exhibiting a compound annual growth of 6.9% during the forecast period (Dhapte, 2024). While it is difficult to know the exact scale of AI-rec-tech usage globally, largely due to a lack of transparency from employers, it was estimated in

[1] There is sometimes cross-over between how scholars understand 'screening' and 'assessment' technology. Some companies offer a 'screening' service by using assessments to screen out candidates (Ajunwa & Schlund, 2020), and therefore some scholars refer to 'screening' tools as those which include assessments (Bogen & Rieke, 2018; Bursell & Roumbanis, 2024). I use the terms as outlined here, keeping the two distinct.

2020 that 98% of Fortune 500 companies use AI or data-driven systems in at least one of these stages of hiring (IFOW, 2020).

The Covid-19 pandemic also escalated the use of this technology even further, as companies sought ways to stay competitive in the changing economic climate and accommodate working from home (Forman et al., 2020; K. Reynolds, 2021; Seiler, 2021; Sloane, 2023; Sonderling et al., 2022). For example, a survey in the US conducted by Gartner on HR leaders in 2020 (n = 334) found that 86% of companies were using virtual interview technology to interview candidates during the pandemic (Gartner, 2020).

While there will always, to an extent, be ambiguities around how largely black-box AI-rec-tech systems work – although I aim to demystify aspects of this, to some extent, throughout the thesis – their place in the market, in terms of what they provide to employers, is less complicated. It is no surprise that, within a decade where gender and racial equality has risen to the forefront of peoples' minds, for example through movements such as #MeToo and Black Lives Matter, that the use of these systems has surged. It is more important than ever for employers to be perceived as considerate of equality and diversity, equality and inclusion (DEI), not only to avoid scrutiny from the public, their investors, and/or their board, but also so that they avoid potentially costly discrimination lawsuits and all the bad press that entails. Therefore, these employers might see that they're able to utilise and advertise the narrative that AI-rec-tech systems allow them to 'reduce' or 'remove' bias from assessments by giving equal weight to each candidate, ignoring gender, race, and other primary sources of bias and therefore "neutralizing human mistakes and biases" (Hmoud & Varallyai, 2019, p. 25, also see Black & van Esch, 2020). As will be shown in Chapter 6, it is likely that employers see

these systems as a key to be ‘compliant’ with anti-discrimination regulation, which is appealing, for example, to a company without a large legal team.

Moreover, their use yields more immediate financial incentives for employers. The efficiency of AI-rec-tech systems means that employers can reduce time to hire, decrease cost to hire, and remove the manual work so that HR teams can supervise larger strategic initiatives (Niehueser & Boak, 2020; Ryu, 2019; Upadhyay & Khandelwal, 2018). Generally, employers also believe that these systems select best-fit candidates who will drive innovation and growth for the company (Altemeyer, 2019; Kantar et al., 2018; Michailidis, 2018). Therefore, AI-rec-tech systems appeal to small, medium, and large companies – all of whom are subject to market-pressures in a time where most of the globe is on the precipice of recession. For example, this would appeal greatly to a venture capital funded start-up who need to scale rapidly to return to their investors and wish to achieve market dominance (Shestakofsky, 2024).

Therefore, the use of AI-rec-tech is often presented by employers and vendor companies as a win-win – improving diversity and providing companies with recruits who match their mission (Allal-Chérif et al., 2021; Altemeyer, 2019; Mondragon, 2023; Upadhyay & Khandelwal, 2018).

However, multiple academics have countered this, arguing that AI-rec-tech systems increase power (and information) asymmetry between workers and employers (Ajunwa & Greene, 2019; Crawford et al., 2019) and are likely to amplify and perpetuate bias and discriminatory treatment of certain groups (Drage & Mackereth, 2022; Kelan, 2023; Raghavan et al., 2020; Tilmes, 2022; Vassilopoulou et al., 2022). For example, the data

used to train these systems might simply reproduce ideas of 'fit' in the workplace (Halpin & Smith, 2019; Raub, 2018) especially if the training data is selected to replicate the most successful employees (Ajunwa & Schlund, 2020). We see this in the example of Amazon's screening tool, which was discontinued in 2018. Using ML, the tool had been trained on successful resumes over the past ten years and had learnt to rank lower resumes which contained the word 'women's', due to males being ranked as the best performers in the training data (Tambe et al., 2019).

However, often, the claims on both sides are grounded in exploration of publicly available data or a theoretical analysis of the technology. There has been a rise in the use of data-driven algorithms to assist employers with recruitment, but unfortunately our understanding of how these systems work has not grown in tandem.

Generally, there has been a lack of empirical exploration into AI-rec-tech, and a recurring call for research concerning the thinking which underpins AI-rec-tech design (Dencik & Stevens, 2021; Köchling & Wehner, 2020; Pan et al., 2021; Stone & Dulebohn, 2013; Tambe et al., 2019). As Dencik & Stevens (2021) note, "we know little about how the actors developing and providing these technologies define and frame the tasks they claim to be fulfilling" (p. 3662). Sloane et al. (2022) have, similarly, said that there should be more social research into the "narrative roots" of what these systems rank and measure, because this can help us to understand "constructs that are then operationalised" (Sloane et al., 2022, p. 5).

Without such insight into the assumptions and epistemologies underlying these systems (Sloane et al., 2022) we cannot understand where to direct our focus for future research,

and whether inequalities arising from these systems might be a result of, or despite, vendors design decisions. Additionally, all productive debates should offer space for multiple opinions and, here, it is those of the vendors which I explore, with an aim to provide balance and nuance to the current conversation on bias in AI-rec-tech.

This thesis therefore offers the first in-depth, qualitative, empirical exploration of how AI-rec-tech vendors understand bias and how they aim to address this bias. Therefore, it fills the gap in existing work and answers the call for more research into the intersection of AI-rec-tech vendor companies and processes of design surrounding inequalities and biases (Sloane, 2019a; Sloane & Moss, 2019). In doing so, this thesis contributes not just to the literature on these technologies, but it also contributes to our societal knowledge concerning the design of the systems and the rationale behind this design. This is important; AI-rec-tech has a real impact on people's lives – people just like Kyle – but so far, it has been starkly under-researched.

1.2 Framing gender bias in the design of AI-rec-tech

In this thesis, I analyse the way(s) in which eight international AI-rec-tech vendor companies understand gender bias and how they aim to address it in their system design.

Throughout the thesis, some of these companies are anonymised and some are named. During my negotiations with companies, it was clear that while some were happy to be named (for marketing or transparency reasons), others wished to be anonymised, and some were unsure. My solution was to allow AI-rec-tech vendor companies to opt-in to being named within this research. This was done via a signed form after they had received

a full final draft of the thesis around six weeks before thesis submission. This solution enabled flexibility and was invaluable in terms of access; it meant I could cater to the wants and needs of the companies. Moreover, this enabled me to comply with copyright, which I discuss further in Chapter 3.

To explore how these companies understood gender bias and how their systems aimed to address the phenomenon, I use qualitative online interviews with both AI-rec-tech vendors and regulators, walkthroughs and autoethnography of the assessments, and content research on conferences, vendor company websites, vendor company reports, and regulatory/legal documents. Through this data, I answer the following research question(s):

RQ: How do AI-rec-tech vendors frame gender bias in hiring?

> **Sub-RQ#1:** How do AI-rec-tech vendors *conceptualise* gender bias in hiring?
>
> **Sub-RQ#2:** How do AI-rec-tech vendors attempt to *operationalise* addressing gender bias in their system design?

This illustrates how frames are constituted by both understanding (sub-RQ#1) and action (sub-RQ#2).

Specifically, I focus on assessment/interview AI-rec-tech systems.

I chose to do this for four reasons. Firstly, this allowed me to narrow the scope of the project and discuss systems which are similar to one another in nature. Secondly, assessment systems are underexplored in comparison to sourcing and screening

technologies, which have been critically reported on or studied with regard to how they can perpetuate existing gender and racial bias through facial recognition (Knight, 2021), ML-based text analysis (Dastin, 2018), and gender-based ad-delivery (Lambrecht & Tucker, 2019). I felt that these topics had already been covered and wanted to address the lack of research on assessment systems specifically. Thirdly, I anticipated that the assessment aspects of the systems would exhibit a wealth of unidentified design which would carry interesting intersections with issues of bias. Fourthly, I decided against focusing on selection systems. From my initial research, these systems seemed to be used largely to scan social media or criminal history of shortlisted candidates. This would have focused my thesis less on design of hiring systems and more on how these systems relate to platforms and/or the criminal justice system.

These assessments have emerged from a long tradition of using cognitive or behavioural tests to determine someone's 'fit' for a job. As Raghavan et al. (2020) note, in the early 1900s, the idea that assessments could reveal cognitive abilities and behaviours gained traction in Western industrial and academic circles, for example through the use of Myers-Briggs or the Big Five personality traits. This was used within workplaces, for example, aptitude and behavioural testing and training was used during World War I in German civil society to increase the productivity of a depleted workforce (Duttlinger, 2022).

However, cognitive assessments have been thought to impose adverse impact on minority populations since their inception, and critics have long contended that group differences in outcomes indicate flaws in the tests themselves (Raghavan et al., 2020). There is a long-standing caution about the potential lack of racially and ethnically diverse norms, especially if these assessments are normalised on white, middle-class, English-language

samples (Walters, 2021). This is commonly known as the Cultural Test Bias Hypothesis (CTBH), which encompasses the way in which people from diverse groups perform differently on mental tests due to inherent biases in the test through a flawed psychometric methodology (C. R. Reynolds et al., 2021). While many AI-rec-tech systems supposedly don't use Myers-Briggs, particularly after recent backlash concerning its origins and use for hiring, some of them do use models such as the Big Five to formulate their tests (Drage & Mackereth, 2022).

While the exact nature and functionality of each AI-rec-tech assessment system varies, most of these systems utilize tests through multiple-choice assessments, gameplay (e.g. puzzles or video games), simulations, or exercises to measure the skills, abilities, or behaviours of candidates (Fabris et al., 2023). For video interviews, candidates record answers to questions and this video can be algorithmically analysed with the aim to assess their skills and personal qualities.[2] These tests can generate a substantial amount of data which aims to capture cognitive skills, behaviours, decision-making processes, and situational awareness, measuring qualities such as problem-solving, creativity, system thinking, and critical thinking (Kantar et al., 2018; Raghavan et al., 2020). Candidates' results are often compared either to the score of an exemplary employee or a purposefully designed job profile which can be specified by the employer or by the vendor. Often, the assessments are customisable and can be adapted to the client's data or specifically named job requirements, and if not, they use human expertise to determine the competencies most relevant to the job, for example through IO psychologists or subject experts (Raghavan et al., 2020).

[2] It might also be that these systems simply facilitate synchronous video interviews, where there is no algorithm attached.

Sometimes, companies exclude humans entirely during this process, allowing the systems to dictate who proceeds to the next or final stage. Other companies require that humans look over the applications and choose who advances, using the assessment scores and measurements as guidance.

These small variations in system design complicate analysis slightly; AI-rec-tech assessment/interview systems cannot be referred to as one consistent phenomenon. However, this research - while recognising that every AI-rec-tech company, system, and team differs regarding their ideas on bias and the major affordances of the technology – analyses these systems in aggregate. I chose to do this because the data showed evidence of consistent trends, themes, and even design techniques, across systems.

In addition to focusing on assessment/interview systems, I concentrate on gender bias. Therefore, this work sits within a much broader field of scholarship concerning the wider ethical implications of AI.

From my initial probing of AI-rec-tech vendors' websites – 'bias' seemed to be a frequently used term. I found it was much more commonly used by these companies than terms such as 'inequality', 'discrimination', or 'fairness' when they were advertising the social value of their products. The dominant use of the term made me recognise the pertinence of exploring it's meaning in relation to system design; it is clearly a large part of how vendors think, design, and market their systems. Moreover, as 'bias' is often construed as a largely technical term (Zhou et al., 2022), this allows vendors to frame it as an issue which can be 'fixed', rather than drawing attention to social factors such as

‘inequality’, ‘injustice’, or ‘discrimination’, which are largely out of reach from their direct influence.

Moreover, I chose to concentrate on gender bias largely because of my positionality as a young, white, middle-class female scholar; this being a topic important to me. And, of course, this focus also allowed me to narrow the scope of the project. However, as I flag throughout the thesis, many of the issues uncovered in relation to ‘gender bias’ are also relevant to biases in relation to other protected characteristics.

To analyse how vendors understand and design systems in relation to gender bias I use Goffman’s theory on ‘frames’ (Goffman, 1974) to illustrate and unpack vendors’ understanding (their conceptualisation) of gender bias, and how this is transposed to system design (their operationalisation). In *Frame Analysis,* Goffman discusses frames as shedding light on how people interpret and reconstruct reality, using the concept to observe and theorise about how people organise and structure reality and experience (Persson, 2022).

Framing also provides me with analytical provocations to evaluate the understanding and design separately and in relation to one another, for example through Goffman’s concepts such as ‘breaking frame’, ‘out-of-frame’, and ‘keying’. These enable me to analyse the ‘vulnerabilities’ of the frames, which, for Goffman, is about frames being transformed. Based on this, I define vulnerabilities as complexities within or surrounding the frame which impact its integrity or function. These vulnerabilities are certainly not always due to faults of the vendor; sometimes they arise because these systems are part of a wide and unequal social ecosystem, which encompasses a range of societal inequalities.

While I do also outline the strengths of these frames, I focus more on the vulnerabilities. This enables me to discern the action and future research needed to ensure bias is addressed as effectively as possible. In general, this highlights the limitations of these technologies against a backdrop which so often portrays them as the solution to human inadequacies, while also shedding light on the fact these issues of bias are not fixable solely through system design and will require broader structural change.

As Goffman pointed out, during any one moment an individual can apply several frames at once (Goffman, 1974). I interpret that across all eight companies there are three dominant frames of gender bias in hiring. These three frames are:

(1) The intrapersonal frame (Chapter 4) – gender bias is *understood* in relation to the inequalities related to human preferences and perceptions. Vendors attempt to *design* human biases out through anonymisation, re-defining fit, and bias reminders. I draw attention to the vulnerability of the frame through the analogy of 'breaking frame' (Goffman, 1974), which sheds light on the inconsistency between vendors' understanding of bias and their ability to address it through system design.

(2) The structural frame (Chapter 5) – gender bias is *understood* in relation to structural injustices in education and opportunity. Vendors attempt to account for these injustices in the *design* by neutralising questions and assessing skills/behaviours instead of qualifications. I analyse the vulnerability of this frame to 'out-of-frame activity' (Goffman, 1974), which highlights the external or unknown factors which impact the function of the system design.

(3) The statistical frame (Chapter 6) – gender bias is *understood* in relation to statistical parity between demographic groups regarding success. Vendors *design* to test this through impact assessments and audits. Using Goffman's concept on 'keying' (1974), I highlight that this frame is understood differently by vendors and regulators and therefore vulnerable in practice.

Of course, while I cannot say that these frames apply to all AI-rec-tech vendor companies, both within or outside of this sample, they still show prominent trends in thought within these companies and beyond.

While I discuss these frames discretely, I argue that they are not mutually exclusive. Indeed, I make the case that while they can be held separately, they are intrinsically linked, and that overall, when held together, they present a socio-technical perspective of bias. Gender bias, or indeed bias in general, therefore, cannot be addressed solely through technical alterations to design, auditing, and law or regulation, but will also require a broader systematic and societal shift.

It is important to note that this thesis does not argue 'for' or 'against' the use of AI-rec-tech systems, but instead recognises them to be an inevitable part of our society and our workplaces. In addition, as I recognise throughout, vendors – contrary to popular opinion – do attempt to reflect upon and address issues of gender bias, and often display complex and thorough understandings of the term. Therefore, I aim for this thesis to be a constructive contribution, analysing what lies behind the design and noting areas where action and future work might improve the impact of AI-tec-tech on the most vulnerable in our society.

1.3 Key contributions

In focusing on how AI-rec-tech vendors frame gender bias and how this is mobilised in system design, there are several key contributions of this thesis. I outline these only briefly here and provide more detail on the contributions of this thesis in Chapter 8.

The key contributions of this thesis are:

- **Rare empirical data:** to date, most work on AI-rec-tech has used publicly available data to explore themes of bias, fairness, and discrimination (Drage & Mackereth, 2022; Raghavan et al., 2020; Tilmes, 2022). While this has been crucial for mapping the field, this work contributes vital empirical work concerning system design. AI-rec-tech companies are extremely difficult to access, so obtaining this data is significant.
- **Widely applicable methodological concept:** in Chapter 3, I present a new methodological concept for qualitative research on elites which entails access issues: 'the hustle'. This shows how access issues impact research and researcher which, in turn, enables reflection of solutions and ways around these issues (Collett, 2023). Moreover, this helps to capture how and why research which encounters access issues can still be rigorous and meaningful.
- **Adaptation of frame theory:** this thesis is the first piece of work to adapt and apply Goffman's theory on 'frames' to AI design. This fresh use of theory contributes a new way of analysing the vulnerabilities and strengths of system design in relation to specific concepts such as bias.

- **Taxonomy of gender bias in AI:** this thesis provides three frames of gender bias as perceived by vendors: the intrapersonal frame, the structural frame, and the statistical frame. This contributes to future conversations around bias and AI and is open to addition and revision.
- **Short-term action recommendations:** in Chapter 7, I lay out a set of action recommendations which emerge from analysing the vulnerabilities of the frames. These surround design, regulation/law, auditing, claims, and marketing.
- **Long-term future research agenda:** the detailed future research agenda, set out in Chapter 8, will enable a more tailored, targeted concentration for ongoing research on the design, use, and impact of these systems.

Overall, this thesis supplies an important contribution in the endeavour to better understand what lies behind the design of AI-rec-tech, particularly in relation to gender bias. This is significant, due to the wide scale of system usage and the impact they have on people's lives.

1.4 Definitions

There are multiple terms used throughout this thesis which, here, I define and justify.

To refer to the systems I am studying, I use the term 'AI recruitment technology' (AI-rec-tech). I have coined this previously (Collett, 2023) and, similarly, Tilmes (2022) has used 'AI recruitment'.

AI-rec-tech differs from terminology used by other scholars to describe similar technologies. For example, in the past, these systems have been described as ‘automated decision systems for recruiting’ (Sloane, 2023), ‘algorithmic recruitment systems’ (Bursell & Roumbanis, 2024), ‘algorithmic hiring’ (Raghavan et al., 2020), or ‘automated hiring platforms’ (Ajunwa & Greene, 2019; Ajunwa & Schlund, 2020).

I decided against using ‘automated decision systems’ because, as I will show in Chapter 4, these systems do not make the decisions, but are used in a hybrid manner with humans, who generally make the decisions guided by system recommendations. Moreover, all the systems in my sample use AI to some extent, meaning that simply ‘automated’ or ‘algorithmic’ didn’t capture their nature as precisely. I felt these terms were too narrow to capture the broad and complex techniques for data capture and usage in the systems.

Within my terminology of AI-rec-tech, I define AI as follows:

Artificial intelligence (AI): Broadly speaking, I follow the OECD definition of AI systems, which reads, “[a]n AI system is a machine-based system that can influence the environment by making recommendations, predictions or decisions for a given set of objectives. It does so by utilising machine and/or human based inputs to: i) perceive real and/or virtual environments; ii) abstract such perceptions into models manually or automatically; and iii) use model interpretations to formulate options for outcomes” (OECD, 2019, p. 20). More specifically, these systems use algorithmic techniques such as natural language processing (NLP), machine learning (ML), speech or image recognition, neural networks, deep learning, or decision trees (Broussard, 2023; Charlwood & Guenole, 2022; R. Iqbal et al., 2020; Miller, 2022).

In addition, I use ‘recruitment’ to communicate the range of stages that these systems can be used for (from sourcing to selection) and define it as the process which helps to determine whether a candidate is suitable and desirable for a specific employment. I use ‘recruitment’ and ‘hiring’ interchangeably in this thesis, and, indeed, ‘hiring’ could just as easily have been used in place of ‘recruitment’.

‘Technology’ was chosen for its broad encompassment of the application of knowledge for a particular purpose, allowing me to account for many different types of system. But similarly, ‘system’ could just as easily have been used, and I use ‘technology’ and ‘system’ interchangeably throughout this work.

Moreover, throughout this thesis I refer to the ‘design’ of AI, which I define as follows:

AI design: the devised feature(s) of an AI system that can be (1) experienced (seen or interacted with) on the front/back end of the AI system by users or subjects, or (2) invisible to users but constitute or influence how the system functions under the hood, e.g. decisions around algorithms or data.

This definition is broad and offers flexibility – design can constitute anything from the appearance of avatars to the wording of questions on tests, to the data chosen to train algorithms, to the impact assessments which influence and feed back into the questions or algorithmic design. This enables me to explore multiple design mechanisms which vendors use to address gender bias.

Next, it is crucial that I define gender. I centre the definition of gender around power and socialisation, so that it might be used intersectionally to capture other protected characteristics such as race, class, and ethnicity. Experiences surrounding protected characteristics are therefore not treated discretely (Crenshaw, 1989) but rather race, gender, sexuality, nationality are all seen to intersect (D'Ignazio & Klein, 2020). These intersections "remind us that oppression cannot be reduced to one fundamental type, and that oppressions work together in producing injustice" (Hill Collins, 1990, p. 18). The intersectionality at the heart of my definition of gender means the findings of this thesis are relevant to other demographic groups.

Therefore, my definition of gender is influenced by those scholars who focus on power distribution and social construction such as Butler (1990, 1993, 2011), Foucault (1991), and Guyan (2022):

Gender: a socially and culturally constructed facet of identity formed through power dynamics which shapes and maintains normalised appearances, roles, and relationships (Butler, 2011). In binary terms, this construction is normalised as female and male and is usually aligned with sex assigned at birth through genitalia. However, gender does not have to be restricted to this binary conceptualisation; it can be something between or beyond these concepts and does not have to conform to the sex assigned at birth or on legal documentation (Guyan, 2022). Gender, therefore, is socially learned and alterable (Hopkins, 2009).

In terms of the stakeholders discussed in this thesis, it is important for me to define both 'vendors' and 'clients/employers', which are used commonly throughout:

Vendors: companies which design, market, and sell AI-rec-tech systems.

Clients/Employers: these words are used interchangeably to describe companies which buy and implement AI-rec-tech systems.

In this thesis, I do not include companies which design recruitment technology which they then use solely for their own purposes to recruit their own employees (and therefore, these companies are simultaneously vendors and employers), although these companies do exist.

Lastly, I do not define 'gender bias' or 'bias' upfront in this thesis; I use gender as the qualifier to primarily discuss bias with the vendors. While in Chapter 2 I discuss how gender bias has been defined previously in relation to AI and hiring, I have chosen not to define it rigidly here. The purpose of the thesis is to explore the definitions and understandings of the vendors, rather than to impose a desirable definition which should be met.

While I don't define it precisely, I do assume throughout the thesis two properties of gender bias: (1) it is a **negative phenomenon**, and therefore not desirable, and (2) addressing or countering bias seeks **diversity of demographics (class, race, ethnicity, gender),** not just diversity of thought. The narrative of 'diversity of thought' can be weaponised to justify demographically homogenous workplaces, as has been in the past by the claim that cognitive diversity – in terms of how people feel, think, and act - is 'deep diversity', whereas demographic diversity is 'surface diversity' (Chamorro-

Premuzic, 2018). Demographic diversity is crucial to bringing about meaningful diversity of experience, background, thought, and to represent a multiplicity of identities in workplace cultures, decisions, and values (Bastian, 2019). This is not just important for organisations, but crucial for inclusion, belonging, and parity of treatment between individuals and groups. These properties were consistent with the views of the vendors, who all spoke about bias as a negative phenomenon and something which doesn't facilitate demographic diversity.

1.5 Chapter outlines

The chapters of this thesis proceed as follows.

Chapter 2: Literature Review

The literature review aims to contextualise the thesis by outlining the areas of research to which this work contributes and by demonstrating how existing literature has helped to discern my research questions. I also lay out my theoretical framework, which centres around an adapted version of Goffman's 'frames' (1974).

Chapter 3: Methods

Content from this chapter has been published in an altered form as: Collett, C. (2023) 'The Hustle: How Struggling to Access Elites for Qualitative Interviews Alters Research and the Researcher', Qualitative Inquiry, pp. 1–13. doi: 10.1177/10778004231188054.

This chapter outlines my data sample, sources, and methods, which included qualitative online interviews with both AI-rec-tech vendors and regulators; walkthroughs and autoethnography of the assessments; and content research on conferences, websites, reports, and regulatory/legal documents. I discuss my decision to use a combination of framing analysis and thematic analysis, aided by NVivo, to analyse the data. Next, I explain my concept of 'the hustle', which details the struggle I went through to access AI-rec-tech companies and how this changed the nature of the research and me as a researcher.

In the second half of the chapter, I outline several considerations which accompany my methods, including reflexivity and positionality, data sufficiency, and data configuration. I also explore interview dynamics such as power, rapport, and honesty. Lastly, I discuss the ethical considerations of this research.

Chapter 4: The Intrapersonal Frame

Chapter 4 is the first chapter summarising and analysing my findings. In this chapter, I argue that the first way in which vendors frame gender bias in hiring is 'intrapersonal' – they understand it to be intertwined with the human mind.

First, I answer sub-RQ#1 for this frame. I show that they conceptualise human perception as biased due to three intrapersonal tendencies: pattern recognition, homophily, and subjectivity.

Second, I answer sub-RQ#2 for this frame. I outline how the vendors aim to address this form of gender bias by 'designing out' elements of human perceptual biases through features such as anonymisation, re-defining fit, and bias reminders.

Third, in the discussion section, I argue that while the vendors show awareness and effort to address human perceptual biases, there are vulnerabilities in the form of 'breaks in the frame' (Goffman, 1974): inconsistencies between the conceptualisation of gender bias and the ability to fully address this in the technological design of AI-rec-tech. These breaks therefore emerge where human perception does, or can, infiltrate within hiring decisions. This means that their conceptualisation of gender bias can never be fully addressed, and so the frame is broken. I argue that this is inevitable, given that some of these cracks can be avoided and some are almost impossible to avoid due to the necessity of human choices up and down stream.

Chapter 5: The Structural Frame

Chapter 5 is the second chapter summarising and analysing my findings. I argue that the second way in which vendors frame gender bias in hiring is 'structural' – they understand it to be related to injustices in education and opportunity in society.

First, I answer sub-RQ#1 for this frame. I argue that vendors define bias as inequalities in opportunity but see their systems as allowing people to break free of this due to hard work, ambition, and innate skill, which resonates with narratives of 'the American dream'.

Second, I answer sub-RQ#2 for this frame. I demonstrate that vendors operationalise addressing this type of bias through attempting to account for structural injustice in design. They do this in two ways in particular: neutralising questions and assessing skills/behaviours (rather than qualifications).

Third, in the discussion section, I utilise Goffman's notion on 'out-of-frame activity' (Goffman, 1974) to highlight that this frame in design is vulnerable due to uncertainty on whether and how measuring skills and behaviours might correlate with background, education, and opportunity. Additionally, I note that this frame assumes that these systems are measuring the skills they claim to measure and argue that more transparency and research is needed on the underlying 'science' and training data of these tests.

Chapter 6: The Statistical Frame

Chapter 6 is the third and final chapter summarising and analysing my findings. I argue that the third way in which vendors frame gender bias in hiring is 'statistical'. Mostly, this chapter applies to the US-based companies.

First, I answer sub-RQ#1 for this frame. I describe how vendors conceptualise gender bias through the statistical idea of adverse impact, generally measured by the 4/5ths rule from the Uniform Guidelines on Employee Selection Procedures (1978), itself based on Title VII of the 1964 Civil Rights Act.

Second, I answer sub-RQ#2 for this frame. I demonstrate that this statistical measurement is operationalised through impact assessments to calculate ratios between different demographic groups, which is usually conducted on accumulated client data.

Third, in the discussion, I use Goffman's idea of 'keying' (Goffman, 1974) to demonstrate that vendors exhibit a different understanding of this legislation to the regulators. While vendors understood themselves to be 'EEOC compliant', regulators said there is no such possibility; Title VII applies only to the employers. Additionally, while vendors perceived current law and regulation to be somewhat insufficient, the regulators tended to view the laws as sufficient, but pointed to issues around enforcement.

Chapter 7: Discussion

In this chapter, I summarise the three frames of gender bias which I have discussed. Additionally, I re-visit how this addresses my research question(s) and discuss five further points:

1. *Notes on the nature and characteristics of the three frames.* First, I discuss the applicability of frames to other protected characteristics, the fact that there were no major gaps in the vendors' conceptualisation of bias, and that these multiple frames allow vendors to 'frame-shift' if one frame is failing (Lorino et al., 2017).
2. *Common 'vulnerabilities'.* Second, I explore the way that certain vulnerabilities apply to more than one frame.
3. *'Laminating' frames.* Third, I describe the theory of frame 'lamination' (Goffman, 1974), considering how more than one frame might be active at once. I conclude

that the frames can be separate but are also deeply related, sequential, and, together, form a sociotechnical view of gender bias.

4. *Frame consistency.* Fourth, I address the fact that these frames were generally consistent throughout vendor companies.
5. *Action recommendations.* Lastly, I consider how the frame vulnerabilities highlight short term and immediate action that can be taken around design, law/regulation, auditing, and claims/marketing.

Chapter 8: Conclusion

The conclusion ties the thesis together by summarising the argument and how the research questions have been answered. It also assesses the contribution of the thesis, limitations of the thesis, and maps out an agenda for future work.

CHAPTER 2: LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 Introduction

In this chapter, I review the literature surrounding gender bias, both in relation to hiring and to AI. This research crosses a range of disciplines, particularly psychology, sociology, business studies, economics, and gender studies. Broadly speaking, I demonstrate the importance of these two areas of research and emphasise how my work aims to unite them to study a topic which sits at the cross-section, but which is seldom given the attention it deserves: AI-rec-tech and gender bias.

To further validate the exploration of gender bias in AI-rec-tech in this thesis,[3] I review the existing literature on AI-rec-tech. Again, this covers a range of disciplines from HR studies to science and technology studies (STS), to law, psychology, and sociology. I discuss how, so far, work on AI-rec-tech has assessed the motivation(s) for system usage, candidate perception, law, and bias. Zooming in on the last of these areas – bias in AI-rec-tech – I demonstrate that, so far, there has been limited empirical exploration of vendors, and none in relation to vendors' understanding of gender bias and how they aim to address it through system design. I discuss how this presents a significant gap in the research which my project aims to fill.

[3] While I do still refer to these systems as 'AI-rec-tech' in this chapter, in some cases in the literature these systems are referred to using different terminology such as 'algorithmic hiring' or 'automated hiring platforms' (as mentioned in the introduction). In these cases, I always adopt the terminology utilised by the author, so as not to impose my terminology onto the work of others.

In addition, in this chapter, I expand upon my theoretical framework. I outline the theoretical and ontological foundation of my thesis, the 'social shaping of technology' (MacKenzie & Wajcman, 1999), which sheds light on the mutual relationship between society and technology. Moreover, I justify my use of Goffman's theory on 'frames', and the sociological terms of 'conceptualisation' and 'operationalisation', which I use to structure the discussion of this thesis. I demonstrate how this theoretical framework shapes my research questions, and how these questions have also emerged from the gaps in the literature reviewed.

Overall, this chapter contextualises my research, discerns the precise gap which this thesis aims to fill, and shows how I have shaped my research design and research questions. In addition, it briefly summarises the contribution of this project in relation to the literature. Largely, this centres around adding empirical work concerning how vendors understand gender bias, how they aim to address it in their systems, and the novel application of using Goffman's theory on 'frames' to study and analyse AI design.

2.2 Uniting work on gender bias in relation to hiring and AI

In this section, I review the existing literature on gender bias in relation to two key areas: hiring and AI. This provides a broad contextualisation for my work. In addition, it shows the pervasiveness of gender bias in both hiring and in AI, therefore highlighting the value in marrying these topics to study AI-rec-tech in relation to gender bias. As I will show in the section following, gender bias in AI-rec-tech is a topic which is seldom given the attention it deserves.

2.2.1 Gender bias in hiring

In approaching this research, it was crucial to reflect upon literature concerning sources and patterns of social inequality. Within psychology, sociology, and business/organisational studies, human biases (including gender biases) have been defined and discussed in a variety of ways, both in the context of hiring and more generally speaking.

Biases are sometimes labelled as 'implicit' (operating outside of awareness and intentionality) or 'explicit' (operating within conscious awareness and sometimes intentionally) (Espinosa & Ferreira, 2022; Greenwald & Krieger, 2006; Régner et al., 2019; Wynn & Correll, 2018). This dichotomy might also be referred to as 'unconscious' and 'conscious' biases (Agarwal, 2020; Njoto et al., 2022), or indeed as 'subtle' (comfort with one's own in-group, exclusion and avoidance of out-groups) or 'blatant' biases (hot, direct, unambiguous) (Fiske, 2002).

Along with these terms which outline the nature of bias, there are also many specific labels for bias which cover a range of causes and manifestations. This includes 'correspondence bias' (drawing inferences about a person's dispositions from their behaviour which is really down to a particular situation) (Gilbert & Malone, 1995), and 'cognitive bias' or 'judgement bias' (human cognition produces representations distorted in comparison to objective reality) (Haselton et al., 2015; Kahneman et al., 2021). Black & van Esch (2020) also discuss 'anchoring bias', when information unduly influences our judgement, along with 'similarity bias', when we favour those similar to us, and

'confirmatory bias', when people only consider information which confirms their initial judgements. This has also been referred to as 'conclusion bias' by Kahneman, Sibony and Sunstein (2021), when we seek to reach a particular conclusion because of 'confirmation bias' or 'desirability bias', meaning we will tend only to consider particular evidence which we believe to be true.

As I show throughout the thesis, some of these specific terms around the nature, causes, or manifestation of bias certainly resonate with the vendors' interpretations of gender bias in hiring. All in all, the multiplicity of qualifiers shows that bias has long been seen as a complex and multi-faceted term, and that it takes many different forms. It is because of this complexity that we must not assume the intricate details of how people interpret the term. In the context of technologies which aid or carry out decision-making, therefore, it is crucial that we investigate the assumptions these systems are built upon, so we know how they try to address them.

Having said this about the nature, causes, and manifestations of bias, as I mentioned in the introduction, I do assume bias to be a negative phenomenon. Not only did this fit with the views of vendors, but it fits with common consensus in the literature.

Whichever way it arises or manifests, bias, generally, is theorised in psychology and sociology as an unfair or unjust treatment or outcome for an individual or group, resulting from particular – often unjustified – human assumptions. These assumptions can be related to protected characteristics such as gender, race, ethnicity, religion, class, and so on. Bias has been previously defined as, "reacting to a person on the basis of perceived membership in a single human category" (Fiske, 2002, p. 123), and therefore gender bias

is thought to occur when "widely held beliefs about gender affect how men and women are evaluated in achievement-oriented contexts such as school and work" (Wynn & Correll, 2018, p. 509).[4]

Much existing literature assumes that bias is a negative phenomenon, with many complexities regarding its definition. On top of this, these biases are viewed as playing a crucial role in shaping our society and, as part of this, our workplaces. Across many disciplines, from economics to business studies, it has been emphasised that there are inequalities between men and women in the workplace. As noted in the UNESCO et al. report (2022), women around the world are paid less (ILO, 2019; Ortiz-Ospina & Roser, 2019; World Economic Forum, 2021), occupy fewer senior positions (Catalyst, 2020; UN Women, 2021), hold more precarious jobs (European Parliament, 2020; M. C. Young, 2010), spend more time undertaking unpaid child and elderly-care jobs, and participate less in STEM fields (Bustelo et al., 2019).

Hiring is a crucial part of shaping workforces around the world. It is an overarching apparatus of control in terms of diversity and culture (Halpin & Smith, 2019), and so these inequalities between men and women in the workplace are often linked to biases in hiring, alongside inequalities in education, societal norms, and workplace culture or belonging, which effect things such as promotion and retention.

[4] While it is generally perceived as a negative term, biases can be seen as a phenomenon which favours *or* disadvantages people. For example, as Hmoud & Varallyai (2019) define bias, it is "when an individual or group receives a favourable or disadvantaged treatment based on their characteristics such as age, colour, education background, ethnicity, religion" (p. 23). In other words, bias might be seen as unfair treatment, or as the equivalent of affirmative action to redress an imbalance. However, in this thesis I find that all vendors perceived bias as a negative phenomenon. Indeed, as seen above, this is a common way of interpreting the term and is, therefore, to be expected.

Biases in hiring are therefore a crucial topic for discussion – these decisions shape workplaces and society more broadly, contributing to the outputs of organisations and economic and social equality. Due to this, there have been attempts to re-structure hiring practices, including nudging decision-makers to be more accountable, changing the language used in marketing materials (Nandigama & Shyamsunder, 2021), and unconscious bias training, which has, unfortunately, been found to have minimal effect on hiring decisions (Dobbin & Kalev, 2016; Isaac et al., 2009).

In the workplace – and concerning hiring specifically – literature across psychology, sociology, and gender studies, which discusses this relation to bias, is split into two broad categories: gender stereotypes and homophily. Broadly speaking, these are the two main ways in which biases are seen to manifest to impact hiring and produce inequalities.

Firstly, gender stereotypes. Implicit and explicit gender biases at work are seen to be linked with gender stereotypes when it comes to hiring, promotion, and treatment in the workplace (Dunham, 2017; Luzadis et al., 2008; Régner et al., 2019; Whysall, 2017; Wynn & Correll, 2018).

Madeline Heilman defines gender stereotypes as "generalizations about the attributes of men and women" (Heilman, 2012, p. 114), and divides them into descriptive and prescriptive gender stereotypes. Descriptive gender stereotypes are argued to create a perceived 'lack of fit' between the attributes women are thought to possess, and the attributes thought necessary for success in traditional male positions. Prescriptive gender stereotypes are argued to create normative standards for behaviour that induce disapproval and social penalties when they are directly violated or when violation is

inferred because a woman is successful (Heilman, 1983, 2001; Wynn & Correll, 2018). It is argued to be this lack of fit and these normative standards which promote gender bias in the hiring context as well as at other stages of organisational decision-making, for example surrounding pay, promotion, and role (Heilman & Caleo, 2018; Wynn & Correll, 2018). Subjectivity in decision-making has been theorised as one way for stereotype-based negative performance expectation around lack-of-fit and role congruity to influence decisions and result in biased evaluations (Heilman & Haynes, 2008).

Gender stereotypes, therefore, are often related to assumptions and patterns around 'fit' between a candidate and a job. Eagly and Wood (2012) argue that stereotypes frame women as communal, warm, cooperative, and caring. Men, in contrast, are often stereotyped to possess agentic qualities such as assertiveness, competitiveness, and independence (Heilman, 2001; A. Li et al., 2017). This characterisation of women as communal conflicts with the perceived demands of many male gender-typed jobs and fields, promoting perceptions of women's incompetence in those areas and causing penalties when they exhibit counter-stereotypical attributes and behaviours (Heilman et al., 2024).

It has also been argued that gender stereotypes are related to the perceived fit of certain genders to certain jobs and sectors, and this is often referred to as 'occupational stereotypes' (Dutz et al., 2022; Martinez, 2022). This is what Eagly & Karau (2002) might also refer to as 'role congruity theory', which links bias to congruence between stereotypes held about job requirements and stereotypes about gender. The argument goes as follows: the greater the incongruence between stereotypical gender traits and the gender stereotype of a job, the greater the gender bias (Koch et al., 2015). For example, female dominated fields (such as education and care) are stereotyped as more communal

(Dutz et al., 2022). In contrast, men are more often associated with innovative behaviours at work (Luksyte et al., 2018), leadership positions (Dennis & Kunkel, 2004; Eagly & Karau, 2002; Martell et al., 1998; Wille et al., 2018), and within certain fields such as STEM (Carli et al., 2016; Schmader, 2023).

Therefore, gender stereotypes are thought to play a role in promoting gender bias, resulting in discrepant outcomes for women and men in the workplace and in hiring or promotion (Heilman et al., 2024).

Alongside theory, there is some evidence supporting these ideas. Koch et al. (2015) did a meta-analysis of gender stereotypes and bias in experimental simulations and found that men were preferred for male-stereotyped jobs, whereas no strong preference was found for female-stereotyped jobs. Overall, the research reviewed showed that people who are pursuing jobs that have stereotypes consistent with their own gender will be evaluated more positively than those inconsistent with job stereotypes (Koch et al., 2015).

In addition, some studies have shown that when candidates violate stereotypical norms, women and men are penalised in terms of how they're ranked at CV stage (Friedmann & Efrat-Treister, 2023; González et al., 2019; Hoover et al., 2019). This has also been exhibited at selection stage where negative biases have been found against women being evaluated for positions traditionally or predominantly held by men (Isaac et al., 2009). A meta re-analysis of 70 studies across 26 countries which look at how fictitious applicants are measured based on gender found that in relatively better paying occupations, which are dominated by men, the effect of being a woman is negative, while in the relatively lower paying occupations, which are dominated by women, the effect is positive. Therefore, this also shows the maintaining of the status quo concerning gender distributions and the earnings gap (Galos & Coppock, 2023).

Secondly, homophily. It has often been noted that – generally - humans tend to "recruit in their own image" (Singh, 2002, p. 3). This is sometimes referred to as 'similar to me' bias, relating to the implicit tendency of humans to feel a strong degree of affinity with and trust of others who we categorise as similar to ourselves (Whysall, 2017). This might also be referred to as 'similarity bias' (Black & van Esch, 2020) or 'affinity bias' (Heidari et al., 2023). While not always referred to specifically in relation to gender, much of this research refers to homophily or 'similarity bias' in relation to gender, race, ethnicity, and other protected characteristics. Given that I am treating gender intersectionally, I see this as deeply relevant.

Recruitment is often discussed as displaying bias which stems from the human longing for similarity. In *On Being Included,* Sara Ahmed speaks about how, in majority white societies, "[r]ecruitment functions as a technology for the reproduction of whiteness" (Ahmed, 2012, p. 39) by creating the idea of the institution and therefore what it imagines as the ideal. She discusses how this reflection, this 'likeness', can create comfort, satisfaction, ease. In contrast, discomfort involves a failure to fit.

Sociological literature labels this phenomenon as 'homophily': the tendency of individuals to associate or bond with those similar to them. Homophily literally means 'love of the same' (Lazarsfeld & Merton, 1954). This is a major concept in network studies and denotes how humans prefer to be with people similar to themselves rather than different so that their core beliefs aren't challenged. This similarity is theorised to breed connection and to structure networks of every type – marriage, friendship, work, support networks, and so on (Mcpherson et al., 2001).

It has been argued that interviewers tend to favour candidates who are more culturally similar to them, including in ways such as gender, race, educational background and physical appearance (Kahneman et al., 2021). A study done by Edo, Jacquemet and Yannelis (2019), for example, argued that ethnic homophily was present in the hiring discrimination process – it found that recruiters with European names were more likely to call back French-named applicants. Higgins and Judge (2004) also discussed the role of similarity in the effect of applicant influence-tactics in hiring. They found that ingratiation - the use of specific actions such as opinion conformity and favour-doing - had a strong, positive effect on perceived fit and recruiter hiring recommendations. Moreover, this is found in relation to gender specifically, where studies have found in-gender-group preference during hiring decisions (M. Carlsson & Eriksson, 2019; Friedmann & Efrat-Treister, 2023).

This literature shows the prevalence of biases in our workplaces, and, in doing so, demonstrates the importance of studying bias to address it as a phenomenon. Human-based decision-making can contain biases regarding stereotypes or homophily, and these biases have consequences. Recognising this, it is vital to consider afresh the nature, causes, and manifestation of bias in hiring within the context of emerging technologies. As will be shown in section 2.3, gender bias in AI-rec-tech is a topic which is underexplored.

In addition, this literature provides context for discussions around human biases when it comes to considering vendors' understandings of the term. As I will show, many of these ideas have been considered in the design of these systems. Not only do the vendors' understandings align with this literature, but they develop and contribute to it, too.

2.2.2 Gender bias in AI

Studies and research on gender bias are not limited to human behaviour, but also address technological instantiations. Particularly, this research occurs within sociology, gender/race studies, computer science, and HCI. Overall, while some argue that AI design and implementation can encode inequity through explicitly amplifying bias and discrimination (Benjamin, 2019; Buolamwini & Gebru, 2018; Noble, 2018) others believe AI is capable of reducing or even eliminating bias (Mondragon, 2023; Upadhyay & Khandelwal, 2018).

Like gender bias in hiring, bias in AI systems is often defined as the unfair treatment or outcome for individuals or groups (Masiero & Aaltonen, 2020; Rovatsos et al., 2019). But in this context, it is due to, or facilitated by, the use of complex algorithmic decision-making (Rovatsos et al., 2019).

Further to this, bias in AI is often divided into:

- **Technical bias:** concerning the data input, design, or output of the AI system, for example, technical biases are spoken about in relation to ‘data bias’ and ‘algorithmic bias’ (Zhou et al., 2022). A common way of dealing with this is through audits (Hall & Ellis, 2023).
- **Social bias:** AI is seen to reflect societal structural inequalities which exist in our culture and language (Sartori & Theodorou, 2022). This is thought to stem from algorithmic design and human-assembled training datasets, and many believe it can be tackled by improving the diversity of AI teams (Hall & Ellis, 2023).

This division between technical and social bias is often discussed in literature on bias in AI (Hall & Ellis, 2023; Lopez, 2021; Ulnicane & Aden, 2023).

Similarly, the taxonomy of Friedman & Nissenbaum (1996) outlines the three types of bias in computer systems, which, they argue, can systematically and unfairly discriminate against certain individuals and groups. They label these as pre-existing bias (e.g. has its roots in social institutions, practices, and attitudes which exist prior to the system), technical bias (arising from technical constraints or consideration e.g. from limited technology, hardware, software, de-contextualised algorithms, or the formalisation of human constructs to computers), and emergent bias (occurs in the context of use because of new societal knowledge that isn't incorporated). As I shall argue later, these divisions resonate with the views of the vendors, who exhibit both social and technical understandings of bias.

Overall, existing work on gender bias in AI, very broadly, is related to two characteristics: stereotyping and exclusion. These bodies of literature show the importance of studying gender bias in AI. Moreover, they are crucial for contextualisation; as will become clear, these themes arise in conversations with vendors about gender bias in AI-rec-tech.

Regarding stereotyping, AI has been shown to incorporate and reinforce gender roles and stereotypes (Hopkins, 2009) and this has been linked with bias. This has particularly been highlighted in relation to virtual personal assistants (VPAs) such as Siri, Alexa, and Cortana. The gendering of these AI devices as female has raised concerns over gender bias in the real world, particularly around aggression towards women and the

encouragement of the stereotype that women conduct most of the domestic chores and caretaking duties (Bergen, 2016; UNESCO, 2019; Woods, 2018). VPAs have been argued to "transfer to the digital realm the gendering and stratification of labour found in the real world, with working women principally confined to jobs of lower power, status and pay, often within service industry roles" (Dillon, 2020, p. 10). There has also been concern that it might spread gender bias through normalising the tolerance of sexual harassment and verbal abuse because the systems "give deflecting, lacklustre or apologetic responses to verbal sexual harassment" (UNESCO, 2019, p. 106).

Along with VPAs, the word embeddings within ML and NLP systems have often been thought to exhibit male/female stereotypes which can amplify biases (Bolukbasi et al., 2016). Leavy (2018) discusses how gender ideologies are embedded in language and are still embedded in text sources and therefore in ML algorithms. For example, it was found that Google Translate, when fed gender neutral languages, exhibited a strong tendency towards male defaults in fields typically associated with unbalanced gender distribution or stereotypes such as STEM roles (Prates et al., 2019).

Gender bias in AI has also been discussed in relation to exclusion. For example, if the data is skewed or labelled incorrectly, this might exclude certain demographic groups from being recognised or counted. On ImageNet, 45% of the data comes from the US and only 3% from India and China, this means a photograph of a traditional US bride is labelled 'bride', 'dress', 'woman', whereas a North Indian bride is named as 'performance art' (Zou & Schiebinger, 2018). Evidently, the exclusionary biases – as demonstrated by this example – do not just affect women, but also intersect with ethnicity. It is people from non-Western countries who are being excluded from accurate

descriptions due to these data disparities. Similarly, Gorska & Jemielniak (2023) argue there is gender bias in AI-generated images of professionals in law, medicine, engineering and science, because it was found that men were represented in 76% of the images and women only in 8%.

Another example is facial analysis/recognition and image labelling services which might perform consistently worse on transgender individuals (Scheuerman et al., 2019), showing that these systems consistently operationalise gender in a trans-exclusive way and treat gender as immutable and physiological, "essentialising the body as the source of gender" (Keyes, 2018, p. 11) and excluding individuals from datasets and/or system design. As has been argued previously, this exclusion is important to note, but it is not just that trans people, for example, need more inclusion and representation – it is broader than this (Rincón et al., 2021). There is a need for more thought concerning privacy, safety, and participatory design. In the interest of bearing in mind the intersection between gender, race, and other protected characteristics, it is important to note that facial analysis systems have historically performed worse on racial minorities. For example, a study evaluating three commercial gender classification systems (MSFT, Face++, and IBM) showed that the darker-skinned females are the most misclassified group, with error rates of up to 34.7%, whereas the maximum error rate for lighter-skinned males was 0.8% (Buolamwini & Gebru, 2018).

Literature thus reveals that gender bias is apparent both in human behaviour and the technologies we build to replace or augment human activity. The topic of gender bias in AI is an important one to consider – AI systems are capable of reflecting unequal treatment in the real world and perpetuating it through widespread system usage.

This thesis therefore aims to unite these two pertinent topics: gender bias in hiring and gender bias in AI. It does this through an exploration of gender bias in AI-rec-tech. Indeed, it sees these discussions as crucial for reflecting upon the implications for minority and vulnerable groups in an age where the scale of AI usage is growing rapidly, and when our workplaces are changing as a result. AI-rec-tech plays a particularly important role in these changes and in maintaining power disparity (Halpin & Smith, 2019). As Sloane discusses, while automation and labour is often explored in terms of deskilling, surveillance, and exploitation, what has remained in the shadows is recruiting (Sloane, 2023).

Moreover, this thesis recognises the lack of work which has been done to unite these topics and study gender bias in AI-rec-tech, as I will expand upon in the following section. Currently, there is no academic work to date specifically on gender bias in AI-rec-tech. This thesis therefore offers a context- and system-specific discussion, drawing together these crucial and salient topics, but noticing their mutual lack of relation to one another.

2.3 The gap in research on AI-rec-tech

Above, I showed how this project aims to unite two bodies of literature to explore gender bias in AI-rec-tech. This is a crucial area for research; AI is increasingly being used in HR. The AI recruitment market specifically was valued at $540.4mn in 2022, and is projected to grow from $590.5mn in 2023 to $942.3mn by 2030 (Dhapte, 2024). However, there is currently a debt of attention being paid to the design of these systems.

In this section, I detail the current literature surrounding AI-rec-tech,[5] showing that it comes from a variety of disciplines including HR/business studies, STS, law, and sociology. I discuss four prominent areas of the literature on AI-rec-tech to date: motivations for use, candidate perception, law, and bias. This review allows me to assess where the gap appears in existing scholarship and enables me to evaluate where more research on gender bias and AI-rec-tech is needed for this body of work to progress. This is what Sandberg & Alvesson (2011) might refer to as research design through 'gap spotting' and specifically 'neglect spotting': identification of an under-researched area where there is a lack of empirical support.

I conclude that the existing scholarship on bias in AI-rec-tech lacks empirical work, especially surrounding the process of design. So far, there has been no work which explores, from the AI-rec-tech vendor company perspective, how gender bias is defined and how they attempt to address the phenomenon through system design.

2.3.1 Looking backwards: motivation for use, candidate perception, legal implications, and bias

Existing work concerning AI-rec-tech falls into four distinctive categories.

Firstly, there is a body of work within HR studies and business/organisational studies literature which explores why employers are motivated to use AI-rec-tech systems. This

[5] While some of this literature does look at specific types/stages of technology (which is always specified), other literature looks at AI-rec-tech more generally speaking. As noted in Chapter 1, I only study assessment/interview technology, but this broader literature on these systems is still relevant.

body of work tends to make the case for the use of these systems but does not approach them through a critical lens to consider their relation to social inequality and biases.

AI-rec-tech is often seen to provide major affordances to companies and HR managers. One of the greatest challenges for medium to large companies is hiring at scale (Adler, 2021) but this technology often provides a solution which is both time and cost effective. It is seen to remove the manual work from the hiring process, offering employers efficiency and accuracy (Bauer et al., 2011; Black & van Esch, 2021; Dickson & Nusair, 2010; Jia et al., 2018; Ore & Sposato, 2022; Ryu, 2019; Upadhyay & Khandelwal, 2018). For example, the use of an AI-enabled screening tool by L'Oréal made the time to review a resume decrease from 40 minutes to 4 minutes, a reduction of 90% (Black & van Esch, 2020). Overall, it is the view of many that "AI in recruitment allows faster and more accurate hiring" (Guenole & Feinzig, 2018, p. 12). Moreover, some argue that it can 'remove bias' and increase diversity of recruits, although they do not provide evidence to support this statement (Altemeyer, 2019; Hemalatha et al., 2021).

Some argue that time-saving allows those in HR to commit energy to more worthwhile investments, such as aligning the values of candidates with those of the company or promoting employee happiness and welfare (Allal-Chérif et al., 2021; Jia et al., 2018; Ryu, 2019; van Esch & Black, 2019). They claim that these systems mean that "[w]ith AI taking care of boring and repetitive tasks, recruiters now have more time to be creative and can focus on strategic issues" (Upadhyay & Khandelwal, 2018, pp. 255–256). Others assert this also frees up the time of HR professionals who can then focus on building and nurturing relationships with candidates (Guenole & Feinzig, 2018).

In terms of accuracy specifically, this is often discussed in terms of how systems help companies to select the best-suited candidate. As Hmoud and Varallyai (2019) discuss, the "goal of the recruitment and selection process is to provide the organization with necessary human resources at minimum cost with focus on core tasks and behavioural competencies needed to fulfil job requirements" (Hmoud & Varallyai, 2019, p. 22). AI-rec-tech is thought to help managers identify new employees with the greatest fit, potential for success, and suitability into the right team, which will in turn help their productivity (Jia et al., 2018; van Esch & Black, 2019). Therefore, this 'accuracy' in terms of candidate fit is seen to save companies a lot of money. Businesses are motivated to use AI or automated systems to increase retention of employees through better matching successful candidates – especially regarding behaviours and values - to the organisation (Allal-Chérif et al., 2021). Otherwise, hiring decisions can cost money through a loss of productivity and high turnover, which results in re-hiring and entails the expense of training a new employee for their post (Lahey, 2015). Using AI or automated systems to hire is thought to contribute to the development of employees who are highly devoted to their job, or 'best fit' (Michailidis, 2018). This is something that humans aren't always seen to be capable of doing, especially if, for example, there are hundreds or thousands of applications for each post (Hmoud & Varallyai, 2019).

On the other hand, though, there has been work which highlights the potentially problematic implications of motivations for using these systems. Some have argued that these affordances encompass power asymmetry between candidates and employers. Often, people don't have a choice about using these systems – they must apply for jobs in order to afford food and housing and other basic things for living (Anderson, 2017). These systems, some argue, are therefore continually dehumanising employees and

devaluing human characteristics that cannot be easily translated into data (Lee, 2018) and offering managers increased legibility of candidate/employee activity which produces information asymmetry and frames workers as fungible human capital (Ajunwa & Greene, 2019).

As is clear, there has been much work conducted on AI-rec-tech regarding the motivation for use. While it would have been interesting to explore how gender bias features in client's motivations for using these systems, and while this was initially something I wished to explore, I was limited by access to client companies. I discuss this more in Chapter 3.

Secondly, another prominent body of work concerning AI-rec-tech surrounds candidate perception of these systems. This research comes from management/organisational studies, STS, and psychology. While this work does consider AI-rec-tech systems in relation to how those subject to them feel and perceive them, it offers no practical steps on how system design and use might address these perceptions.

Some of these studies have focused on justice perceptions of using AI in hiring. For example, a sample of students (n = 153) experienced three types of screening technique. It was found that interactive voice screening was rated lower than face-to-face interviews and telephone screenings in relation to procedural justice factors including interpersonal treatment, two-way communication, and openness (Bauer et al., 2004). Similarly, it was shown by Acikgoz et al. (2020) through an online survey of MTurk workers (n = 320) that AI-based interviewing was generally viewed as less procedurally and interactionally

just than traditional human-based interviewing, because the ability to have a two-way communication was very attractive to applicants.

Regarding fairness, there were comparable findings. Lee (2018), also using MTurk participants (n = 228) who took an online survey, reported that partakers thought human decisions were fairer than algorithmic decisions for tasks involving human skills. They believed that the algorithms would ignore qualities which were hard to quantify and would lack human intuition. Participants expected the human manager would be able to identify top candidates based on "skills and experiences" or "merit," and would have the qualifications and authority to do so. In an ethnographic study of fairness perceptions of AI-rec-tech in a large multinational company, van den Broek et al. (2020) found that candidates often contacted them and said they didn't feel they'd been given a fair chance to prove themselves, complaining that they did not recognize themselves in the AI output they received upon completing the games. Furthermore, Zhang & Yencha (2022) explored the public perceptions of algorithms used in resume and video interview screening using a nationally representative survey in the US (n = 4245). They found that most respondents did not feel these systems were fair and/or effective.

Additionally, there have been multiple studies which have focused on the perception of AI-rec-tech by candidates in relation to other factors such as general preference, anxiety, or respect. Blacksmith et al. (2016) did a meta-analysis of 12 studies and found that interviewer ratings and applicant reactions are generally worse in technology-mediated interviews. They speculated that this could be because applicants might feel they were not able to demonstrate their social skills when impression management was restricted. A vignette study among young adults in Germany (n = 129) also found that candidates

negatively perceived the quality of AI assessments which seemed complex and intangible (Schick & Fischer, 2021). Concerning anxiety, using participants recruited through an online survey platform (n = 532), van Esch et al. (2019) found that candidates do generally feel anxious when AI is part of the job application, but that there are benefits for candidate anxiety to explicitly mentioning that the company uses AI-enabled recruiting tools. And concerning respect, one study (n = 446) found that human decision-makers were generally perceived better in terms of respectful treatment (Bankins et al., 2022).

Generally speaking, candidate perception research shows that recruitment technology, in a variety of different forms, tends to be perceived by candidates less favourably than human recruiters regarding justice, fairness, respect, and associated anxiety levels. This is important to consider, and it is crucial for future research to further explore how candidates perceive the presence and cause of gender bias in these systems. Ideally, this research would take a more participatory, co-design form, where candidates might be able to feed back into how systems are designed and used. Moreover, the above studies largely utilise MTurk workers or students as participants, and future research would benefit from broadening this sample.

Thirdly, there has been a lot of work on AI-rec-tech within legal studies. Overall, literature within law and STS has explored how existing legislation and regulation might be applied to AI-rec-tech, and, in doing so, it points out strengths or limitations. While this work offers crucial insight into the shortcomings of current legal terrain, it is missing an important piece of the puzzle in terms of how these systems are designed and how they work.

The majority of attention in this literature has been given to US law. Raghavan et al. (2020), for example, discuss how Title VII of the Civil Rights Act of 1964 forms the basis of regulatory oversight regarding discrimination in employment. The Uniform Guidelines on Employee Selection Procedures, issued in 1978, stem from Title VII by prohibiting discrimination with respect to protected attributes. This means that an assessment can be challenged on the grounds of disparate impact or disparate treatment (Raghavan et al., 2020; Raub, 2018). Disparate impact can be determined through the 4/5th rule, which requires employers to justify the use of apparently neutral hiring practices if these practices result in a selection rate for one demographic group which is less than four-fifths (0.8 ratio) of the 'most successful' demographic group (Kelly-Lyth, 2021). In addition to the Uniform Guidelines, employers could also be charged for discrimination using AI-rec-tech under Title I of the Americans with Disabilities Act of 1990 (ADA) and Section 503 of the Rehabilitation Act of 1973, both of which prohibit discrimination on the basis of disability in hiring and employment (Moss, 2021).

However, federal legislation is often criticised. For example, with the 4/5th rule, it has been noted that employers might be able to easily justify a lower ratio than 0.8 if they can outline a 'business necessity' (Ajunwa, 2021). Current employment and anti-discrimination laws are sometimes seen as inadequate to address the discrimination attributable to emerging workplace technologies (Ajunwa, 2021, 2023b).

Outside of the US, the EU AI Act is in the process of becoming the first risk-based regulation for AI systems. It regards AI for hiring, promotion, and pay decision-making as 'high-risk' (Charlwood & Guenole, 2022; Commission, 2024b). This will therefore

mandate extensive risk assessments before and after deployment within the EU internal market (Sloane, 2023). There has, however, been criticism that there is a lack of enforcement mechanisms, for example, how those affected will complain? And how will standardisation bodies with no fundamental rights experience write the rules for self-assessment (Veale & Zuiderveen Borgesius, 2021)?

Moreover, there have been comprehensive overviews of how AI-rec-tech systems intersect with UK law. Kelly-Lyth (2021) and Sanchez-Monedero et al. (2020) discuss how equality (UK Equality Act 2010) and data protection law (Data Protection Act 2018) can be applied to automated hiring. Both papers outline how the current legislation will not protect job applicants' rights, especially surrounding the lack of definition of 'legitimacy' in the Equality Act when it says that bias can be justified if there is shown to be a 'proportionate' means of achieving a 'legitimate' aim.

Overall, work from within legal studies generally assesses the gaps and shortcomings of existing and future law and regulation. There is certainly more work to be done within academia in relation to the law, particularly in relation to gender bias. However, given that the law is not my area of expertise, this was not an area on which I chose to focus this thesis.

Fourthly, there has been some work on AI-rec-tech and bias emerging from STS, HR studies, sociology, and gender studies. This work is crucial, but much of it has been based on publicly available data or is theoretical in nature. It has left a gap for more empirical work on gender bias in AI-rec-tech, specifically on studying the vendors' understanding of the phenomenon.

Using publicly available data, for example, Raghavan et al. (2020) assessed 18 vendors of algorithmic pre-employment assessments through analysing their websites, including their whitepapers, reports, and webinars, in order to study what claims are made regarding fairness and bias. Around bias claims, they found that 15 vendors made abstract references to bias and several claimed to be able to 'fix' bias when it appeared. They also found that some of these vendors said they were actively de-biasing algorithms or claimed to be naturally unbiased. While analysing these claims is important, it is necessary for research to delve deeper into these companies' structures to see what lies behind them – an area which my research addresses directly through interviews with vendor companies.

In contrast to the vendors' claims, many have argued that it is clear that these systems are not bias-free (Raub, 2018). There have been concerns surrounding bias in AI-rec-tech platforms, and this work has especially concerned sourcing and screening technologies. Most will be familiar with the 2018 case where Amazon discontinued their screening tool. Using ML, the tool had been trained on successful resumes over the past ten years and had learnt to rank resumes which contained the word 'women's' lower (Dastin, 2018). Therefore, there is widespread scepticism about the pre-defined standards of a 'good employee' or 'good fit' which these systems might pedal (Halpin & Smith, 2019; Raub, 2018). Lambrecht & Tucker (2019) investigated how algorithms delivered ads promoting job opportunities in STEM and found that fewer women than men saw the advert, largely because women are the more expensive advertising demographic. Therefore, there has been some work on the potential impacts of AI-rec-tech, but little specifically regarding assessment technology.

Theoretically speaking, in terms of what bias means in the systems, bias in AI or automated recruitment technology has been categorised by some scholars. For example, Fabris et al. (2023) discuss how bias conductive factors in algorithmic hiring are largely about stereotypes (beliefs about who is most suitable for the job, as discussed earlier) and sensitive attribute proxies (when demographic or background can be detected from other variables). Bogen & Rieke (2018) argue that bias can be embedded in these tools through data, for example through choices about labelling and structuring, inaccurate data, or unrepresentative data. They argue that this could embed social bias (interpersonal prejudice against marginalised groups), structural bias (legacy of historical disadvantages), or internalised biases (thought around whether one sees themselves as suited to a job, which might impact their actions). Similarly, Köchling & Wehner (2020) discuss how algorithmic decision-making might lead to discriminatory or biased treatment of certain groups of people because of inaccurate, biased, or unrepresentative input data, highlighting that current knowledge about the threats of unfairness and discrimination by algorithmic decision-making is mostly limited.

Moreover, some scholars have used theory to deconstruct the underlying aims and assumptions of these systems, arguing that, in fact, these technologies lend themselves to biased decisions. Vassilopoulou et al. (2022) argue that algorithmic systems used in HR have a pseudoscientific quality that prevents them from being scrutinised for bias; they carry the illusion that they're objective, scientific, and bias-free when, really, these systems could perpetuate inequalities through biased data, proxies, algorithmic specification of fit, segregation of individuals, and technical design. This pseudoscientific quality of AI-rec-tech systems is related to scepticism around the ability to infer emotion

and character from facial movements (Barrett et al., 2019) and human behaviour as it relates to job success (Ajunwa, 2021; Sloane, 2023; Sloane et al., 2022).

Gender, race, and disability scholars have also critiqued these systems and their ability to address bias. Tilmes (2022) discusses how de-biasing measures in these algorithms are likely to exclude data of people who might have disabilities effecting their facial expressions, movement, or speech differences, and this risks entrenching ableism and worsening outcome disparities in hiring. Drage & Mackereth (2022) also evaluate the systems utilising gender theory and critical race theory. They interrogate some of the general claims about bias mitigation made by AI recruitment companies through studying marketing and promotional materials. For example, the claim that these tools remove bias from the hiring process, provide a more objective assessment, promote a more diverse workforce, identify internal characteristics, and find the ideal candidate, all of which Drage & Mackereth argue are misleading. They draw attention to the fact that these attempts to 'strip' gender and race from AI recruitment systems misconstrue these factors as isolatable attributes. Instead, we must recognise gender and race to be ingrained in broader systems of power (Drage & Mackereth, 2022). This is crucial for realising the ways in which these protected characteristics intersect with one another and therefore thinking holistically about the impacts of these systems on a diverse population.

However, contrary to this, some literature in HR/business studies does argue that these systems are in fact more objective, claiming that AI-rec-tech systems 'reduce' bias or even 'eliminate' bias altogether (Florentine, 2016) and that they bring about no adverse impact between demographic groups (Mondragon, 2023). Upadhyay & Khandelwal

(2018) argue that because the technology can ignore gender and race, it gives equal weight to all candidates and strengthens the diversity of new hires.

Many also argue these systems are at least less biased than humans. Black & van Esch (2020), for example, contend that humans are beset with cognitive biases which limit the reliability and validity of their judgements. Allal-Chérif et al. (2021) also argue that the unconscious bias of a recruiter is potentially eliminated using AI-rec-tech, because while a human being is likely to be influenced by his or her personal interests, discriminate against certain candidates, or be guided by attraction, a computer system will optimize the results solely according to the needs of the company. However, like the above, all claims seem to be based on the potentiality rather than the realities of these technologies. This research is based on speculation and not upon empirical data.

Many of these critiques and counter-critiques already noted in existing literature on bias in relation to AI-rec-tech raise important points which continue to emerge as I explore vendors' understanding of gender bias. However, largely, research on bias in AI-rec-tech is speculative or theoretical, rather than based on empirical data. This pocket of research on bias in AI-rec-tech so far shows very little empirical work to back up any of these claims, and no work on gender bias specifically in the design of these systems. Moreover, there is no work specifically on assessment technology.

Naturally, therefore, this is where I aim to situate my work. My research aims to use empirical data to draw this debate back to its roots and explore the way(s) in which AI-rec-tech vendors understand gender bias and attempt to address it. This should lay the groundwork for future research to explore the impacts of these systems.

In focusing on gender bias in the context of design, I also address multiple gaps within the existing work on AI-rec-tech concerning client motivations, candidate perception, and law and regulation. My work offers a critical examination of how systems might function in relation to social inequality – a gap left by client motivation research; it offers practical steps on how system design might be altered – a gap left by candidate perception research; and it offers a missing piece of the puzzle from legal literature by offering some insight into how they work in relation to bias.

It is striking that there has been this plethora of work concentrating on 'bias' in relation to AI-rec-tech systems, rather than in relation to 'inequality', 'discrimination', 'fairness', or 'injustice'. This is likely due to the common use of the term 'bias' by vendors themselves. As mentioned in Chapter 1, when carrying out an initial review of AI-rec-tech websites and reports, I found 'bias' to be the primary way vendor companies framed the problem of gendered inequalities in hiring. This made it important for me to understand the meaning of the term 'bias', but the dominance of the term also carries underlying implications.

I suspect the common use of the term 'bias' by vendors - as something which they can address through their systems - appeals to their prospective clients. It enables the clients to advertise the social value of the system with the claim that it is 'bias-free' or that it 'reduces bias', which perhaps sounds more achievable than saying the system 'lowers discrimination' or 'tackles injustice'. Moreover, it likely appeals to them for legal and regulatory reasons. As I demonstrate later, bias is often spoken about as a signifier of

discrimination in hiring, and so reducing or eliminating bias might aid their rebuttal of any legal claims concerning discrimination.

In addition, bias is a largely technical term (Zhou et al., 2022), allowing them to frame it as an issue which can be 'fixed'. On the other hand, terms such as 'discrimination', 'injustice', 'inequality' are more social terms (Joyce et al., 2021; Sloane, 2019b), which carry admission of complications regarding structural and social factors out of their reach. However, as I show throughout the thesis, it is the social factors which are important to consider in tandem with the technical. For example, structural injustices surrounding education and opportunity, or discrimination stemming from stereotypes, homophily, subjectivity. However, the common understanding of 'bias' largely forecloses this fuller picture from view, instead framing it as an issue which can be addressed technically.

2.3.2 Looking ahead: empirical work on gender bias and AI-rec-tech design

So far, I have shown that, broadly speaking, this research aims to unite two wider bodies of literature on gender bias in hiring and gender bias in AI by studying gender bias in AI-rec-tech. This is especially important considering the rise of AI-rec-tech usage. Additionally, I have demonstrated that there is a lack of empirical work which critically explores the design of assessment AI-rec-tech systems in relation to gender bias. These systems are crucial to study because they impact people's lives and livelihoods, and so more attention must be paid to the technology companies shaping these systems (Raghavan et al., 2020).

Having said this, though, there is a small body of empirical work emerging within sociology and media and cultural studies which has engaged with AI-rec-tech

stakeholders including vendors and clients. I use this emerging work to further focus the subject of my own work.

Dencik & Stevens (2021), for example, interviewed eight data-driven hiring system vendors in the US and Europe about their systems to study the way they situate the provision of tools in relation to existing hiring practices. They found that the companies legitimise the use of these systems by outlining how existing practices are ineffective and unfair for candidates, and that data-driven knowledge was more objective. While this study is similar to mine in its approach, it does not delve into the design of the systems in relation to inequality or bias, something which my work aims to address. In addition, Sloane (2023) held 55 semi-structured interviews with predominantly low-volume recruiters and found that automated decision systems (ADS) continually reconfigure professional discretion, especially as they cannot see everything that's going on within the system. Sloane's study is vital to see how these systems play out in practice, and while I intended to work with vendors on how they interpret their systems to interact with issues of bias, I was restricted by access. Most recently, Bursell & Roumbanis (2024) studied one of Sweden's largest food retail companies and compared the outcomes of traditional (human-based) recruiting with that of algorithmic job recruiting. They found that, contrary to intention, algorithmic recruiting decreased diversity because of the unintended human-algorithmic interaction effect. Again, this work is cutting-edge and vital, and the first piece of work to show the potential impact of these systems on the demographic diversity of employees. While again restricted by access from this sort of study on employers themselves, I hope my work will lay the groundwork for continued and tailored research in this area.

All three of these papers have been crucial in gathering empirical data from vendors and clients of AI or automated recruitment technology and in showing how these systems make claims, reconfigure workplaces, and have the potential to decrease demographic diversity. Many themes from these papers re-emerge throughout thesis – for example, human interaction with systems in relation to bias (Bursell & Roumbanis, 2024) is noted by vendors in Chapter 4, as are claims about other alternative arguments which vendors use to justify their approach (Dencik & Stevens, 2021).

However, these three studies also continue to note the recurring gap in empirical work on these technologies, noting that there has been little reflection from the providers themselves, and "we know little about how the actors developing and providing these technologies define and frame the tasks they claim to be fulfilling" (Dencik & Stevens, 2021, p. 3662). Sloane also notes that "[d]espite the extensive recruiting ADS market growth […] the field of critical technology studies in hiring and employment is still nascent" (Sloane, 2023, p. 3). A similar message is communicated by Bursell & Roumbanis (2024), who discuss how the social impact of these systems has been given limited attention, and the contributions have been primarily non-empirical in nature, but given the potential effects on inequality and discrimination, studying the consequences of algorithm-based hiring is of tremendous importance.

There is still, therefore, much work to be done in providing empirical work on AI-rec-tech (Vrontis et al., 2022). This thesis aims to use empirical data on eight AI-rec-tech vendors to contribute to this emerging body of empirical literature on AI-rec-tech. It does this through exploring how vendor companies understand gender bias and how they

attempt to address it in their system design, something which previous work has not endeavoured to do.

It has been noted before that there is a lack of research which engages with vendors to unpack their definitions of terms such as bias, and how this underpins the logic of their product design (Köchling & Wehner, 2020; Stone & Dulebohn, 2013; Tambe et al., 2019). As Noble (2018) points out, the people who make the decisions surrounding the algorithms are important to explore so that we can think about the implications of these systems in more details and see which values shape their choices. Therefore, while this thesis does not demonstrate or prove the impact which these systems might have in terms of bias, it does aim to lay the groundwork to explore the impacts in a more targeted way. This should enable academics and industry actors to explore, together, the intricacies and impacts of this design in relation to bias.

This work is crucial, not only to bring vendors into the conversation from which they have so far been largely sequestered, but crucially, it is only when we know how these companies are defining and attempting to mitigate bias in the design that we can consider whether the impacts of these systems are arising because of, or despite, AI-rec-tech design.

To explore this topic, which was born from my literature review, I had to consider an appropriate way to discuss questions of how vendors think about and address bias in the design of their systems. Next, therefore, I discuss my theoretical framework.

2.4 Theoretical framework

In this section, I overview the theoretical assumptions and tools I use to study the topic outlined above. This theoretical framework, along with this gap in the literature, helps to construct my research questions.

Firstly, I discuss the underlying theoretical foundation to this work: the social shaping of technology (MacKenzie & Wajcman, 1999). This is the underlying epistemological assumption to my work concerning the relationship between technology and social understanding. Secondly, I outline the theory which I use to structure this study. I needed an approach which would allow me to capture the understanding of the vendors, show its translation to action, and analyse both stages. I explain and justify the use of 'frames' (Goffman, 1974) to study of this topic. Thirdly, I explain my adaptation of framing which encompasses 'conceptualisation' and 'operationalisation', which gives me a vocabulary to discuss the frames of the vendors in understanding and design respectively.

Generally, this section shows that my theoretical framework constitutes work from STS and sociology. Recently, there has been call for the increased application of sociological approaches for studying AI technologies (Joyce et al., 2021; Sloane, 2019a; Sloane & Moss, 2019). Overall, sociology provides a key to anticipating the possible social influences and social consequences of technology (Sartori & Theodorou, 2022; Sloane, 2019b). My work takes inspiration from these calls, and from Zajko (2022), who argues that in studies of bias, we should aim to move beyond the technical and think about the

social meaning of technology, and further than this, we should use this to make normative and practical messages from our findings.

2.4.1 Laying foundations: the social shaping of technology

Underlying this thesis is the assumption that technology and society have a dialectical relationship: aspects of technology can reflect ideas, inequalities, and ideologies present in society, and aspects of society can be shaped by technology (MacKenzie & Wajcman, 1999; Wajcman, 2009). It is because of this social shaping that we can study technological systems, such as AI, as they are 'knowledge machines' (Gillespie, 2014) which, in turn, can reveal aspects of thinking which have been incorporated into the design. Moreover, it is likely that this design will impact society.

If we consider gender specifically, itself an integral part of society, gender can be used as a lens to demonstrate this dialectical relationship between technology and society. As Hopkins (2009) points out, technological use, access, and design can reinforce a gender system or allow groups to subvert an existing gender arrangement. As we saw earlier, AI systems such as VPAs can embed restrictive and stereotypical gender ideologies which reflect societal views. This can be seen in the case of BIA, a female-gendered chatbot in Brazil's largest bank Bradesco. The bot received over 95,000 morally or sexually offensive messages in one year, including many regarding violence against women. However, when the responses became firmer and challenged this language, saying it was unacceptable, there was a decrease in offensive messaging (UNESCO et al., 2022).

Rather than a one-way system where technology influences societal changes, I employ the underlying assumption of the 'social shaping of technology' (MacKenzie & Wajcman, 1999; Wajcman, 2009):[6] technology and society shape one another. I demonstrate this theory throughout my work as I look to explore how designers are shaping technological systems around their own social views. Moreover, in the discussion section of each findings chapter, I consider what role technology might play in changing, reflecting, or reinforcing certain aspects of our society. Therefore, I do not assume that technology and society are separate spheres, but rather that they are, or can be, "mutually constitutive" (MacKenzie & Wajcman, 1999, p. 23).

This theoretical foundation can also be seen within my research design. This research aims to study how those people designing AI-rec-tech understand the concept of gender bias and embed these understandings into the systems' function. As Hughes argues, technological systems "are both socially constructed and society shaping" and because they are invented and developed by system builders and their associates, the components of technological systems are socially constructed artifacts (Hughes, 1987, p. 45). This is also argued by Forsythe (1993), who points out that in undertaking the selection and interpretation involved in designing and constructing expert systems, these knowledge-based systems, in a sense, replicate the makers perspectives. Gillespie (2014) notes, therefore, that we "must unpack the warm human and institutional choices that lie behind these cold mechanisms" (Gillespie, 2014, p. 169).

[6] In adopting these theoretical foundations I am not assuming a stance of technological determinism. Technological determinism encompasses the argument that technologies evolve autonomously due to their own intrinsic reasoning and this, in turn, impacts society (Papageorgiou & Michaelides, 2016; Peters, 2017; Wyatt, 2007).

As sociologists, we must study the designers of technology; they direct the visions and therefore the potential uses and impacts of emerging technologies. This project therefore aims to discern these underlying perspectives and think about the impact they have on the design and therefore on our society, as well as what they might reflect about our society.

2.4.2 Structuring the research: Goffman's 'frame analysis'

To study how AI-rec-tech vendors define and attempt to tackle gender bias through their system design, I required theoretical scaffolding which would allow me to structure this conversation and capture both the vendors' understanding and how this is translated into design. To do this, I construct an adapted version of Goffman's theory on 'frames'.

Applying the concept of frames to AI design is a novel contribution of this work.[7] 'Frames' has become a colloquial term which is used in many ways (Cukier et al., 2021; Fisher, 1997; Vliegenthart & van Zoonen, 2011). Frames have been used widely in communication studies, cognitive psychology, and news/media studies to analyse the dominant narratives communicated through news media (Sullivan, 2023; Vliegenthart & van Zoonen, 2011). While the concept of frames has been applied to news coverage of AI (Bunz & Braghieri, 2022; Cools et al., 2022; Luo et al., 2023) and to policy surrounding AI (Ulnicane & Aden, 2023),[8] thus far it has not been applied to study the design of AI technology.

[7] I use 'framing' to refer to the process of understanding and designing, and 'frames' as the interpretable result of this activity.

[8] Some of these uses of 'frames' do stray from the original Goffmanian definition. For example, in their paper Ulnicane & Aden (2023) use policy frames as put forward by Schön & Rein (1994), which sees frames as a way of making sense of reality and providing guideposts for analysing and knowing. Although, arguably, this takes inspiration from Goffman's usage.

However, I aim to take it back to its sociological roots while also creatively and innovatively adapting it for future work. This thesis makes the case that this adapted version of 'frames' is a valuable approach to studying AI design. The value lies in its flexibility, its ability to capture how understanding manifests in design, and its analytical tools and provocations which enable analysis of frame 'strengths' and 'vulnerabilities'.

'Frames' was a term initially used in sociology by Bateson (1955) who influenced Goffman (1974). In *Frame Analysis,* Goffman discusses frames as shedding light on how people interpret and reconstruct reality. He used the concept to discuss meaning-making and the organisation of interpretation in people's experience of the social world. His core aim was to "try to isolate some of the basic frameworks of understanding available in our society for making sense out of events and to analyze the special vulnerabilities to which these frames of reference are subject" (Goffman, 1974, p. 10). The subtitle summarises it all, argues Persson, as it reads: *An Essay on the Organization of Experience* (Persson, 2022); 'frames' allow us to observe and theorise about how people organise and structure reality and experience. These frames can be either individual or collective, or both. While the frames are generally consistent and collective, I recognise throughout that there are signs of differing individual and personal views in some cases, too.

Overall, I aim to explore the thoughts and understandings of vendors in relation to bias, and how these realities are structured. This enables me to look at one way of the bi-directional relationship between society and technology, allowing me to focus on how vendors' thoughts might be shaped by societal or personal realities and how these feed into the systems they design. In addition, it enables me to consider how this design might

reflect into society, and to direct future research which might discern this with more certainty.

While *Frame Analysis* is one of Goffman's less utilised pieces of work, as Persson argues, it lends itself to interpretation and adaptation because of its own uncertainty and messiness (Persson, 2022). I adapt Goffman's theory on frames to encompass how these social interpretations are transposed to system design. I therefore use frames as a key to investigate the construction of meaning and discourse (Björnehed & Erikson, 2018; Johnston, 1995), looking at their collective understandings, as well as how this discourse is mobilised in system design (Mendonça & Simões, 2022).

As Sullivan (2023) outlines, there are three levels of framing which focus on language, cognition, and communication: semantic framing, cognitive framing, and communicative framing. This thesis aims to look at all these levels, combining analysis of language used in communication (both public and from interviews with vendors) to interpret the thoughts of the vendors surrounding gender bias. Indeed, as Sullivan herself notes, we could not have communicative frames or semantic ones without cognitive ones.

This thesis, then, interprets frames within technological design to constitute:

(1) how vendors interpret social concepts – their 'frame in understanding', and,

(2) how vendors translate these interpretations into reality – the 'frame in design'.

Understanding and design together provide the frame. This affords us as sociologists with a fresh way to look at the relationship between our social worlds and whether or how this is transposed into technological systems.

Moreover, I needed a theory which would enable me to critically examine these understandings and how they are mobilised. I argue that discussion of frames prompts analysis of their strengths and vulnerabilities, as it has been argued this aspect of 'frame' theory is not made use of enough (Björnehed & Erikson, 2018). This vocabulary enables me to evaluate the understanding, the design, and the relationship between the two when thinking of the potential social influences and implications of AI-rec-tech in relation to gender bias. I use many of Goffman's tools to do this throughout the thesis, including 'breaking frame', 'out-of-frame activity', and 'keying'.

Goffman's idea of 'vulnerabilities' is defined as ways in which social activity can 'transform', causing alterations to frames. I translate this to the context of technological design, adapting this theory to define vulnerabilities as: complexities within or surrounding the frame which impact its integrity or function. This definition is broad because, in practice, these vulnerabilities can manifest in multiple ways. Crucially, it means that either the vendors' understanding, or their system design, is impacted by factors either within or outside of their control.

Vulnerabilities do not allow a definitive conclusion about the effects or impacts of the frames, but they allow an evaluation of the understanding and design in relation to external factors or in relation to one another. This provides scope for discussing the potential effects of the frame. In doing this, vulnerabilities highlight some of the pressure points which need further exploration in future research, and therefore point to the unanswered questions surrounding gender bias in AI-rec-tech.

While recognising these pressure points, I also aim, in some cases, to highlight the 'strengths' of the frames both in terms of the vendors' understanding of bias, and the potential of how this is addressed by design. Strengths highlight the opposite of vulnerabilities: factors which might reinforce the integrity or function of the frame. This might be in relation to understanding alone (for example, if a concept is supported by data/theory and is therefore well-grounded), or if design effectively addresses this concept through counteraction/mitigation. Therefore, while an understanding of gender bias can be strong in its own right, design can only be strong in relation to a strong understanding.

Of course, my own data does not enable me to conclude with certainty whether their design is strong, so I can only discuss the strengths of design in theory. However, I do often find that there are strengths in the vendors' complex and thorough understanding of bias. Overall, therefore, I focus more on frame vulnerabilities in this thesis. This is largely because these vulnerabilities point the way to action and future research which might aid reduction of gender bias surrounding AI-rec-tech further.

Although I do not use the term in this way, frames could, in theory, be applied to how users, or those subject to certain technologies, interpret systems. They could also be applied to study the impacts of these frames. In frames literature, these impacts are termed the 'frame effects', an umbrella term to discuss repercussions of certain frames on particular subjects (Björnehed & Erikson, 2018). These topics are not the focus of this thesis, excepting identification of the potential uses or impacts of these systems. However, as I outline in Chapter 8, framing effects will be important to consider within future work concerning AI-rec-tech and, indeed, this focus is the intended next step of this research.

In this thesis, therefore, I concentrate on how those who design AI-rec-tech systems frame gender bias. I study their institutional 'framing process' (Björnehed & Erikson, 2018) looking at their collective understandings and the interplay between this social meaning-making and actions concerning technological design. Overall, by using frames, I aim to highlight the nature, vulnerabilities, and, at times, the strengths of vendors' interpretation(s) of gender bias and how this is translated to system design. I hope this will shed light on how these understandings, embedded within technology, might reflect into society. In addition, emerging from these vulnerabilities, I aim to provide action recommendations and crucial directions for future research to mitigate any negative impacts of these systems going forward.

Therefore, this theoretical structure shapes the overall research question which this research aims to answer:

RQ: How do AI-rec-tech vendors frame gender bias in hiring?

My findings demonstrate how, across multiple companies, these frames are largely consistent. I show that across all eight companies there are three dominant frames which, while discrete, are not mutually exclusive. As Goffman points out, during any one moment, an individual can apply several frames at once (Goffman, 1974). These three frames, which I explore separately in Chapter 4, 5, and 6, and then bring together in the discussion, Chapter 7, are:

1. The intrapersonal frame
2. The structural frame

3. The statistical frame[9]

I argue that these frames are not mutually exclusive and show how, while they can be held separately, they also are intrinsically linked, sequential, and together, present a socio-technical perception of gender bias in hiring.

Lastly, it is crucial to note two caveats: firstly, these frames are my own interpretation of the vendors' frames; they are my framing of their frames. This includes interpretation about how vendors attempt to address certain understandings of gender bias through the design. While my own interpretation, I demonstrate throughout the thesis how this qualitative approach is valuable; it allows me to get closer to the interviewees points of view, and draws our attention to many interesting aspects of thinking and design which will be crucial to consider in future research into AI-rec-tech. Secondly, while these frames have emerged from studying these eight companies, this does not mean they can necessarily be translated to other companies. This is not to say they could not and do not resonate with other AI-rec-tech vendors, only that we cannot be certain of this.

2.4.3 Back to the building blocks: conceptualisation and operationalisation

Given that frames constitute 'understanding' and 'design', a suitable way to discuss frames is through the lenses of conceptualisation (the frame in understanding) and operationalisation (the frame in design). These terms are embedded in my sub-research questions:

[9] These frames become progressively less theoretical and more practical and exhibit interesting differences. For example, the first two refer more to sources of gender bias, whereas the third focuses more on measurement to define gender bias. I discuss this more in Chapter 7.

Sub-RQ#1: How do AI-rec-tech vendors *conceptualise* gender bias in hiring?

Sub-RQ#2: How do AI-rec-tech vendors attempt to *operationalise* addressing gender bias in their system design?

This terminology structures the discussion of how gender bias is defined by those designing AI-rec-tech and how bias reduction is measured or put into practice in the system design. In using these as my sub-research questions, it enables me to focus on the individual building blocks of this framing structure.

I borrow and adapt elements of some sociologists work to re-fashion the meaning of the terms conceptualisation and operationalisation in the setting of technological design.

Conceptualisation is usually referred to when forming theory in sociology (Alford, 1998; Edel, 1979; Merton, 1957). Generally, it is seen as an idea that, once tagged and generalized, can effectively guide inquiry into seemingly diverse phenomena. As Merton points out, "[a] good part of the work called 'theorizing' is taken up with the clarification of concepts – and rightly so. It is in this matter of clearly defined concepts that social science research is not infrequently defective" (Merton, 1957, p. 114).

In this thesis, I focus on conceptualisation as a process of naming and creating a definition for an understanding which would otherwise be abstract (Swedberg, 2014). Conceptualisation, therefore, captures the vendors' understanding(s) of gender bias, their cognitive naming process. This adopts Shoemaker, Tankard and Lasorsa's (2004) definition that a concept "describes a portion of reality" (Shoemaker et al., 2004, p. 2), and Swedberg's basis of conceptualisation as naming and creating a definition for our

observations in the social sciences in order to theorise (Swedberg, 2014). This is important, says Swedberg, because it gives the concept a new identity and makes it easier to see the phenomenon for what it is. However, one can't simply pin a name to a concept and not explain it; it requires being 'turned into a concept', and this happens by deciding what is essential to it. I aim to do this throughout the thesis by expanding upon what constitutes their conceptualisation(s) of gender bias.

Operationalisation is defined by Shoemaker, Tankard and Lasorsa (2004) as the form of experimental measurements which test particular variables which make up constituent elements of the concept. This is something also argued by Steidl and Werum (2019) who define operationalisation as the process by which social scientists create empirical indicators, for example through surveys, to measure broader (theoretically-based) concepts: "to operationalise means to develop measures for a concept" (Steidl & Werum, 2019, p. 1). In this thesis, I will not be using empirical measures when it comes to operationalisation. Here, I use the term operationalisation to capture the mobilisation of the vendors' concepts of gender bias. I am defining operationalisation, therefore, as how the companies themselves set out to measure or design addressing their conceptualisations of reducing gender bias, whether qualitatively or quantitatively.

The relationship between conceptualisation and operationalisation for the vendors is not always a conscious process. For example, they do not always label or define gender bias as [x] (conceptualisation) and explicitly say they mitigate it through [y] (operationalisation). It is my self-proclaimed job as the researcher to interpret this connection at times. For instance, even if not measured statistically, I ask, through

operationalisation, whether and how the company is translating the concept of gender bias into a method or an approach to operationalise its reduction.

Fundamentally, through using these terms, I am conceptualising their conceptualisations and operationalisations of gender bias. This is one way in which I depart from the above theorists, who see it as the theorist's role to conceptualise and operationalise research variables. Instead, while I do this in part and am aware of my own views being implanted within this research, I am primarily using these terms to represent their thinking and action and to get behind and beneath technological design to look at the way they define, label, name, and measure gender bias in their systems.

Overall, then, my theoretical framework consists of several parts. The foundation of the 'social shaping of technology' (MacKenzie & Wajcman, 1999) notes my assumption of the mutually reinforcing relationship between society and technology; my theoretical structure of 'frames' (Goffman, 1974) enables me to discuss the vendors' understanding and design and to analyse the frame; and 'conceptualisation' and 'operationalisation' aid me in outlining and discussing 'understanding' and 'design' of gender bias in these systems, and therefore studying the aspects of the frame in more detail.

2.5 Conclusion: research questions and literature contribution

In this chapter, I have progressed from a broad to a narrow review of the literature surrounding my topic. In doing so, I illustrated the gap in the literature which my research aims to fill. Broadly, this thesis aims to bring together literature on gender bias in hiring and gender bias in AI to discuss the important and under-researched topic of gender bias

in AI-rec-tech. More narrowly, work on AI-rec-tech shows a gap concerning empirical research, particularly around how vendors understand gender bias and how they attempt to address it in system design.

To explore this research topic, I discussed the theoretical framework which enables me to structure this work, and which helped to discern my research questions:

RQ: How do AI-rec-tech vendors frame gender bias in hiring?

> **Sub-RQ#1:** How do AI-rec-tech vendors *conceptualise* gender bias in hiring?
>
> **Sub-RQ#2:** How do AI-rec-tech vendors attempt to *operationalise* addressing gender bias in their system design?

The contributions of this thesis will be discussed more in Chapter 8, but there are several ways in which this research will contribute to the existing bodies of literature discussed:

- Broadly, this thesis draws together the literature on gender bias in hiring and gender bias in AI. These are important areas of research which will need renewed concentration given their merging in the rise of AI-rec-tech. This literature also helps to contextualise this research. This is demonstrated particularly in Chapter 4, which shows how the views of vendors resonate with literature surrounding stereotypes and homophily.
- Directly, this thesis contributes to work on AI-rec-tech and bias, and to a small but emerging body of work which uses empirical data to study these systems. It does this by filling a gap in this work; the focus on the rationale and understanding of vendors regarding bias is something previous work has not addressed.

- Indirectly, this thesis contributes to other areas of work on AI-rec-tech such as law (in Chapter 6) and, through assessing design, might shed some light on clients' motivations for use (in Chapter 4).
- Theoretically, it contributes to literature on Goffman's 'frames' by being the first piece of work to apply frames to highlight the social processes and thoughts behind technological design, and to use this to analyse the vulnerabilities and strengths of these frames.

In the following chapter, I outline my methods before embarking upon the empirical chapters which explicate my findings.

CHAPTER 3: METHODS

Note: content from this chapter has been published in an altered form as -

Collett, C. (2023) 'The Hustle: How Struggling to Access Elites for Qualitative Interviews Alters Research and the Researcher', Qualitative Inquiry, pp. 1–13. doi: 10.1177/10778004231188054.

3.1 Introduction

At the heart of my research questions and my theoretical framework is the objective to understand how those designing and producing AI-rec-tech systems frame gender bias. To explore this, I used several methods.

Primarily, I carried out qualitative, online, synchronous interviews with 18 employees across eight AI-rec-tech vendor companies. Naturally, to explore how these systems are conceptualised and designed, I needed to go straight to the source. I also examined the content from company websites and reports, as well as from industry conferences and/or webinars. In addition, with six vendor companies, I experienced demos (a walkthrough of the system by a company representative or detailed video, usually intended for prospective clients to explore how the system works) and/or completed the online assessment myself. Through demos and experiencing assessments, I gained familiarity with aspects of system design and saw how this matched up to what I had been told in interviews.

As I carried out this research with AI-rec-tech vendors, particularly in the US, I found that definitions of bias were constantly referring to the existing legislation on anti-discrimination. This preliminary finding led me to interview six experts from relevant governments or regulatory bodies in the US and EU. In addition, I analysed legislative or regulatory documents including the Uniform Guidelines based on Title VII of the Civil Rights Act 1964 (US), along with the UK Equality Act 2010, the GDPR (EU), the EU AI Act, and local state law in the US including New York, Illinois, and Maryland. This was not included in my initial research design, but as I discuss later, it was an extremely important avenue for me to investigate in relation to how gender bias was framed.

Given the precautions of AI-rec-tech companies surrounding IP and competition, alongside the ethical controversy concerning these companies and their systems, gaining access to vendors was incredibly difficult. As I discuss in this chapter, while the number of interviews doesn't seem large, for this kind of research, the sample size is a significant achievement, and the data is sufficient for me to answer my research questions.

In this chapter, I outline my methods of data collection and analysis, considering their complexities. In addition, I argue that while there was a significant struggle to access companies (a concept which I term 'the hustle') (Collett, 2023) the data is rich and pioneering in this field. This chapter presents not only a novel way of considering data access and sufficiency, but also analyses the dialectical relationship between the researcher and the research setting.

3.2 Data sample, sources, and methods

Here, I explain my data sources (vendors and regulators) and within that, each of the methods I used to collect data from these sources.

3.2.1 AI-rec-tech vendor companies

Overall, I collected data from eight AI-rec-tech vendor companies. I did this using several methods including qualitative interviews, experiencing free assessments or demos of the systems, and analysing their reports and websites. These methods matched my research questions, and the aim of my theoretical framework based on Goffman's 'frames'; they enabled me to explore how vendors were framing gender bias in hiring. Through interviews, I could hear their thoughts, test their rationale, and delve deeper into their understanding and the design of the systems. The assessments, demos, and reports/websites also allowed me to see how they were framing gender bias outwardly and enabled me to compare this material to what was said in interviews. I also attended online industry conferences surrounding AI-rec-tech so that I could analyse how this technology was being framed and discussed in the public sphere. I go into each of these methods in more detail below.

While each of these resources were aimed at different people – for example, the interviews were solely for me (and therefore my thesis and any academic papers), the conferences, demos, and websites primarily aimed at clients – I still found that messaging was consistent across the resources (perhaps because, arguably, this is all one outward-facing message), albeit that the interviews allowed me to delve deeper into design and their thoughts. I try to bear this in mind throughout the thesis.

As I discuss when outlining the concept of 'the hustle', access certainly shaped my research design. Nonetheless, I still aimed to use 'purposive sampling': a sample chosen by the researcher to include representatives from within the population being studied who have a range of characteristics relevant to the research project. This is also sometimes referred to as stratification, choosing layers or groups of people relevant to research. In this case, I used purposive sampling and stratification both in terms of the companies I approached as well as selecting the most relevant participants within companies (Gorman et al., 2005).

The vendor companies in my sample all design and sell (or plan to soon sell) AI-rec-tech. All companies used AI technology, as I defined AI in Chapter 1, even if on a small scale. For example, their systems use natural language processing, machine learning, speech or image recognition, neutral networks, deep learning, or decision trees.

All companies provide or facilitate assessment/interview technology, or both. This focus on assessment and interview technology meant I was comparing like with like; for example, I wasn't also studying systems which advertise jobs, source candidates, or screen CVs. This enabled me to compare companies, and therefore gave me the option of configuring the data as either case studies or as an aggregate sample.

Each company was at a different stage of developing their system. Some were early-stage start-ups in the process of designing their technology, but most were mid- to late- stage companies. The stage of development was not a priority for my criteria when it came to purposive sampling, partly because of access, and partly because my primary interest concerned the different ways they were dealing with issues of bias, rather than their stage

of development. While I was expecting to see a difference between the age of companies and the issues they were considering, this was not something I found to be the case.

The assessment/interview companies in my sample supply a mixture of employers, from small, to medium, to large companies. Many of these systems were also used across sectors (e.g. manufacturing, education, finance, technology, retail) and across contract types (full-time, part-time, zero hours), while some were more specific to a particular sector such as technology, where they mainly offered software and coding-based tests. In some cases, there was an extremely broad range of tests offered, with some even specific to jobs such as baker, bartender, carpenter, electrician, dental hygienist, teacher, or administrator. This thesis does not differentiate between these but looks at the systems more broadly. However, I do point out that future work should focus more on the differences between types of tests specifically. Generally, I was told that companies purchasing these systems, whether large or small, had a common primary goal: to find employees who are well suited to the role. It is crucial to note the place of AI-rec-tech in the market; there is large demand from all size of companies across many sectors.

3.2.1.1 Company anonymity

Some companies in this study are anonymised and some are named.

The anonymised companies have pseudonyms formed by using different colours to distinguish between companies, for example ‘Company Green’ or ‘Company Yellow’.

Colours have been used rather than numbers as a way of making it easier for the reader to keep track of each vendor company. While my preference would have been to use pseudonyms resembling company names, it was a challenge to find pretend company names which didn't resemble real company names, and I was aware this could carry legal risks.

Each participant is also anonymised, and an abbreviation for their role is used, such as 'Company Green CEO'. The full list of participant abbreviations is in Table 2 below. The named companies are referred to by their names: Pymetrics, Code Signal, and Skillspace.ai.

There are multiple reasons why I have a combination of named and anonymised companies in the thesis. Firstly, this was part of my negotiations with companies when it came to access. While some vendor companies thought they might want to be named for marketing or transparency purposes, others were certain they'd want to be anonymised, and some were unsure and wanted to decide once they'd seen the thesis. I used an opt-in solution which enabled companies to give their informed consent to being named after having seen a full draft of the thesis. They received a copy six weeks before submission and were given three and a half weeks to respond via a signed form, where company leaders could tick a box to remain anonymous or to be named. This accommodated the needs of multiple companies by offering flexibility, and this approach was invaluable to my ability to access these companies.

Secondly, this was important to me ethically. It allowed companies to decide on whether to share their identity once they had seen the contents of the thesis. To me, this enabled

informed consent and full transparency. In addition, it also meant the companies could see the 'action recommendations' which I propose in Chapter 7, some of which cover design and auditing. One contribution of this thesis is the recommendations, and for this to be read by these companies is significant. The recommendations were received well by some companies, some have said they will consider these suggestions or invited me to present these findings and suggestions via talks or webinars.

Thirdly, it was important due to copyright. This opt-in solution allows me to credit the websites and reports of companies who were happy to be named, as in line with copyright. It also allows me to still directly quote the websites and reports of those companies who wanted to remain anonymous without citing them, as this complies to 'fair use'. I was advised by the Copyright & Licensing specialists at Oxford University that I could still directly quote these sources without a citation, as the requirement in the legislation that you acknowledge the 'author' doesn't apply where this is "impossible for reasons of practicality or otherwise", and maintaining anonymity falls under this 'otherwise' category (Copyright, Designs and Patents Act, Section 30, 1988). To maintain this anonymity for these companies more stringently, I have added two further companies, within which I did not interview, to the websites and report data, and to the demo/free assessment data. This helps to double anonymise these companies. When citing a website, report, demo, or free assessment of an anonymised company, I do not provide the company pseudonym, and this means that if it were traced back to a company, there is no guarantee that I interviewed at that company.

Table 1 indicates the company codes/names used in-text as well as the types of technology they offer, with slightly more detail on the named companies. To maintain consistency

throughout the thesis, I do not offer much more information on those named companies than the ones which are anonymous. Of course, there are points where I can go into slightly more depth about those companies who have opted-in to being named, but I try to maintain a regular weighting in terms of discussion of each vendor company.

While all the companies offer assessment/interview technology, the specific ways in which they do these tests differs. Sometimes, they offer this alongside other elements such as screening or onboarding, although I do not focus on these aspects of their system. The table below outlines these variations without giving too much detail which could enable identification of anonymised companies. Broadly speaking there are tests which assess behaviours and those which assess skills.

AI-rec-tech Vendor Company	**Company Code**	**Type of Company**
Anonymous Vendor Company 1	Company Green	Screening, assessments (behavioural and skills), interviews, onboarding.
Anonymous Vendor Company 2	Company Yellow	Assessments (behavioural and skills), interviews.
Anonymous Vendor Company 3	Company Blue	Assessments (behavioural and cognitive).
Anonymous Vendor Company 4	Company Orange	Early-stage start-up – skills assessment product in development.
Anonymous Vendor Company 5	Company Red	Cultural analytics, assessments (behavioural and psychometric).

Anonymous Vendor Company 6	Company Pink	Assessments (skills). *(Website, reports, demos/free assessments only for double anonymisation, no interviews)*
Anonymous Vendor Company 7	Company Purple	Assessments (skills). *(Website, reports, demos/free assessments only for double anonymisation, no interviews)*
Pymetrics	-	Pymetrics is a US-based company which produces gamified and "audited AI behavioural soft skill assessments" developed through cognitive science. This matches people to jobs using this behavioural data (Pymetrics, 2022a, 2022f).
Code Signal	-	Code Signal is a US-based company which produces skills assessments, largely for coding and engineering jobs. They offer pre-screening and interview technology. This enables employers to "hone in on skilled candidates early with our top-of-funnel evaluations" (Code Signal, 2024).
Skillspace.ai (by AI Planet)	-	Skillspace.ai is a Belgium-based company which produces "automated skill-based assessments". These largely focus on technical roles, and they offer coding/programming

		assessments, "professional" assessments (concerning data science and project-based tasks), and interview technology.

Table 1: Vendor company and interviewee codes and roles

In terms of where the companies at which I interviewed are based, five of them were in the US (two of those being Pymetrics and Code Signal), one in the UK, one in Singapore, and one in Belgium (Skillspace.ai). Of the two companies at which I did not interview, but used data from their websites, reports, demos, and free assessments for double anonymisation purposes, one was based in Singapore and one in Australia.

Therefore, this sample mainly pertains to the Global North, particularly the US. Largely, this is down to the fact that these are mainly the places where these companies are based, and I could not find many companies like this in the Global South. Certainly, it would be remiss to ignore the fact that China and India, both global leaders in AI deployment, are not within this sample (IBM, 2022). This was due to a lack of information on the existence or nature of AI-rec-tech companies in these countries.

3.2.1.2 Online interviews

The primary method of data collection from AI-rec-tech vendors was synchronous, video-based, semi-structured, qualitative digital interviews via Zoom. Before carrying out these interviews, I held three pilot interviews with peers, which helped me to shape my pacing and questions.

Overall, I had 18 interviewees across the eight companies. Regarding purposive sampling within companies, I aimed to speak to people in a variety of roles, from Sales and Marketing, to Design, to Leadership and C-Suite. While I did interview people in different roles, I mainly spoke to people who had decision-making or leadership roles. This allowed me to get the broad overview of systems, company values, leadership visions, and their design and implementation decisions.

Table 2 gives a breakdown of each of the participants, their roles, which company they worked at, and the date of our interview. Some of the titles for the individual's 'role' have been altered for anonymity reasons.

Within named companies, some of the participants are named and some anonymised. This was because they decided individually, prior to their interview and via an informed consent form, whether they would want to be named if the company was de-anonymised.

Each interview lasted between 45 and 90 minutes depending on the availability of the participant. Regardless of interview length, I always managed to get through all the questions.

Company and Interviewee Code / Name	**Role**	**Date of Interview**
Company Green CEO	CEO *(or equivalent)*	31/05/2022
Company Yellow IOP1 Company Yellow IOP2	IO Psychologists	09/12/2021

(2-2-1 interview)		
Company Blue CSO	Chief Science Officer *(or equivalent)*	02/11/2022
Company Blue HOS	Head of Sales *(or equivalent)*	15/02/2023
Company Blue CPO	Chief Product Officer *(or equivalent)*	01/03/2023
Company Orange CPO	Chief Product Officer *(or equivalent)*	03/03/2022
Company Red Partner	Partner *(or equivalent)*	20/01/2022
Company Red HOS	Head of Sales *(or equivalent)*	08/02/2022
Company Red CSM	Customer Success Manager *(or equivalent)*	22/02/2022
Company Red CEO	CEO *(or equivalent)*	25/02/2022
Company Red CTO	Chief Technology Officer *(or equivalent)*	11/04/2022
Dr Frida Polli, CEO of Pymetrics	CEO	15/12/2021
Head of Impact, Pymetrics	Head of Impact *(or equivalent)*	12/01/2022
Alex Vaughan, Former CSO of Pymetrics	(Former) Chief Science Officer	20/01/2022

	(Alex was in this role at the time of our conversation in 2022[10])	
Tigran Sloyan, CEO and co-founder of Code Signal	Co-founder and CEO	09/11/2021
Sophia Baik, COO and co-founder of Code Signal	Co-founder and COO	01/03/2022
Chanukya Patnaik, CEO of Skillspace.ai	CEO	26/05/2022

Table 2: Vendor company and interviewee codes and roles

As can be seen in Table 2, there is one notably different interview. In Company Yellow, I conducted a 2-2-1 interview rather than a 1-2-1 interview, and I spoke to IO Psychologists rather than C-Suite or those in leadership roles. This was their condition for the interview, and every opportunity to speak to people was too good to turn down. This interview was certainly different to others, and I'm sure that this format changed what the participants said to me. In this interview, one of them tended to answer first and then the other would agree, rather than necessarily saying what they thought. However, it was still valuable data.

Zoom with end-to-end encryption was used for interviews to ensure privacy and security. Interviews were recorded on Zoom and downloaded to my computer, as well as being recorded on my personal dictaphone as a backup. These details were all contained in the information sheet, which each participant was asked to read and fully understand before

[10] While some of the other participants have also left these companies since being interviewed, I only specify this for Alex because he is a named participant. I do not specify this for anonymised participants in case it might play a role in identification.

signing a consent form. A big advantage of doing this on Zoom was that it allowed my research to be internationalised without travel costs (both financial and climate-related) (Deakin & Wakefield, 2014; O'Connor et al., 2008).

In terms of the questions asked, broadly speaking, I focused on three topics:

- How the system deals with gender bias.
- In what way(s) these systems compare to other recruitment methods.
- How success is measured in these systems in relation to bias.

The interviews were semi-structured, so there was flexibility in terms of going question to question to get the most out of the interviews and being led by participants. I have put a more detailed interview guide in Appendix 1. The script before these interviews, also contained in Appendix 1, was largely drawn from O'Connor *et al.* (2008). This gives a mock script including a brief introduction to the research, an explanation of the procedure, aims, length and types of questions. While I didn't go into too much depth (I didn't want to waste precious time with these interviewees) I did include a short introduction before commencing the interviews.

As Brenner (1985) suggests, I attempted to ensure that questions weren't directive so that I wasn't suggesting a 'right' answer – I also tried to follow up questions non-judgementally (Brenner, 1985). While I endeavoured not to be leading, I did sometimes play devil's advocate, repeating things back to the participants, or challenging them on certain things to see their reaction. I found that, in general, this tended to build rapport and therefore more openness in the interview.

Interviews were largely about knowledge transfer regarding personal and collective conceptualisations of gender bias. In this sense, the interviewee is characterised as teacher (Foley, 2012) and the interviewer as miner (subjective or objective knowledge is buried, and the interviewer unearths it) rather than traveller (the interviewee takes the interviewer on a journey and this allows the interviewer to tell the story, and to create new knowledge) (Kvale, 1996). My aim was to excavate the subjective epistemologies of these employees, which was another reason why this open but sometime provocative technique worked well.

After interviews, I would ensure to always email my thanks to the participants.

In order to allow for my frame analysis and thematic analysis, as Strauss (1987, p. 267) recommends, I wrote notes and ideas in my interview journal immediately after each interview, demo, or free assessment. I also transcribed each interview shortly after it occurred, doing this manually rather than using a transcription service. This was partly due to privacy reasons and partly because, from experience, I found that this gave me time and space to reflect upon the content. An extract of my interview journal can be found in Appendix 2. Usually, my notes would be evaluative as well as descriptive, and they would brainstorm ideas of themes or patterns in the data, perhaps linking it to other interviews. As Gorman *et al.* (2005) notes, preliminary analysis can be led by 'hunches' and they contribute to the conceptual framework for gathering data.

In terms of the philosophical approach to interviews, as spoken about by Edwards and Holland (2013), I took a mixed approach between positivist and interpretive interview stances. In a positivist approach, the interviewer stresses objective reality and truth. All

interviewees are asked the same questions, and there is minimal influence from the researcher. The interviewer seeks independent and objective reality, and therefore attempts to separate out value-free knowledge gained through systematic procedures from beliefs, feelings, and moral stances. The objective truth I sought in these interviews was about how the systems functioned, what they aimed to do, and how the functionality addressed their own thoughts and understandings. That is where the interpretive stance comes in. This is loosely concerned with understanding social phenomena from the perspectives of those involved. Knowledge takes the form of explanations about how others make sense of their day-to-day lives and interactions. Therefore, it's the interaction of the participant and interviewer that creates this knowledge. This is appropriate because as well as seeking the objective knowledge about how the systems work, I was looking for what was underneath in order to understand their thoughts surrounding particular concepts (Edwards & Holland, 2013).

Arguably, this relies on the epistemology of constructivism, outlined by James and Busher (2009). This explains that truth and meaning are constructed, not discovered. This research notes that all knowledge emerges from constructivism – people construct ideas about concepts, both independently and collectively, and these feed into the technological operationalisations of these concepts (James & Busher, 2009).

Interviews were therefore a suitable method for this research; they allowed me to address my research question(s) by exploring how employees conceptualise and operationalise terms such as bias and gender. This was crucial; I wanted to understand not only the processes, but the logic behind these processes, and "qualitative research is ideally placed to understand the process of events – how ideas become actions, the reactions to those

actions, and so on…" (Gorman et al., 2005, p. 6). Interviews allowed me to gather in-depth information, opinions, and explore people's thinking and motivations from "studying people's understanding of meanings in their lived world" (Kvale, 1996, p. 105). In this case, their 'lived world' being their organisation, and the processes within. As Grigoropoulou and Small (2022) discuss, in-depth interviews help us to understand decisions that affect the nature of algorithms (or systems). They write that, "[t]o understand those human roles, interviews and field observations are among our most powerful tools: interviews are designed to unravel how people make decisions" (Grigoropoulou & Small, 2022, p. 905).

Therefore, these interviews allowed me a window into the decisions that happen behind closed doors. Gerson and Damaske (2020) talk about exactly this in their book *The Science and Art of Interviewing.* They note, quite rightly, that interviews allow a glance into private settings and some of society's most entrenched institutional arrangements, showing us how structural and cultural forces interact with individual meaning making and decisions to produce a range of outcomes.

3.2.1.3 Walkthroughs and autoethnography: demos and free assessments

In some companies, I was walked through the system via a Zoom call or detailed video (what AI-rec-tech companies refer to as a 'demo') and in some cases, I also completed a free assessment. The demo is used by the vendor companies to show their prospective clients the back end of the system and to explain how it works. The free assessment is also generally used so that prospective clients can experience these tests or assessments as a user.

While this wasn't possible with all the companies, I was able to do these exercises within four companies where I interviewed employees, and two companies where I didn't interview employees (Company Pink and Company Purple). Doing free assessments at these two companies allows an extra layer of anonymisation. In cases where I am restricted by NDAs or company anonymity (as can be seen below, all companies where I completed demos and free assessments requested to be anonymised), I provide re-phrased words or replicas of demo visuals or free assessments, rather than exact screenshots or wording from the tests. This enables me to provide the same or similar meaning without giving away the identity of the company or any sensitive information. Table 3 shows the companies where I took demos and/or free assessments.

Company	Demo	Free Assessment
Company Green	Done 29/09/2022	-
Company Yellow		Done 22/09/2022
Company Blue	Done 14/02/2023	Done 02/11/2022
Company Red	Done 19/04/2022 (demo video)	Done 21/09/2022
Company Pink	-	Done 28/02/2023
Company Purple	-	Done 28/02/2023

Table 3: Vendor company demos and free assessments

In terms of the methods used to collect this data, I would class these methods of data collection as: [11]

[11] Demos and free assessments were done after interviews. I chose to do them in this order to ensure they didn't change my ideas about the systems before speaking to participants.

- Walkthroughs (demos)
- Autoethnography (free assessments)

Walkthroughs are a method used mostly when studying apps. They are a data gathering procedure which involves the researcher engaging with the app interface, working through screens, tapping buttons, and exploring menus (Light et al., 2018). Of course, in this case I am not studying an app, but rather an assessment system. However, the idea remains the same: walkthroughs, or demos in this case, enabled me to explore the interface and become familiar with how it feels to use, with the added ability to ask questions to the person showing me to explore the rationale behind it, and therefore to get to the heart of my research question(s). This enabled me to understand more about how the system worked – for example, what information can employers see about each candidate? It also aided my analysis of the decisions behind the design. As Ritter (2022) points out, walkthroughs reveal the sociocultural assumptions embedded in interfaces and enables the researcher to engage with the construction inherent in communities of practices.

Free assessments allowed me to incorporate elements of autoethnography into my methods. Ethnography is traditionally about immersing oneself within the culture to produce a holistic analysis: "[e]thnography is prized as a method for getting to the heart of meaning and enabling us to understand, in the round and in depth, how people make sense of their lives" (Hine, 2015, p. 1).

In a similar way, autoethnography uses a researcher's personal experience to describe and critique the cultural beliefs, practices, and experiences of the 'insiders' using deep and

careful self-reflection to name and interrogate the intersections between the self and the research subject(s) (T. Adams et al., 2015). Relatedly, doing these assessments myself, I was able to reflect on how they were constructed and presented, and to think about how this relates to gender bias as spoken about in interviews. In addition, it allowed me to experience how it might feel as a candidate to complete these tests and encounter elements of design. For example, how does it feel to be asked my gender at the beginning or end of the assessment? Or to see cartoons or avatars on screen which are gendered? How does the assessment relate to, or contradict, my interview material? How does it feel to be given a profile which tells me about my skills, behaviours, or persona in the workplace? Free assessments allowed me to observe how activities happen, rather than solely relying on selective retrospective accounts from interview participants (Hine, 2015).

As Crawley (2012) points out, autoethnography is a kind of self-interview. It allows the researcher not to be boxed into categorical notions of method. If we see the interview as a site of joint action between respondent and interviewer, then "why pretend to ignore or obfuscate that relationship of the researcher to the site, content, and lived experience of the issues being investigated?" (Crawley, 2012, p. 19).

The type of data which demos and free assessments provided me with was also interesting. While I do use it explicitly in some instances, in most cases it became tacit knowledge. By tacit knowledge, I mean knowledge that is not formally recorded (e.g. written within the thesis) but which is contained in my own thoughts and shaping of the arguments (Taylor, 2007). Tacit knowledge is defined by Kabir (2013) as subjective, informal, and internalized. It is not tacit in the sense that I cannot articulate it, but I *should not* put it in writing, largely because incorporating elements of tests would often give away the identity of the companies.

3.2.1.4 Content research: conferences, websites, and reports

I also collected and analysed data from publicly available sources from AI-rec-tech vendor companies and beyond. This included:

- Company reports and whitepapers.
- Company websites.
- Public webinars or conferences.

The reports or whitepapers which I analysed were published on company websites. I analysed those which seemed to be relevant to my research questions. I looked at these during September 2022, in total analysing 32 white papers, reports, journal articles (produced by the companies), user guides, manuals, and sample assessment reports.

I also analysed the content of the websites of all eight companies where I interviewed, plus two other companies in which I didn't interview, Company Pink and Company Purple. The addition of these two companies again allows an extra layer of anonymisation.

While Company Orange had no website yet and some of them had more extensive websites than others, I took and analysed 413 screenshots from websites during September until November 2022, and in February 2023 for the additional two companies. I looked at the language surrounding bias in their sales material, their blogs, their case studies, and so on, analysing how they presented it to their customers, and reflecting upon

how this matched up with interview material. Generally, their marketing material was often identical to what they told me during interviews.

In addition, I attended or watched back six online industry conferences on the topic of recruitment technologies. These conferences were aimed at marketing different technologies to potential clients or providing an interesting discussion to the HR community or the general public. This data helped me to see how these technologies are being presented in the public domain (for marketing purposes) as well as the private one. These conferences were not necessarily attached to the companies I was studying. Given that these are publicly available, I cite these throughout the text where appropriate.

Conferences on AI-rec-tech seem to be becoming much more commonplace (Heather, 2018; Holbein, 2020). Most of these conferences are now online (AIHR, 2021) which offered me the opportunity to attend. I selected conferences based on relevance and cost. The conferences I attended are listed in Table 4:

Conference	**Date**
Using Psychometric to Make Better Hires, More Quickly	02/02/2022
IHR Recruitment Technology Virtual Conference	29/04/2021
Recruitment Automation Community Conference on Bias, Diversity, and Inclusion in a Digital Work	04/06/2020
The Talent Board Conference: How Technology Empowers Recruiting and Candidate Experience	27-28/05/2020
The Impact of AI and HR on Future Work - CogX	27/11/2018
The Future of Recruitment Technologies - HR.com	07/12/2022

Table 4: Recruitment technology conferences attended

The purpose of collection from these data sources was threefold: firstly, it allowed me to see how AI-rec-tech companies are selling this technology. Secondly, it enabled me to contextualise my findings within a wider discourse formed by a network of HR professionals and those in HR automation and technology. Thirdly, it showed me how these companies were outwardly presenting themselves in comparison to how they appeared in the interviews. In many cases, I found that the things which were being said to me in interviews appeared to be almost 'party lines' which were often included in website or report material. Again, all these methods enabled me to get to the heart of my research question and investigate how AI-rec-tech vendors are framing gender bias in hiring, both in terms of their own understanding and how this is translated to design.

3.2.2 Law and regulation experts

In addition to focusing on AI-rec-tech vendor companies, I also collected data surrounding law and regulation related to these technologies. After interviews with some of the AI-rec-tech vendors, I found that regulation kept emerging as a theme related to gender bias. Considering this early finding, I decided to carry out interviews with regulators and to analyse relevant law and regulation documents.

Small and Calarco (2022) talk about the five indicators to distinguish empirically well-executed research from poorly executed data collection. One of these is 'follow-up': the extent to which the researcher collected data to answer questions that arose during the data-collection process itself. Gathering additional data to answer unanticipated questions

is a form of responding to the field and one means through which good fieldwork gets closer to the experience of those interviewed.

3.2.2.1 Online interviews

From October 2022 until May 2023, I approached regulators and legal experts in the UK, US, and the EU for interviews. In general, it was easy for me to select the sample because there are very few regulators or government bodies working on this technology.

I found that regulators in the UK were particularly evasive. I often got the response that there was no-one working on this technology specifically:

> Dear Clementine,
>
> Thank you for your email. Unfortunately, [our company] does not have the correct expertise to be able to help with your research.
>
> I'm really sorry that we can't be of more assistance.
>
> Many thanks,

Figure 1: Reproduced rejection email - example 1

Or:

> Hi, we have no one available to speak with you.
>
> King regards

Figure 2: Reproduced rejection email - example 2

I even arranged an interview with someone from the CDEI only for them not to show up and then stop responding to my emails altogether.

In contrast, I did find that in the US, the regulatory bodies were willing to speak with me much more readily. Consequentially, Chapter 6, which details how gender bias was at times perceived in relation to legislation, is mostly relevant to the US.

Table 5 below shows the interviewees, the dates I spoke to them, and their organisations and pseudonym/name. All experts indicated on their consent form whether they were willing to be named in the research or not. Overall, I interviewed six experts from the EU Commission, the EEOC (US), and the OFCCP (US). In Chapter 6, I give more detail on each of these government bodies and on the individual interviewees.

Pseudonym/Name	**Institution/Government Body**	**Date of Interview**
AI Regulation Expert at European Commission	European Commission [EU]	14/02/2023
Keith E. Sonderling *(Commissioner)*	Equal Employment Opportunity Commission (EEOC) [US]	17/11/2022
Senior Staff at OFCCP	Office of Federal Contract Compliance Programs (OFCCP), US Department of Labor [US]	03/11/2022
Bradford J. Kelley *(Former Chief Council to Commissioner Sonderling)*	(Formerly) EEOC [US]	06/02/2023

Brandi Meredith *(Senior Attorney Advisor to Chair Charlotte A. Burrows)* Eunice Ikene *(Deputy Chief of Staff to the Chair)* *(2-2-1 interview)*	EEOC [US]	02/05/2023

Table 5: Law and regulation experts: pseudonym/name and institution

It was powerful speaking to these experts and regulators, and it has been a big part of shaping my thesis and shedding light on the data I collected from vendor companies. It must again be noted that I interviewed Brandi and Eunice simultaneously, which, in a similar way to the 2-2-1 joint interview noted above, will have changed the dynamic and to some extent the answers given.

Interviews with these experts were entirely different to one another and to the interviews with vendors; I asked very structured and specific questions about the relevant law and regulation. This was mainly because I wanted to question the regulators on specific elements of text, or concepts raised by the vendors. I've included a sample of my interview script with the AI Regulation Expert at the European Commission in Appendix 3.

I used Zoom with end-to-end encryption for these interviews and recorded them on Zoom and on my personal dictaphone as a backup. I provided the experts with an information

sheet and individual consent forms before the interviews which outlined how the data would be used and which asked them if they wanted to be anonymised. All of them gave their informed consent.

All interviews lasted between 45 and 60 minutes. They were incredibly generous with their time and passionate about the subject.

3.2.2.2 Content research: regulatory and legal documents

In addition to interviews with regulators, I analysed key regulatory and legal documents. These were selected as they were referenced in interviews both from the AI-rec-tech vendors and the regulators themselves. They included the Uniform Guidelines based on Title VII of the Civil Rights Act 1964 (US), the UK Equality Act 2010, the GDPR (EU), the EU AI Act, along with local state law in the US including New York, Illinois, and Maryland. I use this data mostly in Chapter 6, which discusses gender bias in relation to the current regulatory landscape. Using these documents, I looked at the relevant sections and focused specifically on how it would relate to gender bias in these technologies.

While I am not a lawyer by training, I felt this was a crucial aspect of my research. Discussing law and regulation of this technology is vital to advancing the field and providing a full picture of the factors relating to gender bias in AI-rec-tech.

3.3 Data analysis

3.3.1 Frame analysis and thematic analysis

In the previous chapter, I discussed Goffman's frame analysis. It was important to me that this theoretical framework also informed my methods. However, while frame analysis has consistently been used to discern how perception and interpretation are organised, there remains no systematically consistent way to use it as a method of analysis (David et al., 2011; Johnston, 1995; Mendonça & Simões, 2022).

Therefore, within this thesis, I chose to use a combination of frame analysis and thematic analysis.

Frame analysis, a flexible and context-specific method, allows me to closely study the micro-aspects of the text to discern narratives and interpretations, and thematic analysis, a more structured method, enables me to have a broader macro-process through which to string together interpretations. This combination also allows me to keep my analysis flexible and therefore to apply frames within the newly trodden context of technological design.

The aspects of frame analysis utilised within this research included close and in-depth analysis of text, speech, and other media for sequences and patterns (Linström & Marais, 2012). Scholars such as Johnston (1995) have spoken about how frame analysis allows us to see how interpretation and perception are formed through a 'micro-frame analysis'

which encourages close reading of the text (including verbal texts such as interviews or speeches) to interpret meaning. Looking in close detail at the text is something which I utilised in my analysis, paying special attention to words and discourse within interviews, websites, reports, which were recurring or clashing. In this sense, I also used part of analysis methodology put forward by Van Gorp (2007) which suggests reconstitution of patterns of interpretation (frames) while being attentive to three things: the framing devices (which words, metaphors, cases, visuals are used), the reasoning devices (which justifications, causes, consequences, and equivalences are presented), and the implicit cultural phenomenon (what does this reveal about broader patterns of interpretation and social values). In addition, I concentrated on how frames were transferred from understanding to practice – how frames are mobilised and how they interact with other frames (Mendonça & Simões, 2022).

Thematic analysis provided me with a more structured approach and enabled me to convert this closer analysis to interpret the macro-level patterns, themes, and ideas which emerged from the data (Lapadat, 2010). In many ways, it encapsulates the same end goals as frame analysis, and so these methods are well-paired. This method is also a constructionist method – much like frame analysis, it examines the ways in which realities, meaning, experiences (in this case, the design process, and the underlying understandings of certain concepts) effect a range of discourses operating within society (and within organisations). To convert this closer analysis into broader themes, I followed the formulaic structure for process of analysis, guided by the suggested 'phases' of analysis from Braun and Clarke (2006) show in Figure 3. Generally, I employed 'frame analysis' within the first phase.

Phases of thematic analysis	
Phase	Description of the process

1. Familiarizing yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Figure 3: Reproduced from Braun and Clarke, 2006, p. 87

While Braun & Clarke write that there is no hard and fast rule about what proportion of data is needed to display evidence of a theme, I tried to listen to the data and to be guided by the most prominent patterns which emerged surrounding my research questions. Once organised into themes and patterns, I moved on to interpretation, where there is an attempt to theorise the significance of the patterns, their broader meanings, and implications (Braun & Clarke, 2006).

In addition, thematic analysis prompted me to go beyond the semantic content and identify underlying ideas, assumptions, conceptualisations, and ideologies informing this content. In this sense, "thematic analysis can be a method that works both to reflect reality and to unpick or unravel the surface of 'reality'" (Braun & Clarke, 2006, p. 81). Guest, MacQueen and Namey (2012) also notice that thematic analysis requires more

involvement and interpretation from the researcher, moving beyond counting words and phrases to identifying and describing both implicit and explicit ideas within the data.

Braun and Clarke (2006) also advocate that themes do not 'emerge'; this creates an illusion of the passive process of analysis. Instead, we must admit the active role of the researcher, who identifies these patterns and themes, selecting and reporting them. They argue that a theme captures something important about the data in relation to the research question and represents patterns within the data set. This recognition is important: my positionality, my own reflections, and my own selections and interpretations of data shape this research.

3.3.2 NVivo: seeing and coding

As can be seen above, part of thematic analysis requires coding. To code my data, I used NVivo.

This helped me with the richness and complexity of the data as I could separate it out into multiple codes and reflect upon how these related. As Guest, MacQueen and Namey (2012) point out, this is an iterative process where the text is re-read and analysed via segmenting and coding into categories, types, and relationships. We deal with the messiness of the text by minimising the messiness in the code, they argue.

NVivo was particularly helpful and suitable for my research because of the richness of the data I collected. Richards (1999) argues that NVivo is suited to rich data and is designed for researchers who wish to display and develop rich data in dynamic

documents: “In NVivo, rich text can stay rich, and rich documents develop as ideas do. As analysis proceeds, researchers can expand data from summaries to detailed descriptions. As interpretation changes, they can reject and add, revise and rethink data, interviewing data and commentary” (Richards, 1999, p. 414).

As Bazeley and Jackson (2013) talk about, coding is part of knowledge production – it is an abstract representation of a phenomenon or object and can range from being purely descriptive to being labels for themes, and they can be broad or detailed. I used both, eventually sorting my codes into broader themes with sub-themes, using parent codes to distinguish between these. Eventually, this helped me to structure the chapters of my thesis.

While this is a useful tool to aid analysis, it is still up to the researcher to interpret, conceptualise, examine relationships and develop theory, not the computer (Bringer et al., 2004). I recognise with this method of analysis, someone else might see different codes to the ones I perceived. Therefore, the analysis becomes inseparable from me as a person, and acknowledging this is vital to my methods.

3.4 ‘The hustle’: the struggle to access companies

The term ‘hustle’ has been used both academically and non-academically. In academic work, Carter (2016) has referred to the hustle, defining it as how influencers on social media work to make their brand visible. Ticona (2022) has also used the term ‘digital hustle’ to refer to strategies for how technology is used in the setting of getting, maintaining, and carrying out precarious work, allowing people to piece together paid

gigs across many different labour markets but, ultimately, re-producing inequality between high and low wage workers. In non-academic settings, the hustle is a term often used within popular culture. In the 2019 film 'The Hustle', Anne Hathaway and Rebel Wilson star as two scam artists who aim to con an internet millionaire, and in 'The Real Hustle' TV show, con artists perform distraction scams and proposition bets on members of the public.

Here, I re-coin the term 'the hustle' to refer to something other than scamming, social media influencers, or precarious work. Instead, I want to define 'the hustle' to refer to the difficulty accessing desired research subjects and the implications this carries for researchers. The hustle captures the movement in methods: the pushing or jostling of the researcher in the face of resistance when it comes to accessing elite subjects.

Therefore, I define 'the hustle' as: the struggle to access research settings or participants in the face of resistance, which requires the researcher to transform their role to networker, and additionally alters research design, data collection, and interview dynamics (Collett, 2023).

Pokémon helps to explain this definition. In Pokémon, some have the ability 'hustle'. This increases the user's attack stat by 50% but lowers the accuracy of the user's physical moves by 20%. The Pokémon ability of 'hustle' summarises that while my effort to access participants (or 'attack') was increased in terms of how much I was contacting companies and requesting access compared to other types of research with more easily accessible research subjects, my success (or 'accuracy') remained low. In other words, while I was trying hard to access companies, I had a low success rate.

Accessing AI-rec-tech companies in this research required more effort than would be the case with many other research subjects. As Li (2021) notes, this is often the case with elite interviewees; it is hard to reach them and even harder to get them to agree to an interview. Although the definition of elites has been long debated and uncertain, in this article I loosely follow Harvey's definition of elites as "those who occupy senior management and Board level positions within organizations" (Harvey, 2011, p. 433). As McClure and McNaughtan (2021) talk about, elites are hard to pin down because often they don't share their contact information publicly, and often utilise gatekeepers, such as personal assistants or administrative staff.

In my case, the hustle can be explained by a few factors: in these companies, there is little time for employees to do things outside of their job scope; trademark and IP are sensitive because they worry about competition with other companies; and there is a particular nervousness because of negative media attention and/or recent legal action (Harwell, 2019a). This also might have been harder for me to access these companies because I was a doctoral student as opposed to a professor, who might have had more credibility or value in the eyes of these companies.

The struggle to access elites could also be explained by not appearing as an 'insider'. Mears (2020) touches on this in her book, *Very Important People*, a sociological and ethnographic exploration of the party circuit for the world's 'very important people' which attempts to understand what they do with their money. Mears notes that, as an ex-model herself, there was an element of her seeming to fit into these scenes which allowed her to research them (Mears, 2020). In the opposite sense, Cerón-Anaya (2019), in his

ethnographic research study conducted in exclusive golf clubs in Mexico, draws attention to how these clubs erect firm social boundaries with the outside world. He gives the example of one interviewee, Fernando, a young golf player who no longer seemed willing to provide contacts of other golfers to Cerón-Anaya for his research after it became clear that he did not own a car. Cerón-Anaya hypothesises that, due to the class symbol of having one's own car in Mexico, this established him as an outsider, subsequently making access more challenging (Cerón-Anaya, 2019).

The difficulty accessing companies was evident from the avoidance tactics and the barriers which these companies erected. As Mikecz (2012) talks about, gaining access to elites must be carefully negotiated and this can take longer and have higher costs because they "purposefully erect barriers, which set them apart from the rest of society" (Mikecz, 2012, p. 483) and which allow them to manipulate information and deny access to it.

While many companies completely ignored my request, others would reject it, perhaps claiming they didn't even use AI in their systems, even though often their marketing said that they did. For example, one company wrote to me:

Hi Clemi,

Ok, it really depends on the amount of work involved here as to whether I'd be available to contribute.

Worth noting that our platform doesn't use AI in its decision-making so we may not be the right vendor to speak with/about.

Thanks

Figure 4: Reproduced rejection email - example 3

In terms of other barriers, two companies which initially agreed to be interviewed asked me to sign NDAs. One of them was approved and signed by the Oxford University Contract Team, the other NDA was so strict that it wouldn't have allowed me to write and publish any research regarding the company, even if anonymised. Of course, this could not be signed and so the company was not included in the research.

The hustle signifies how it was not only difficult to get hold of companies, but in the agreement and negotiation stages, companies often slipped through the net. This happened during my research despite my best attempts to be flexible and accommodating within the boundaries of my research questions and aims, for example, allowing companies to opt-in to being de-anonymised after seeing the final document.

To recruit these companies, I used a combination of emails to existing contacts, cold emails, open call through my knowledge partner (the Women's Forum for Economy and Society), snowballing, and LinkedIn posts from professional HR contacts. I had already reached out to a few companies prior to the doctorate and got some positive responses, so I was confident it would not be difficult to recruit companies. This was naïve.

After the first term of my doctorate, in January 2021, I started trying to recruit participants. Two years later, by January 2023, I had reached out to 58 companies, which, from my research, were the only relevant ones I could find (e.g. they designed and sold assessment and/or interview systems). I got eight companies to agree to interviews, and one other company agreed in theory but asked me to sign an overly restrictive NDA. I never interviewed people at this company.

While 12 other companies also responded to me, they eventually rejected the request or pulled out last minute. Sometimes, this was after months of correspondence and negotiation. In many cases, as Empson points out, people can string you along and then can just stop answering emails (Empson, 2017). This process was long, repetitive, frustrating, and did not feel like research. Ironically, I felt like someone applying for, but failing to obtain, a job.

Initially, I had also aimed to conduct research on client companies purchasing and using AI-rec-tech systems to understand the implementation process, their motivations behind their usage, and why they purchased certain technologies. I found 222 client companies online through the vendor companies' websites client pages, or through LinkedIn posts of contacts. I managed to email or LinkedIn message HR contacts at 92 of these companies and got responses from only ten people. Out of those ten, I managed to interview only four. Some of these participants claimed their company didn't even use the AI-rec-tech, but they were interested to discuss why. I have not used this client interview data within my thesis; it was insufficient and was not rich enough to answer my intended research questions. However, the process of trying to collect data regarding AI-rec-tech clients and failing to do so showed me that companies using this technology do not want to make it known.

It is worth noting that, in contrast, my success rate with regulators was initially much higher, particularly in the US and at the EU. In the UK, I reached out to five regulators, policymakers, or government bodies which should be working on regulating this technology. Only two responded saying they would be willing to speak – one never showed up to the arranged interview and stopped responding to my emails, and the other

also stopped responding to my emails. I reached out to three governing bodies in the US and managed to interview people at two institutions, and the one email I sent to the EU Commission was followed by a prompt and helpful response which resulted in an interview.

It is difficult to quantify how long I spent attempting to recruit participants for this research. Had I known just how much of a hustle it would be, I would have kept a timesheet. I urge anyone anticipating the hustle with their research to do so. However, if I try to quantify it, I find the following: looking back at my timetable, I was spending, at the *very* least, half a day per week on recruitment between January 2021 and January 2023. This included emailing people, following up, arranging interviews, looking up companies or people on LinkedIn, or having introductory calls with potential interviewees. Given there are 52 weeks in a year, there are 104 weeks in two years. 0.5 multiplied by 104 is roughly 52 days, which comes to about two months of solid work trying to recruit participants. Given this is likely an understatement, this leaves wiggle room for weeks when I was on holiday or not working.

Unfortunately, all of this is what I refer to as 'hidden research time', which, again, is part of the hustle. This hidden research time isn't reflected in my data or my results, but it is a significant part of how my research has been shaped.

All of this demonstrates the hustle in the recruitment stages of my research; I put in maximum effort to gain minimal success and morphed from a researcher into a desperate networker and opportunist. As Ostrander (1993) said, perhaps for research regarding

elites, “luck and a willingness to take advantage of opportunities as they arise have proven [just as] valuable” (Ostrander, 1993, p. 9).

The hustle is an aspect of my research methods which must be recognised. It sheds light on the foundational point of this thesis: AI-rec-tech is a taboo subject and, mostly, those creating or using the technology prefer for it to remain opaque. The reason I continued down this path of research, despite the resistance from research subjects, was because the data I did collect from the vendor interviews was rich and unique, and I am confident this research will make a valuable contribution to the field of AI-rec-tech research. In addition, as I argue in Collett (2023), the hustle signifies that a researcher is likely hitting upon an interesting topic.

Overall, the hustle impacted four fundamental aspects of this research.

Firstly, it forced my own persona, skills, and qualities as a researcher to change into those of a networker. It transformed me into someone who had to be good at convincing people to give me their time, and at arguing that this research is valuable and important. I found it uncomfortable that networking was becoming part of my research success. I wrote in my research journal: “As someone for who networking makes my skin crawl, this is an actively unpleasant experience. I’ve often felt during this process as if I’m constantly asking people to do me favours and this is uncomfortable”. Networking does not come naturally to me. Perhaps this discomfort and lack of access is why some social scientists don’t want to study elites in the first place (Ostrander, 1993). I found it awkward sending three follow up emails to one person if they didn’t respond, but I knew it was necessary

for the success of the research and so I did it. This research, and the hustle, therefore, changed my necessary skills as researcher to those of a networker.

Secondly, the hustle also reduced my ability to gather the amount of data that would normally be thought of as 'enough'. I discuss this in more detail later when exploring the concept of 'rich' data. I propose richness as a benchmark for researchers who go through the hustle, rather than sample size or saturation.

Thirdly, the hustle changed the design of my research. The hustle meant that my ability to carry out the research I had intended was limited. Initially, I had intended to interview client companies which purchase this technology as well as vendor companies. However, as explained above, the hustle made this impossible. This, perhaps, was a good thing; it narrowed my focus, allowing me to look at these technologies from the design perspective. The hustle put me at the mercy of my data access. It's also worth noting that Covid-19 also shifted my research design from what I had originally hoped. I was designing my research during the pandemic, and so although I would have liked to conduct ethnographic research in some of these companies, this was not something which I felt at the time to be an appropriate request.

Fourthly, the hustle altered the dynamics of my interviews. As I touch on more later, it meant that there was a large power disparity, but also, generally my rapport was higher with participants because I felt more grateful for their time.

Recognising the impacts of the hustle is important for future research. Researching AI-rec-tech is difficult and requires maximum effort for access of the most basic kind, but

this methodological concept helps us to consider ways to manage this and how it might manifest, as I demonstrate more below where I pick up on some of these considerations. Ultimately, the hustle highlights the importance of these companies becoming more transparent with researchers in the future and shows crucial impacts of access struggles on research and researcher.

3.5 Data collection and analysis considerations

In this section, I discuss some major considerations concerning my data collection and analysis.

3.5.1 Reflexivity and positionality in collection and analysis

Hammersley (2008) talks about how qualitative inquirers too often attempt to understand people with whom they already sympathise and have been less ready to represent the perspectives of those they regard as playing a more central or dominant role in society, or with whom they have less sympathy.

This is certainly not the case in my own research. I attempted to understand those with whom I didn't sympathise and who play a very dominant role in society. This research is not about trying to give these people a 'say', but rather it attempts to understand their thinking and reflect upon how this might have implications for technology and society.

My own stance and thoughts on AI-rec-tech, as well as my interview persona, are both integral to my data collection and analysis. Our own position in society and our own

views are always going to impact not just what we study but how we study it. Mason-Bish (2019) discusses how we must think of our positionality as researchers and the effect this has not only on our relationship with respondents but on the research in general, including our exchanges and findings. Certainly, as Jacobson and Mustafa point out, the way we see the world around us impacts our research interest, how we approach research and participants, the questions we ask, and how we interpret the data (Jacobson & Mustafa, 2019). Therefore, being reflexive about my positionality and my interview persona are vital to how I collect and analyse the data.

Reflexivity is a process of continual critical self-evaluation on the researcher's positionality, including personal characteristics, gender, race, affiliation, age, and sexual orientation. It can have large knock-on effects on how we think about and frame content, and how we produce and edit material (R. Berger, 2015). Gerson and Damaske (2020) talk about how researchers must attend to their positionality because our conscious and unconscious biases will affect our findings. We must take steps to minimize the influence these biases and their impact on our perceptions about the experiences of others.

Deutsch (2004) realised, for example, that she had to develop her own identity as a researcher, and as a woman who is white and middle class. I had to do the same by looking at myself and my identity in relation to the research. As a young woman, a millennial, a dyslexic who has always struggled with exam conditions, a tech-sceptic, and someone who has had many job rejections, I did at first come from a place of critically questioning these systems rather than welcoming them. I chose to study them because I feel that people should be given equal opportunities in life and the chance to do what they're passionate about.

Therefore, in terms of my positionality, I think it's important for me to notice my original scepticism regarding AI-rec-tech. I have many peers who have had negative experiences with recruitment technology, causing them to feel demotivated and undervalued. They feel confused about what else they can do to obtain jobs, and about what is needed in recorded video interviews. As a result, they try to perform what is needed rather than being their authentic self.

I also think it's crucial for me to reflect upon my positionality in terms of being an Oxford student. I realise that I am in an extremely privileged position, and I cannot fully understand how much these systems could negatively impact the lives of individuals.

It is crucial for me to reflect upon how my positionality may have influenced multiple aspects of this research, and how it also changed throughout. As Small and Calarco (2022) discuss in their chapter on self-awareness in their book *Qualitative Literacy,* field-workers don't just collect data, they co-create it through the transcripts and field notes and the decisions they make. Things like their gender, extraversion, awkwardness, clothing, vocabulary, and so on, can affect not only who agrees to be studied (access), but also what they decide to say (disclosure) and how we analyse the data because each researcher will interpret statements and observations differently (interpretation).

For example, Talmage (2012) talks about 'filtering', which is where we pick and choose what we take from interviews. This can happen because of the self we bring to the interview, the ideology that we hold, what we perceive as shared experience, and filtering for relevant topics. Ironically, I have needed to be wary that the "ever-present danger of

bias may be overwhelming" (Gorman et al., 2005, p. 126) and try to be as neutral as possible.

Having said all this about my original scepticism, I must also acknowledge that during the interview process, I found my opinions changing. I was incredibly taken by the participants, probably due to the hustle and the fact I was grateful for their participation, and I was, at times, quite persuaded by their arguments. The rapport with participants and their own charisma altered how I perceived the data. It was therefore hard for me to see it from a completely objective perspective during analysis. The initial scepticism, mixed with this gradual but conscious swaying, created an interesting tension for me on the topic of AI-rec-tech. I hope that, throughout the thesis, I have balanced their thoughts with the issues these systems raise, and that I maintain a constant consideration of how my positionality impacts my findings.

3.5.2 Me, myself, and my interview persona

It's important to consider my interview persona and how this changed throughout data collection. I found that my persona was altered somewhat during the research.

As noted, when discussing the hustle, initially, I felt so lucky to be able to interview participants because of the difficulty obtaining their time. Therefore, initially, I felt I had to tread carefully, and I tried to come across as warm and non-judgemental in interviews.

However, over time, I became more relaxed and more likely to challenge participants if it would lead us down an interesting path and help to answer my research questions. I

realised that I hadn't been thinking critically in interviews and therefore was not getting the best material out of them. If I was going to do so, I needed to be able to reactively change my interview persona to get the data I needed in order to answer my research questions. Harvey (2011) talks about how, "[e]ffective interviewers are those that are able to easily adjust their style and make the interviewe[e] feel as comfortable as possible" and be able to adjust their behaviour, speaking voice, and mannerisms accordingly (Harvey, 2011, p. 434). This meant adopting a more challenging, provocative stance in some cases, and it seemed to work. For example, in my interview with the Company Red CTO, when they were giving two separate answers to one question, I gently nudged them: "I'm sorry to push you on this but I am going to push you on it, because I am interested to know, if you had to pick one?". I ended up getting a much more fruitful answer.

Gerson and Damaske (2020) talk about how probing for information in interviews is important because it delves beyond the surface and can unearth the process of constructing social accounts as well as many other kinds of information that reveals the multiple meanings which participants attach to their behaviour. Staying alert to and following up on incomplete, unexpected, or vague answers can give an interview depth. Part of this is about listening. Talmage refers to active listening from the interviewer – clarifying, suggesting alternative interpretations, and facilitating significant linkages in order to make sense and understand things (Talmage, 2012).

Overall, I felt this malleable persona enabled me to gather better material. In addition, I tried to balance my challenges with a warm, open, and non-judgemental environment, but it was a big contrast to my initial encouragement and constant agreement with interviewees.

3.5.3 Data sufficiency: is rich data enough data?

There has long been debate about when data becomes 'enough'. Often, data sufficiency is discussed in relation to theoretical or thematic saturation or sample size.

Theoretical or thematic saturation are used to determine whether we have enough data, usually within grounded theory projects and thematic analysis projects respectively. It is often argued to be a determinant to cease data collection and define sample size once no new information or themes emerge from data (Braun & Clarke, 2021; Guest et al., 2006; Nascimento et al., 2018).

In terms of theoretical saturation, some have even tried to define this in quantitative terms, using a lognormal distribution principles instead of relying "wholly on arbitrary judgement" about when to stop collecting (Rowlands et al., 2015, p. 42). Others argue for specific points where theoretical saturation can be reached. Guest, Bunce and Johnson (2006) analysed one of their own research projects that involved the participation of 60 women. They found that the point of no new codes being reached was after the first six participants, and concluded that theoretical saturation mostly occurs after 12 interviews. Hennink and Kaiser (2022) carried out a systematic review of four databases to identify studies empirically assessing sample sizes for saturation in qualitative research and identified 23 articles that used empirical data or statistical modelling to assess saturation. They concluded that saturation is generally reached with a narrow range of interviews (n = 9-17) or focus group discussions (n = 4-8), particularly those with relatively homogenous study populations and narrowly defined objectives. If not prescribing an

exact number, others argue for various indicators of having 'enough' data. For example, Small and Calarco (2022) talk about how exposure is a precondition to all good field projects – the greater the contact, the better the data. In interview data, exposure derives from the number of hours spent talking to respondents, and in-depth interviewers generally agree that more hours of interviewing lead to better data.

While saturation can be a useful concept in some contexts, I wish to challenge the concept of having 'enough' data in terms of saturation, hours spent, or exact sample size. Instead, I hope to re-frame the goalposts to surround data richness.

As Braun and Clarke (2021) argue, attempting to predict the point of data saturation cannot be tied to the number of interviews in which the theme is evident, "as the meaning and indeed meaningfulness of any theme derives from the dataset, and the interpretive process" (Braun & Clarke, 2021, p. 210). I concur. For example, in the case of Guest, Bunce and Johnson (2006), there were no new codes emerging after six people in a sample of 60, but this could be for a number of reasons. Perhaps their sample was too homogenous, perhaps the interviewers were moving the questions in particular ways, perhaps the researchers looking for particular answers, or perhaps after 60, new insights would have - for some reason - started to emerge. As Low (2019) writes, "there is no magic number to ensure that one will achieve saturation. People who would like to have a numerical formula for determining saturation would have it substitute for researcher conceptual insight and skill. In the end, no amount of information or data collected can ensure against poor research skills" (Low, 2019, p. 135).

We cannot ignore the data in relation to the factors which dictate data collection. My own difficulty with accessing companies also cannot be ignored in relation to the data. We cannot dismiss the fact that research is a "pragmatic activity" which is "shaped and constrained by the time and resources available to the researcher" (Braun & Clarke, 2021, p. 211). However, it is the job of a skilled researcher to judge whether the amount of data is sufficient to answer the research questions they have posed, and to navigate whether any limitations in place will stunt the data collection to the point of pointlessness. For example, we see an incredibly rich and in-depth study of just one research subject who faked credentials and invented stories in order to be part of a research project, being conducted by Owens (2022). Owens uses just one case to present a fascinating perspective on issues surrounding the increasing shift to online interviews in the social sciences.

The data I collected, while not a large sample size, was rich enough to answer my research questions, and therefore worth its limitations. The richness of the data here is, as Richards (1999) says, not only about the data itself, but about how it is handled. My own data is rich in the way that she defines 'richness': relevance (just what is needed for the problem to be studied), impact (evocative and involves the reader), complexity (carry many meanings and raise many topics), and fluidity (ideas must be richly developed together) (Richards, 1999). In some cases, rich data with a smaller sample size can even lead to more concise findings, in contrast to rich data in larger sample sizes, which "may make it difficult to examine data in all of their complexity, limiting ability to probe data collection, develop emergent questions, or contextualize quotes" (Roy et al., 2015, p. 250). My own data is rich as defined by Richards (1999). It is relevant because it was sufficient to answer my research questions, it is impactful because these technologies effect so many people applying for jobs, it is complex as can be seen from the multiplicity

of codes and themes the data raises, and it is fluid as my data has been developed in dialogue with my ideas and the findings emerging.

3.5.4 Data configuration

Another consideration with these methods concerned how to configure the data; whether to present the companies as case studies or to aggregate the data.

The data guided me on this decision. I found that many themes were consistent across different companies, interestingly, regardless of whether they were based in Europe, the US, or Asia. Therefore, I concluded that configuring the companies as case studies would create a disjointed piece of research. Instead, in this thesis, I aggregate the data to explore the common themes. I use website material, report material, interview material, and reflections from the demos or free assessments all together. This worked for the data because the themes were largely consistent across companies, and many of the ways that they operationalised the reduction of gender bias in the system design were similar.

While I focus on individuals in the text, I think about them collectively as well. Often in the text, I switch between referring to individual's views and then company's views; it was clear in the data that there were few discrepancies within companies; often, people within companies would offer me similar views. The individual views were consistent with the collective, but I do refer to them to offer additional elements of nuance.

3.6 Dynamic considerations

3.6.1 Power

One of the biggest dynamic considerations was around power dynamics, a common relation to navigate in interviews with elites (Empson, 2017).

Perera's study of interviewing academic elites argues that power is a set of relations, and the interview is a space where power can be exercised by both parties. Gender, social class, level of education, and professional hierarchies, are just a few elements that impact power relations, along with religion, ethnicity, language, and age (Perera, 2021). This is certainly true; power is dynamic and malleable in these contexts and can shift and change during the interview, even if it is also largely dictated by position in society.

It is true that the issue of power is also discussed in terms of whose voices are heard when researching elites. As Hertz & Imber (1995) point out, some social researchers have issues with researching elites because they believe that to understand them is also to empower them. This is a problem on which Rice (2010) focuses, arguing that the difficulty accessing elites places the researcher in the position of having to maintain positive relations with those they are studying, whilst also developing critical perspectives from the empirical material they have maintained. Researchers therefore can be critical of the data they have collected and continue to highlight important issues of power disparity. As Hertz & Imber (1995) suggest, one strategy could be "to expose the reach of power in the

hope of clarifying it for those who are subject to it" (Hertz & Imber, 1995, p. viii). This is certainly something which I hope to do within this thesis.

While I did have seemingly normal experiences in interviews regarding power relations, in many ways, I didn't.

Often, power asymmetry between the interviewer and interviewee can manifest itself in the verbal and non-verbal communication in interviews and can therefore affect the content of the interview itself. Some argue it can be accompanied by a defensiveness and opacity about the nitty-gritty inner workings of organisations, a desired control over the topic of the work and the manipulation or denial of particular information (Empson, 2017; Ma et al., 2021). This can be evident through speech dominance such as interruptions, back channelling, or lengthy monologues (Boucher, 2017). As Boucher (2017) asserts, "power refers to the capacity of the interviewee to make or resist certain outcomes, with regard to responses to questions" (Boucher, 2017, p. 99).

I did find this true in some respects. Sometimes, interviewees wouldn't answer my question, and instead would speak about something else. In some cases, they would answer in a way which felt rehearsed or like a 'party line', and often this was confirmed through what I found on their websites or reports.

However, in light of my previous experience interviewing elites, I felt prepared for this inevitable power asymmetry during interviews and was well-equipped to handle it.

Firstly, I saw this as an opportunity for covert observation of how these elites were constructing narratives, concepts, and potentially how they are trying to come across, as this does "help researchers capture how business elites construct narratives about their actions, associations, and effects" (Ma et al., 2021, p. 83).

Secondly, I found a good technique for handling this was trying to narrow the gap in power by establishing my own professionalism, knowledge about the area, and knowledge about them. Harvey (2011) and Mikecz (2012) suggest that interviewers need to show they've done their homework before the interview because sometimes elites might consciously or subconsciously challenge them on their subject and its relevance, so the interviewer needs to enhance their knowledge on the interviewees background. Li (2021) also says a common way to navigate power dynamics is to refer to one's own institutional connections and professional status. I found that doing my homework and citing my own experience, usually, worked to narrow the power dynamics in terms of their openness and verbal/non-verbal dominance. During my interview with the Company Red CSM, they asked me:

> Company Red CSM: "First off, well, let me ask you this, how familiar you are with the system?"
> Collett: "Yep, relatively familiar. I've also spoken to [P1] and to [P2], and I'm speaking to [P4] on Friday. So yeah, relatively familiar. I know that […]"

Immediately this changed the tone of the interview whilst leaving room for them to add to my knowledge. Later in the interview, when I had three questions left, I said to them:

> "The first [question] is about how you measure success, and I notice, at least on your company profile, that you look at implementation success with the client. But what I'm really interested in is how you *measure* implementation success with the client."

I also found that the 'Oxford' label worked to my advantage here. I felt this, alongside displays that I'd done my research, allowed the candidates to consider me an equal, or at least to respect my work.

These techniques to narrow the power gap seemed to work, and manifested themselves in comments such as the following:

> "I'm curious what your thoughts are on this, as you're doing the research Clemi." [Company Red HOS]
>
> "One really important thing, that you will know well..." [Alex Vaughan, Former CSO of Pymetrics]
>
> "I know that I'm not telling you anything you don't already know…" [Company Red CSM]

Having said this, with many of the participants I didn't find there was initially a big power asymmetry in terms of their respect for me and their withholding of information. Defensiveness was also not something I often encountered. This surprised me, given how difficult it was to obtain their time. Generally, interviewees were open, even if there was certainly a knowledge and time-availability gap between us. As explanation, it could be that the people who agreed to do these interviews had an interest in this topic and felt strongly about this sort of research taking place. I realise that this is a potential limitation to my sample.

3.6.2 Rapport and fondness

Given the difficulty with accessing these companies due to the hustle, people offering me their time for an interview immediately created a fondness towards them. In some cases, this fondness only increased as rapport was established. In other cases, it dwindled amid

lacking rapport. Almost certainly, this will have impacted not only the interview itself, but the way I analysed data from different participants.

There was one particularly interesting case. The Company Red CSM, as quoted above, asked me how much I knew about the system as one of our very first exchanges in the interview. Our interview began in a slightly reserved manner as we tried to establish our dynamic. As the interview went on though, I felt us both relaxing and establishing a rapport. We even exchanged a laugh or two, and it felt like a truthful, open conversation at times. For example:

> Collett: "What do you tend to see [in terms of] the motivations for why clients want to use the recruitment tech?"
> Company Red CSM: "The reason that they're motivated to do… to use it… is um… for money."
> [Pause]
> Collett: "[Laughs] and that's it?"
> Company Red CSM: "That's it. [They smile] Ok? […] So, organisations invest in diversity, equity, and inclusion programmes because they pay off."
> Collett: "Literally."
> Company Red CSM: "Literally pay off."

By the end of the interview, I felt a fondness towards this interviewee who I'd only been speaking to for one hour. When I emailed after to thank them for their time, they responded: "Thanks for listening ☺".

It was fascinating how the rapport transformed in such a short space of time. I suspect this was partly down to the hustle; I felt lucky to have their time, and perhaps gave some indication of this which might have assured the interviewee they could trust me.

On a separate note, I was aware that the online setting might change the ability to build rapport with the interviewees. There was possibility it might cut across the socio-emotional signals between myself and the interview subjects, which might have been

more readily interpreted face-to-face (Edwards & Holland, 2013). Some believe that nothing compares to face-to-face interviews (Gerson & Damaske, 2020). For example, O'Connor *et al.* (2008) write, "[i]n the disembodied interview all the subtle visual, non-verbal cues which can help to contextualise the interviewee in a face-to-face scenario are lost" (O'Connor et al., 2008, p. 8). There certainly can be challenges with online interviews; participants might feel embarrassed or uncomfortable being filmed as well as things like internet connection and the lack of visual cues (Hay-Gibson, 2009).

However, I didn't find this to be an issue. Internet connection was luckily never a problem and, given the elite nature of the participants, they were usually comfortable being recorded. As Harvey (2011) notes, elites tend to prefer doing telephone or online interviews because of time saving. People were so accustomed to doing online calls because of Covid-19. This likely mitigated some of the issues with rapport building that might have occurred pre-Covid-19 with doing video calls. As Deakin and Wakefield (2014) point out, technological advancements have made communicating over large distances easier, and rapport can be built just as well as in face-to-face interviews as in online settings. For example, in some cases it degraded the line between public and private, allowing me a window into their private settings. With the Company Red CSM, they had to go and ask their partner to stop hoovering because "it's like right in my ear". This relaxed both of us, we had a laugh, and it was even good for the rapport. And with Alex Vaughan, Former CSO of Pymetrics, he had to pick up the laptop and take it to the kitchen as we talked because their "coffee is just whistling". Of course, this helped rapport in some ways, and is also a reflection upon the fact that they're fitting me into their busy lives, which is exactly as you would expect from elite participants.

While in some cases rapport felt natural, at other times it felt slightly forced. For many of these participants, their job is to sell and promote this technology. Sometimes, I felt myself being captured by their vision, their pitch, their opinion. I felt myself agreeing. It is hard to communicate how persuasive and convincing these people are. However, I tried to stay aware of the perspective presented to me in interviews, especially when analysing the transcripts. It is true that a person can choose to present a version of themselves, a certain identity or idea, based on who is listening, and it is the job of the interviewer to be aware of this (Beitin, 2012; Holstein & Gubrium, 1995).

3.6.3 Truth and self-censorship

Another dynamic to navigate surrounded how to detect truthful answers. Morris talks about how dishonest respondents have long been an issue when interviewing elites – they might try to present themselves in the best light or portray a particular version of events (Morris, 2009). Gerson and Damaske (2020) talk about social desirability bias; some interviewees might be prone to say what they think the interviewer wants to hear rather than the truth. Most often, I didn't know if they were twisting things to appear in a better light or telling me how the systems worked. As Empson (2017) notes, "[e]lite interviewees are sufficiently powerful to manipulate information and to deny others access to it" (Empson, 2017, p. 59), and gaining the trust of elite interviewees is difficult, meaning their honesty is probably also uncertain.

Were some of them just handing me party lines? In some instances, this was certainly the case, as I could see from repetition between participants and their company marketing material. Did some participants avoid telling me the main issues they find in using this technology around bias? Perhaps. There is no way of knowing for certain.

At times during interviews, it became obvious to me that participants were self-censoring. Before one interview began, the interviewee asked me jokingly, “so I don’t need to put my PR hat on, do I?”. And during a separate interview a participant said to me, “I probably shouldn’t even be on the record on this”. Another said to me, while getting frustrated about the negative perception of these systems, “And again it’s recorded, so whatever”.

As with the power dynamics, my instincts about honesty were different depending on the interviewee, but of course, my own instincts might not have been correct. I could pick up on small cues sometimes when people felt uncomfortable or were considering what they said. I could tell, occasionally, if they felt nervous or were particularly evasive of questions. However, I realise this could have been for several reasons and not just because they were avoiding being truthful. On the other hand, I could often tell when participants were passionate about something, and giving me their personal views, for example, when asked whether these systems could have a positive impact on the labour market, Frida Polli, CEO of Pymetrics responded:

> “They could have a totally positive impact. I mean I think that’s what…ugh… you know…I think this is where I get quite frustrated […] I don’t believe that these systems as dangerous, and they’re so like gone awry and we have to put a clamp on them and like all of this. I don’t believe that.”

Even though I can never know if everything they gave me was completely truthful, in this thesis I take everything said to me by the vendors at face value. I think we must consider the distinction between truth and falsity, especially with matters such as how systems work, but also, it is the vendors’ narratives in which I’m most interested. In addition, I see no good reason why these participants would lie to me in an interview in which they have willingly agreed to participate.

3.7 Ethical considerations

There are certainly ethical considerations relating to both the process of the research and also to the outcomes (Kelman, 1972). Eynon, R., Fry, J., & Schroeder (2008) outline several ethical issues, many of which I considered before conducting this research.

Before the research commenced, I had my CUREC1a form approved by DREC.[12] This mainly addressed three ethical issues: anonymity, informed consent, and confidentiality.

As already mentioned, some vendor companies are anonymised, and some are named. Allowing for this combination was born out of my negotiations with companies. While some companies wanted to be named (for marketing and transparency purposes), some wanted to be anonymised, and some were uncertain. This solution enabled flexibility and was invaluable in terms of access. This was also important to me ethically, as it enabled companies to give their informed consent on whether they wanted to be named or anonymised.

Moreover, this was crucial for copyright purposes. It allowed me to credit the websites and reports of named companies as required by copyright where possible. It also allowed me to still directly quote the websites and reports of those companies which wanted to remain anonymous, as this complies to ‘fair use’; the requirement in the legislation that you acknowledge the ‘author’ doesn’t apply where this is “impossible for reasons of

[12] These have been considered in further detail in my CUREC1a form, which was submitted to DREC in early April 2021 and was approved in early May 2021 (after minor corrections). I made two further minor amendments after this in September 2021 and May 2022, both of which were approved. CUREC reference number: SSH_OII_CIA_21_088.

practicality or otherwise", and maintaining anonymity falls under this 'otherwise' category (Copyright, Designs and Patents Act, Section 30, 1988).

While it might have been better to have all companies either anonymised or de-anonymised, this was the price I paid for entry.

To enable informed consent, there were multiple stages.

First, a senior figure at each company had to read and understand a three-page information sheet about the research. The information sheet enabled them to understand the purpose of the research and the risk. Eynon, R., Fry, J., & Schroeder (2008) define informed consent as when the individual understands the goal of the research and what they are agreeing to do, along with the potential risks and benefits of taking part. The senior figure then had to sign a form which agreed to this information sheet and allowed me to speak to employees at the company. As Gorman *et al.* (2005) talk about, we need approval from a person in authority to access the research setting when doing this type of research. There are copies of the information sheet and consent forms for vendors in Appendix 4.

Second, each individual employee then received the same information sheet and had to sign a form giving their individual consent. This form also specified whether they would like to remain anonymous or be named if, later, their company opted to be named within the thesis.

Third, after the thesis had been written, as agreed, I sent a full draft to each senior figure at the company around six weeks before submitting my thesis. They could then opt to be

de-anonymised through the ‘Anonymisation Addendum Form’ which had to be signed by the senior figure. This meant I had to write the whole thesis under the assumption that all companies might wish to remain anonymous. Therefore, I could have perhaps included more detailed analysis of the companies which are named, and, indeed, I have in some cases since learning which companies are de-anonymised. However, I also think that writing under this assumption kept my analysis and tone consistent throughout and was a price of entry worth paying.

For regulation and legal experts, this process was much simpler: they read the information sheet and then signed a consent form, within which they specified whether they would like to be named or anonymised.

All of this ensured that I was fully transparent with my participants, and that there was informed consent at every stage.

Regarding the analysis of websites, reports, regulatory or legal documents, and demos or free assessments, given that this information is openly and publicly available, I did not seek ethical approval to use this data. However, I did inform participants that if they opted to be de-anonymised, I would also credit the material from their websites and reports in the thesis.

It was important that I notice the risk to companies in terms of IP, competition, and design protection here, and I try to be as careful about this as possible.

Finally, in terms of confidentiality, As Eynon, R., Fry, J., & Schroeder (2008) point out, we must think about the proper and secure transmission of data. Regarding data storage, all data has been handled and stored in compliance with GDPR requirements. All data is stored in OneDrive for Business and documents are password protected and anonymised as far as possible. Interviews were also held on Zoom pro with end-to-end encryption.

3.8 Conclusion

Overall, I have considered in this chapter how accessibility, the data and analysis, and the dynamics in interviews have impacted my methods and how they interact with my research questions. My aim is that these discussions will continue to arise throughout the thesis and that my methods will continue to be seen as integral to my argument throughout.

CHAPTER 4: THE INTRAPERSONAL FRAME

4.1 Introduction: framing gender bias as intrapersonal

There is an image which tends to reappear on AI-rec-tech websites or conferences. It is usually along these lines: a robot – metallic and shiny - standing tall, shaking the hand of a smartly dressed human. Perhaps the human is wearing a suit and tie, or a dress and heels. Sometimes, the image just captures the handshake: the robot meeting the human, the human meeting the robot. It might be that just their fingers are touching – the robot finger and the human finger – a replication of *The Creation of Adam* by Michelangelo, but for the technical age. Or rather than meeting the human themselves, it might be that the robot is staring at a computer screen. The screen contains an image of a person, or the person's assessment results pertaining to a job application. The robot is sometimes touching the screen, calculating whether this human can do the job for which they're applying.

Regardless of the details, these images communicate the same message: the traditional way of hiring - humans assessing humans - is a thing of the past; AI-rec-tech systems make the judgements now. These images also symbolise the first way in which AI-rec-tech vendors frame gender bias in hiring: gender bias is intricately intertwined with the human mind – it is intrapersonal.

In this chapter, I provide the first of three concurrent answers to my research question: how do AI-rec-tech vendors frame gender bias in hiring?

Using interviews with the vendors, along with analysis of their websites, reports, walkthroughs of system demos, and autoethnography of free assessments, I show how all these vendors consistently frame gender bias in connection with the human mind. This occurs both in the understanding which they communicate to me and in the design of their systems. Overall, vendors focus on the superiority of AI-rec-tech compared to people when it comes to addressing gender bias in hiring. Therefore, in this chapter, I outline and analyse the first of three frames: the intrapersonal frame.

Within existing HR and business studies scholarship, AI-rec-tech is often positioned as able to reduce or even eliminate 'human' biases through its processes and functionality (Allal-Chérif et al., 2021; Upadhyay & Khandelwal, 2018). Hmoud & Varallyai (2019), for example, discuss how one of the main advantages of these systems is that they are "neutralizing human biases" (p. 25) and Black & van Esch (2020) discuss how humans are limited in reliability and validity compared to AI-rec-tech due to cognitive biases. However, gender and disability scholars have pushed back on these claims, countering the assertion that these systems can make recruitment fairer by stripping away human ideas on gender and race. Generally, they argue that the core function of AI-rec-tech is to assess and observe, and this is caught up in a wider system of power and categorisation. Additionally, these scholars have argued that AI-rec-tech often reflects human biases present in our society (Drage & Mackereth, 2022; Tilmes, 2022).

In this chapter, I show that the relationship between humans and AI-rec-tech is a complex one. I argue that while the vendors demonstrate a thorough and well-rounded understanding of human perceptual biases, it is certainly difficult to justify the claim that these systems remove, or even reduce with any certainty, human biases.

4.1.1 Chapter outline

This chapter proceeds as follows:

In the first section, I detail the vendors' conceptualisation of gender bias as intrapersonal. I describe the ways in which vendors understand gender bias as intertwined with the human mind, specifically in relation to three tendencies of human perception, which I refer to throughout collectively as 'human perceptual biases': pattern recognition (stereotypes), homophily, and subjectivity. I demonstrate in this section that these views encompass the two main themes present in literature on gender bias in hiring and in the workplace (stereotypes and homophily), and how they also emphasise a topic which is less discussed in existing literature in relation to bias (subjectivity). I note that these bolstering bodies of work legitimise and support the vendors' conceptualisation, and that overall, the vendors' understanding of gender bias in relation to the human mind is thorough and well-grounded. This section therefore answers my first sub-research question in relation to this frame:

Sub-RQ#1: How do AI-rec-tech vendors conceptualise gender bias in hiring?

In the second section, I outline how the vendors aim to address intrapersonal bias through their system design, first analysing the design mechanisms attempting to 'remove' these human perceptual biases, and then those attempting to 'improve' (or in other words, reduce) them. This, I argue, aims to 'design out' the intrapersonal tendencies outlined in the previous section through features such as anonymisation, re-defining fit, and bias reminders.

As I outline these design features, I analyse them and argue that there are certain ways in which the design still enables the infiltration of human perceptual biases. I label these as 'breaks in the frame', which I adapt from Goffman's idea of 'breaking frame' (Goffman, 1974). These breaks are caused by inconsistencies between the vendors' understanding of gender bias and their ability to mitigate, or even reduce with certainty, these biases through the technological design. These breaks therefore emerge where human perception does, or can, infiltrate alongside the use of these systems. Specifically, I outline five different 'breaks in the frame': the optionality break, the hybrid break, the ignorable break, the profile-shaping break, and the contradictory break. This section therefore answers my second sub-research question in relation to this frame:

> **Sub-RQ#2:** How do AI-rec-tech vendors attempt to operationalise addressing gender bias in their system design?

In the third section, the discussion, I review the argument of this chapter and demonstrate how it answers my research question(s). I also argue that these breaks in the intrapersonal frame are inevitable, due to the necessity of human choices up and down stream when using these systems. While some vendors do acknowledge that there is no guarantee these systems will remove or reduce human biases, others assert that they do. These 'breaks in the frame' are especially important to acknowledge for those companies which claim AI-rec-tech can and does 'remove' human biases. In this section, I also discuss how these breaks illustrate important areas for future research.

4.2 Re-stating ideas on gender bias as human perception

Vendors websites, demos, reports, along with the vendors themselves, constantly displayed reasons why their systems are superior to human-based recruitment. Not only do these systems 'save time and money', these sources told me, not only are they 'scalable' in a way which the human brain could never handle, but crucially, they can 'remove' or 'reduce' human bias.

One website quote summarised this well, claiming that AI-rec-tech can "increase accuracy, save time and money, and remove human bias", as did a Pymetrics company report which expressed that "AI can be a powerful tool not just to mitigate bias, but to actually remove it altogether in a way that is not possible with humans" (Pymetrics, n.d.-a). Frida Polli, CEO of Pymetrics also told me:

> "I actually think […] we can do better than humans. […] look at, like, the workforce today and say did algorithms get us there or did humans get us there?"

Consistently, gender bias was written or spoken about in relation to humans, specifically the ability of these systems to outperform humans in terms of bias. For example, Frida told me that one of the biases they remove (across all demographics) is "human bias" [Frida Polli, CEO of Pymetrics], and another participant reassured me that companies were better off using their tool due to human's conscious and subconscious biases:

> "But to me there's much more bias in letting humans do it by themselves […] If you're meeting someone face-to-face or talking to them on the phone, you're much more likely to have a conscious or subconscious bias against that person […] I think the irony is that it's probably the other way around where there's more bias if they don't use a tool like ours." [Company Red HOS]

In this section, I unpack this underlying understanding of the vendors that gender bias is tied to the human mind and, in doing so, I answer sub-RQ#1 regarding this frame: how do AI-rec-tech vendors conceptualise gender bias in hiring?

In Chapter 2, I overviewed literature within psychology, sociology, and gender studies which specifically deals with human's gender bias in hiring and the workplace. I showed that the discussion is split into two broad categories: stereotypes and homophily. Stereotypes are commonly referred to as giving rise to biased judgements in the workplace (Heilman, 2012; Heilman et al., 2024; Hoover et al., 2019; Luzadis et al., 2008; Nadler, 2017). Additionally, gender biases are often defined with reference to 'similarity' or 'in-group' preference (Black & van Esch, 2020; M. Carlsson & Eriksson, 2019; R. Carlsson & Sinclair, 2018; Friedmann & Efrat-Treister, 2023) which is often referred to as 'homophily' in sociology (Lazarsfeld & Merton, 1954; Mcpherson et al., 2001).

The vendors' understanding of gender bias in the intrapersonal frame highlights these two main strands of the literature on human gender biases ('stereotypes' and 'homophily'). While not arguing this to be a causal link between existing literature and their views, I do contend that the conceptualisation of this frame is strong; this legacy of research gives weight to the views of the vendors and shows their well-grounded understanding of gender bias.

More surprisingly, however, I found that the vendors also placed high importance on 'subjectivity', more so than has been previously seen in the literature surrounding gender bias in hiring or the workplace. In the literature, subjectivity has been interpreted as something which leaves room for biased judgements through stereotypes (Heilman & Haynes, 2008), but the vendors spoke of subjectivity in isolation from discussion of stereotypes. I argue that this adds an extra layer to existing work on human perceptual biases and hypothesise about why vendors might be putting emphasis on subjectivity as an aspect of human bias.

In this section, I therefore show that vendors conceptualise gender bias in hiring as deeply intertwined with three human perceptual biases: pattern recognition (stereotypes), homophily, and subjectivity.

4.2.1 Pattern recognition (stereotypes)

> "Human brains recognize and learn from patterns in the environment." (Ingersoll, 2020) [Pymetrics blog]
>
> "Our brain is capable of matching patterns. However, this pattern matching does not always give accurate results." [Company website]

One of the tendencies of human perception which formed part of the vendors' conceptualisation of gender bias was pattern recognition. Tigran Sloyan spoke about how human beings are "pattern matching devices", linking this to biases in the mind:

> "So, like, just like, where does human bias come from? Because we're pattern matching devices, when you observe something in the world, and even if there is no pattern, you just make it a pattern in your head." [Tigran Sloyan, CEO and co-founder of Code Signal]

Later, Tigran clarified to me that it is these 'patterns' which define human biases:

> "To me, bias is a form of association or pattern that you've created in your head that influences your judgement." [Tigran Sloyan, CEO and co-founder of Code Signal]

The conceptualisation of gender bias in humans in relation to pattern recognition chimes with literature on gender stereotypes, which has long been discussed as a form of pattern recognition.

As discussed in Chapter 2, gender stereotypes encapsulate perceived patterns in qualities and roles which pertain to certain genders more than others. Gender bias in the workplace is therefore often perceived to relate to these patterns, for example, whether someone

'fits' the role in light of these gender stereotypes (Heilman, 1983, 2001; Heilman & Caleo, 2018; Wynn & Correll, 2018). This is what Eagly & Karau (2002) refer to as 'role congruity theory', which links bias to congruence between stereotypes held about job requirements and stereotypes held about gender groups. For example, some studies have shown that when candidates violate stereotypical norms, women and men are punished in terms of how they're ranked at CV stage (González et al., 2019; Hoover et al., 2019).

While I'm not suggesting that the vendors have necessarily formed their ideas through engagement with this literature, research on gender bias in relation to stereotypes resonates with the vendors' understandings around gender bias in humans as pattern recognition. In essence, the 'patterns' to which vendors refer resonate with ideas on gender stereotypes surrounding work, and therefore who is more 'suited' to a particular job. As Tigran Sloyan, CEO and co-founder of Code Signal discusses:

> "[B]ecause we're pattern matching devices, because historically more men went into computer science than women, you end up with a situation where like, our brains observe the world and say, like a machine learning algorithm, oh, engineer, men, female, not an engineer, and then you just make that a pattern and then you start believing that pattern."

Frida Polli, CEO of Pymetrics talked about how this relates to system one and system two thinking, referencing Daniel Kahneman's book *Thinking Fast and Slow* (Kahneman, 2012). System one thinking is fast processing, system two is more deliberative. Frida told me that she believes system one thinking underpins a lot of our decision-making, especially when it's being done quickly, for example in the case of hiring. Looking at a resume for six seconds is not deliberative – it's being done quickly, and therefore it is utilising patterns engrained into the human brain, and this, she says, is biased. Frida expanded upon this, saying:

> "We live in a biased world. If women and men were CEOs at the same rate, or black people and white people were proportionately represented in the workforce,

> our biases would be different, and we wouldn't have this problem. But because we learn from patterns that we see in the real world and the real world unfortunately is biased…" [Frida Polli, CEO of Pymetrics]

In order to emphasise the stubbornness of this tendency to perceive patterns, Frida also cited the failure of unconscious bias training in changing the landscape of bias in hiring (Dobbin & Kalev, 2016; Wynn & Correll, 2018):

> "For too long we've been trying to change the human brain and, again, ten years' worth of meta-analysis on unconscious bias training shows it doesn't work. It's ineffective. Because, again, like, you're trying to change behaviour with like a four-hour training. When has that ever happened in the history of behaviour change? You know, never…" [Frida Polli, CEO of Pymetrics]

Tigran Sloyan, CEO and co-founder of Code Signal said something along similar lines, telling me that AI-rec-tech system design is "easy to control and modify because it's just code, right? You go in there and you change it. You cannot reach into a human's brain and change the algorithm that runs in your head".

As well as reference to pattern recognition, there was also explicit acknowledgement of the relationship between human biases and stereotypes. As one blog post on the Pymetrics website said: "[b]iases are the shortcuts used by the human mind to process information, based on expectations that are shaped by personal experiences and social stereotypes" (Ingersoll, 2020), and another website read: "[u]nconscious biases, or implicit biases, are social stereotypes about certain groups of people that individuals form outside of their own conscious awareness. Everyone holds unconscious beliefs about various social and identity groups, and these biases stem from one's tendency to organize social worlds by categorizing".

Overall, pattern recognition (stereotypes) was one aspect of how the vendors conceptualised gender bias in hiring in relation to human perception. The biases affecting hiring outcomes were, to vendors, often related to their tendency to prefer abiding by established patterns or stereotypes relating to gender and work.

4.2.2 Homophily

"Familiarity breeds comforts" read one Code Signal company report (Code Signal, 2020). This quote encapsulates the second aspect of how vendors conceptualised gender bias to be linked to the tendencies of the human mind; the human preference for those who reflect us. As the Company Blue CPO pointed out:

> "Most people are… they prefer people like them. And for that reason, they're not able to make those decisions objectively."

This was also demonstrated in what the Pymetrics website called the 'same as me' bias, which can mean that 80% of candidates are never even seriously considered, and which the system can 'remove' (Pymetrics, 2022c). Often, participants, websites, and reports spoke about the preference for candidates who are 'the same' or 'similar' as a consistent factor which captures bias in hiring; humans often want to recruit those who are like them, and this leads to biased decisions. Reflection of recruiters or the company at large was conceptualised as an innate part of biases stemming from humans, and something, therefore, which they tried to tackle in the technology.

Warnings against this human tendency were often present in marketing materials or interviews. As one of the company websites pointed out "[i]t's human nature to hire people like ourselves, so it's imperative to leverage a variety of networks and tools to

support hiring based on diversity of backgrounds/perspectives". In addition, during one of the company demos, it was said to me that "[i]t's human nature to lean toward people who are just like us" and that their system can be used to "avoid that".

The Company Red CEO also commented that the AI-rec-tech system 'eliminates' this desired reflection in potential candidates:

> "So, it eliminates the human error of hiring who you like and hiring people who are similar to you. Um because currently that's the system. Currently it's all bias and it's all who you like and the best example is, we've all been there: the five people at the company, the boss and let's say the four team members, meet the final five candidates, they interview them and at the end of the day they rank them on a wall like on a blackboard or on Zoom on a whiteboard and they say who do we like the most? And they all agree that Stephanie or John or Mark is the one they like the most well, guess what, that's complete biased." [Company Red CEO]

Further to this, the Company Red CTO told me that "what we don't want is, you know, people to use this [the AI system] to defend their ability to, you know, hire people like them". They went on to say:

> "So, then the product supports that methodology so that in some ways it actually, um, highlights bias, um, you know, when a manager is just hiring everybody that looks like them, right, and sounds like them and thinks like them, it shows where that difference is. […] Um I also feel it's kind of removed some [bias]. […] I'm interviewing people and I'm like, wow, I like these people cos they're all like me, but I may not have known that right?" [Company Red CTO]

Again, the vendors emphasis on how preference for similarity is innately linked to the biased tendencies of human perception matches one of the two prominent strands in existing literature on bias in recruiting and the workplace as shown in Chapter 2. There, I highlighted that humans tend to select people with whom they feel some affinity (Kahneman et al., 2021). This is often referred to as 'similarity bias' (Black & van Esch, 2020) or 'affinity bias' (Heidari et al., 2023) and commonly seen to occur in relation to

gender in recruitment settings (M. Carlsson & Eriksson, 2019; Friedmann & Efrat-Treister, 2023).

4.2.3 Subjectivity

Lastly, gender bias was spoken about by the vendors in relation to the human tendency for 'subjectivity'. 'Subjectivity' was referenced by many of the participants as constituting the biased nature of human perception. One participant discussed how, for humans, it's "hard to bring in consistency overall, especially in recruitment practices, because everyone will have their own subjective opinion about everything" and "recruitment works with the subjective opinion of the interviewer". He continued by saying that this subjectivity means that "it [hiring] is bound to have biases" [Chanukya Patnaik, CEO of Skillspace.ai]. From the other side, another company's website communicated that AI-rec-tech systems will allow you to "eliminate subjectivity in the hiring process", and that they are able to reduce subjectivity by 80%.

Others also described the human-based recruitment process as 'subjective' and linked this subjectivity to the biases within hiring. For example, the Company Yellow IOP1 said that:

> "[A] lot of traditional ways of recruiting candidates where, ok we're gunna do an interview, maybe a phone screen, then we'll invite them to an interview and then we'll make a decision. Those things can oftentimes be very subjective. Because there's one interview and then they meet with the person and then they might have, of course, not even on purpose, but they do maybe have some sort of bias or whatever it might be."

Not only was human perception labelled as subjective in the hiring process, but there was also a perceived 'objectivity' to using AI-rec-tech systems:

> "[T]here's a more objective way of doing it versus, I think, in the technical interview process, a lot of it is based on people's perception of how this candidate

> performs during the interview." [Sophia Baik, COO and co-founder of Code Signal]
>
> "The client can now easily identify those candidates that have the traits of a high performer without subjective judgment." [Company report]
>
> "There are advantages in that this is in some ways more objective. You don't have a human doing the point of soliciting the data." [Company Blue CSO]

The vendors' idea of subjectivity as linked to human gender biases in hiring is not a focus which is commonly seen in literature on gender bias in hiring. Therefore, it is interesting that the vendors seemed to emphasise this.

However, that is not to say it has not been mentioned in the literature surrounding gender bias in the workplace. Psychologists Heilman & Haynes (2008) discuss subjectivity in relation to performance appraisal in the workplace. The pair link subjectivity to bias, asserting that "subjectivity derives from the critical role it can play in giving rise to biased evaluations and their subsequent behavioral expression" (Heilman & Haynes, 2008, p. 127). They define subjectivity as:

> "[A]n evaluative orientation that requires inference because judgements are based primarily on outcomes and criteria that are open to interpretation. This can be contrasted with 'objectivity', an evaluative orientation in which the need for inference is minimal because judgements are based primarily on concrete outcomes that are independently arrived at and measured against a prescribed reference point for comparison." (Heilman & Haynes, 2008, p. 128)

Therefore, while subjectivity is defined by room for 'inference' because judgements are 'open to interpretation', objectivity lowers the need for inference because 'concrete outcomes' which are 'measured' through data and give points for comparison. This resonates with the views of the vendors, who largely connect subjectivity to human perception, or 'inference', and objectivity to data which can be measured.

However, this definition also differs from the views of the vendors somewhat. The vendors tended to refer to 'subjectivity' in isolation from stereotypes and homophily. Subjectivity was understood as another way in which human perception might be biased. However, Heilman and Haynes argue that subjectivity leaves room for bias, specifically stereotype-based negative performance expectation, to influence judgement, for example a lack of fit between stereotypes of female attributes required to perform male stereotyped jobs. Therefore, subjectivity is not communicated by the vendors as linked to stereotypes, but rather as an independent trait of gender bias in humans.

The fact the vendors tended to discuss subjectivity in isolation from stereotypes and homophily – and indeed, discussed all these three tendencies separately but in relation to human biases – is an interesting finding. It shows the new direction of vendors' understanding regarding gender bias in hiring. This focus on subjectivity might be taking place to highlight how their systems, in contrast, are 'objective', which might appeal to many of their clients in emphasising the role of data in measuring candidates, and the superiority of these systems in comparison to humans. Indeed, emphasis on all these three human perceptual tendencies position their systems as a preferential alternative to human-based recruitment.

Subjectivity, therefore, demonstrates another tendency of human perception which the vendors conceptualised to be linked with gender bias, and another piece of the puzzle contained within the intrapersonal frame. It plaits together with the two other strands discussed above to form the underlying understanding of the intrapersonal frame of bias.

Together, these three strands represent a well-grounded understanding of human perceptual biases. The bolstering bodies of work which already exist on these topics legitimise and highlight the thorough nature of vendors' conceptualisation. Overall, the vendors' understanding of gender bias in relation to the human mind is a strength of the frame – it is thorough and well-grounded in theory and evidence, as seen from the overview of this literature in Chapter 2. Clearly, this frame resounds with a broader understanding of bias and is vital to consider when it comes to recruitment practices. Not only do these views resonate with much of the literature on gender bias in hiring and the workplace, but discussion of subjectivity by the vendors brings another layer to this conversation.

In the following section, I outline how the vendors attempt to address this understanding of gender bias in the design of their systems.

4.3 Breaks in the intrapersonal frame: 'designing out' human perceptual biases?

Now that we understand how the vendors are conceptualising bias in relation to human perception, we can move on to discuss how this understanding is operationalised in system design. In this section, I answer sub-RQ#2: how do AI-rec-tech vendors attempt to operationalise addressing these gender biases in their system design?

Here, I describe how the vendors continue to frame gender bias as intrapersonal by endeavouring to 'design out' the tendencies of human perceptual biases described above. I argue that some vendors' design features which attempt to 'remove' human perceptual

biases and that others attempt to ‘improve’ (in other words, reduce) human perceptual biases.

Overall, through analysing this design, I argue that there are ‘breaks in the frame’. I define ‘breaks in the frame’ as: inconsistencies between the conceptualisation of gender bias and the ability of the operationalisation to address this conceptualisation.

These breaks appear when evaluating design features, and they occur due to the presence of human intervention and decisions both up and down stream of the design. These breaks therefore represent places where human perceptual biases (and therefore stereotypes, homophily, and subjectivity) have the potential to influence rankings or decisions surrounding candidates, and therefore impact the design being able to function to remove or reduce these biases with any certainty. These ‘breaks in the frame’ are a type of vulnerability: complexities within or surrounding the frame which impact its integrity or function.

Goffman’s idea on ‘breaking frame’ illuminates this notion of ‘breaks in the frame’ further (Goffman, 1974). His theory is that there are certain things that can disrupt the organisation of activity and therefore break the frame. Goffman discusses this in terms of social events. For example, when someone is in the theatre playing a part on stage, if they laugh or come out of character, this will break the frame. Breaking frame, in essence, is when someone leaves the frame, the interpretation of what is going on, while others are still there. It communicates an inconsistency. However, in this case, we see that another form of breakage is possible. While Goffman speaks of breaking frame by stepping

outside of the frame, in this case, we see breakage occurring through an inconsistency between understanding and execution. We see breakage *within* the frame.

Before I outline the vendors' operationalisations within the intrapersonal frame, there is a caveat which must be noted. Claims to remove or improve human perceptual biases and therefore lessen biases are made by the vendors to varying extents. In the data already displayed in this chapter, the following phrases are used when comparing AI-rec-tech to humans: there might be "more bias if you don't use a tool like ours", the systems are seen to "eliminate subjectivity", "avoid" bias, "eliminates the human error", "remove" human bias. This ranges from claims that these systems can 'remove' human bias – and, indeed, this claim from vendors has been found to be the case within other studies (Drage & Mackereth, 2022) - to claims they can 'reduce' human biases.

Views on the extent to which AI-rec-tech could reduce or remove human biases varied not only between companies, but between individuals. While some informants did claim that these systems can "remove that subconscious bias" [Company Red HOS], others said that it is "impossible to take all bias out" [Head of Impact at Pymetrics], or that it is "really hard to eliminate bias altogether" [Company Yellow IOP2]. As the Company Orange CPO said to me: "but it is only a tool at the end of the day. You could have […] the most fancy deep blue of AI technology […] and you're always gunna have a manager that has to have that interview or that interaction and says I don't like people who wear brown shoes".

While the claims varied, and this is important to note, all participants were unanimous in their argument that their systems were less biased than humans. The caveat, therefore, is to note this variation, and that I use 'remover' and 'improver' to reflect the action that

their design takes on human perceptual biases. It does not unanimously signify the vendors' voice on the extent to which these systems remove or improve human bias.

With this in mind, I turn to explore how the vendors attempt to 'design out' gender bias stemming from human perception.

4.3.1 The remover?

The first way in which vendors attempted to address the gender bias in their system design was by removing the ability for human perceptual biases - such as subjectivity and homophily - to influence decisions. This was done through one key feature: anonymisation.

Often, the word 'blind' or 'anonymous' would emerge in company marketing materials or client facing material. The Pymetrics website said that their system allows for "[a] true blind audition. We don't look at your gender, ethnicity, pedigree, or work history" (Pymetrics, 2022b), while another company's website spoke about how "[b]lind hiring is part of this larger change that aims at minimizing bias in the workplace".

In these assessment technologies, there are two ways in which candidates are, or can be, anonymised:

(1) **The algorithms are blinded to demographics**. While demographic information can in certain circumstances be used to audit and assess the impact of certain tests, it is currently not lawful, at least in US and UK legislation, to use information on protected characteristics within the algorithms decision-making process as this

would constitute disparate treatment (Kelly-Lyth, 2021). Alex Vaughan, Former CSO of Pymetrics emphasised this to me, telling me that while they do collect this information (although optional to complete), companies are not legally allowed to use demographic information at any stage in the actual hiring process. Vendors often argued that this means the algorithm is less likely to discern patterns related to demographic or background.

(2) **The recruiters can be blinded to the demographics of the candidate.** During the post-assessment shortlisting or selection process, the recruiters can be blinded to information which might allow them to perceive any demographic information on the candidate. As I saw from demos of the back end (the employer's view), this usually takes the form of blanking out the candidate's name (which is required information), and any optionally completed information concerning their demographics.

I will discuss both types of anonymisation in turn, examining the first of these only briefly, given that it is not a feature of design, but rather a legal requirement.

The first type of anonymisation occurs when the algorithm is blinded to demographics.

This is something which has led some scholars to claim that AI can "avoid unconscious bias" because it "can ignore primary sources of bias like names, schools attended, gender, age and race" (Upadhyay & Khandelwal, 2018, p. 256). This anonymisation is required by law, because legislators interpret that otherwise the algorithm could learn to rank candidates using discriminatory indicators such as gender or ethnicity (Kelly-Lyth, 2021). This might take a similar form to the Amazon screening tool, which learnt to downgrade

CVs with ‘women’s’ in the text after being trained on successful candidates (Dastin, 2018). Instead, the algorithm only uses data surrounding that person’s skills or behaviours – factors which are argued to be relevant to the role for which they are applying (Allal-Chérif et al., 2021). But due to this, demographic information also cannot be used to reduce the discriminatory nature of an algorithm. Therefore, vendors produce models that do not take protected attributes as input (Raghavan et al., 2020).

The Company Red Partner, for example, explained that this is the case with their tool:

> “[W]hen we run people through our tool, yeah, we don’t know who they are […] right, we don’t know if they’re male, we don’t know if they’re female, we don’t know if they’re Asian, if they’re white, if they’re Hispanic.”

The advantage of blinding or anonymising job candidates to measure primarily based on skill has been touted as a technique for reducing bias for decades. One famous example of blinding as a mechanism to reduce bias is from Goldin and Rouse (1997) who adopted ‘blind’ auditions for symphony orchestra auditions - they used a screen to conceal the identities of the players, and, in some cases, a carpet to muffle footsteps so that their gender couldn’t be inferred. Through testing this at eight major orchestras, they found that the screen increased the probability by around 50% that a woman would be advanced from preliminary rounds. In some cases, then, blinding or anonymising does allow avoidance of making hiring decisions based on occupational stereotypes or homophily.

In the fairness auditing literature, this is referred to as ‘anti-classification’ - an algorithm is blinded to protected characteristics or proxies from which protected characteristics can be inferred (IFOW, 2020; Rovatsos et al., 2019). Often, it is these proxies which can create an issue and can maintain the legacy of human perceptual biases. For example, Kleinberg et al. (2018) found that even when blinding the algorithm and excluding race,

this protected feature could still be reconstructed from other factors. They name this the 'reconstruction problem' and argue that while blinding might feel fairer, it might do more harm than good. Were it not for legal constraints, they argue, one should include variables such as race and gender for fairness reasons. They use the case study of college success to make admissions decisions – underlining how the inclusion of protected variables (and the use of different thresholds to admit white and black students who might have obtained different levels of guidance during school) improves predicted GPAs of admitted students and the fraction of admitted students who are black (also see Gillis & Spiess, 2019).[13]

A generalised technical conclusion about blinding the algorithm in recruiting contexts is outside the remit of this thesis. However, it is interesting to note that there are contexts in which blinding the algorithm does not necessarily remove the legacy of human perceptual biases due to the 'reconstruction' problem which captures the history of human discrimination against certain groups.

The second type of anonymisation occurs when the recruiters are blinded to any potentially identifying information on the candidate during shortlisting.

While not all vendors included this feature, for those which did, the system blinded the recruiter to factors such as age, gender, ethnicity, educational institution, name, picture, address, and email address at the point they are selecting the candidates for interview,

[13] Of course, this idea of different thresholds speaks to the topic of affirmative action in the US, which has long been a divisive subject. This grants special consideration to historically excluded groups, particularly racial minorities, and women. Nine states in the US explicitly ban affirmative action in the employment process, and in June 2023 the Supreme Court ruled that race can no longer be considered as a factor in university admissions as they argued it violated Title VI of the Civil Rights Act 1964, which prohibits discrimination based on race, colour, or national origin (Debusmann, 2023).

shortlist, or the job itself (depending on how the client uses the system). It aims to counteract all the tendencies outlined above by removing the ability for humans to make subjective judgements or those based on stereotypes or homophily. For example, the Company Green CEO said:

> "[T]here's just kind of just a block on all that personally identifying information until you shortlist the candidate at which point the decision has been made and, of course, you need to contact the [candidate]."

As mentioned above, blinding humans to demographic details has, in the past, been advocated as a way of reducing human biases (Goldin & Rouse, 1997). Heilman and Caleo (2018) also argue that leaving a person's name or identification out of materials precludes decision-makers from the influence of gender stereotypical inferences, because they lack knowledge of the person's gender. As one company website reads, anonymisation "encourages hiring managers to be more objective while making recruitment decisions. When they don't know about backgrounds, they tend to focus more on the capabilities of the candidate".

However, I saw how using this design feature was sometimes optional, and therefore candidates can be easily de-anonymised. In one of the demos, the vendor guided me through the back end of the system. There was a table with a candidate ID (in number form), the test date, the score, and candidates are ranked accordingly. To replace the ID with a name, you just click to disable this feature (although it does give you a strongly worded warning).[14] From talking during the demo, I understood that the ability to de-anonymise was there because once the shortlisted people were decided, the candidates must be contacted. However, it was interesting that the candidates could be de-

[14] The algorithm is still blinded to demographics of the candidate, so this is within the law.

anonymised before the shortlist was finalised, and I interpreted this to be a means of giving flexibility to their clients. For all the supposed commitment to reducing opportunities for human biases to emerge, the vendor had chosen to design in the ability to disable the feature before it was necessary. Of course, this is not the same for all systems.

During the demo, I felt curious to de-anonymise the candidates to see who had the top and bottom scores on the tests. While the vendors did assure me that they understand most of their clients use the feature continuously, this did make me wonder about whether the clients might disable it, just for a moment, and it did make me wonder whether this could impact their judgement and enable subjectivity.

This leads to the **first break in the frame - the 'optionality break':** optionality to use design features which sequester human perception. When it comes to removing human perceptual biases through anonymisation, this optionality in the design enables clients to de-anonymise candidates as they wish and, therefore, allows for the infiltration of human perceptual biases.

Some participants did recognise this, telling me that while anonymisation was effective, it wasn't fool proof. The Company Red HOS, for example, argued that it's not necessarily true that these tools can solve the problem entirely, because if someone has conscious bias, using the tool is not going to prevent them when it comes to the final decision, but a tool like theirs can "kind of put those blinders on to gender and race and allow you to look at something else, right?" [Company Red HOS].

Although this thesis cannot claim to know how clients use the systems, it can say that the design means these systems can be used in a variety of ways due to feature optionality. Therefore, the use of anonymisation features is dependent on the variable incentives and choices of the clients and, naturally, these will be assorted.

The Company Red CEO told me that they usually ask their clients not to look at the names:

> "Because you can have a neutral tool that's also creating gender bias the way it's used. Um, ur, we often ask our clients […] not [to] look at names, okay? And, and look at the results as numbers. As literally, regardless of where the persons based, what their skin colour is, demographic, race, religion, or name."

However, there was recognition that once the system was handed over to the client, it was no longer within their control how it is used. As the Company Red CSM told me:

> "[W]e are a ur a vendor or a service provider right […] We can't control you know, ur, how racist the interviewer is or how sexist they are […] there are boundaries drawn between what our customer's responsible for controlling and what we are responsible for controlling […] But we also will work with, ur, let's say as part of our, ur, implementation process we'll run a training programme for recruiters. And in that training we'll talk about subjective bias, we'll talk about the, ur, you know, what to watch out for when you find yourself, you know, being biased, you know. It's a tough nut to crack because, generally speaking, people just wanna get their job done […] and most people don't wanna do the work it takes to, ur, become that aware of their own tendencies around bias um so they just kind of do what's always been done."

While some of the companies do consult their clients about anonymisation features in terms of how and why they should use it, understandably, they cannot control the client's actions. As the Company Blue CSO told me: "how they administer all of it you know, all we're providing is the assessment and guidance. Everything else is on the client".

This optionality and autonomy therefore leave a vulnerability in the frame - a potential gap for human perception of candidates which incorporates subjectivity, stereotypes, or homophily through the ability to de-anonymise.

Additionally, in the context of anonymisation and client usage, vendors often spoke of their systems as being 'support' mechanisms for recruiters. While in some cases the candidate will be anonymised, in others they won't be, and in all cases, candidates will be de-anonymised at shortlist stage. At this point, even with the guidance and use of the system, humans are involved in decisions surrounding candidates. This brings to light the **second break in the frame** which emerges from discussion of removing human perceptual bias through anonymisation: **the 'hybrid break':** the continued human involvement in parts of the selection process such as final selection stages and interviewing.

At conferences, in reports, and during interviews, I would often hear or read the word 'hybrid' in relation to the type of hiring which these systems offer. "Embrace hybrid recruitment" one speaker said enthusiastically, their arms open, at one of the conferences. "This is the hybrid hiring era!" said another person at the same conference (IHR Recruitment Technology Virtual Conference, 2021). Often, these systems were marketed, and designed, as support mechanisms to help humans make the decision. As the Company Blue CSO noted:

> "So, well-designed and monitored assessments […] serve as an objective additional data point to help hiring managers and recruiters get a fuller picture of the candidate."

Therefore, humans are still very much involved in the final decision steps. For example, they often still conduct the interviews when it's down to the final few. While many of the

companies noted that bias stemmed largely from human perception, they would also encourage the human eye when it came to assessing each candidate, watching their videos, or conducting the final interview:

> "[B]asically, you make the decision on your own, it's just giving you the data." [Company Red HOS]
>
> "We give the score, I believe, as a starting point, but you should be reading and watching all these video interviews yourself." [Company Yellow IOP2]
>
> "You've still got to take the decision yourself, yeah. There's no technology that will be able to line up every single factor for example in a recruitment process. We can add a bit, but there is still plenty of other aspects that you should be thinking about." [Company Red Partner]
>
> "Within recruiting it's always 100%, there always should be some sort of human interaction." [Company Orange CPO]
>
> "Now I don't think that any of the decisions you make should be made solely on the data points that come from those tools, but it's an additional data point that we haven't had before and that can help." [The Impact of AI and HR on Future Work conference, CogX, 2018]

As is clear, oftentimes the vendors would encourage this hybrid model so that clients can make decisions which best suit them, or perhaps to maintain the notion of 'human-in-the-loop' which has so often been seen as crucial for the responsible and accountable use of technology (Harris, 2018; Mellamphy, 2021). However, it is impossible to ignore the paradox here. The vendors are framing gender bias as interlinked with human perceptual biases, both through their understanding and their system design, however, they often encourage using the system in a hybrid manner. This, of course, leaves room for the influence of stereotypes, homophily, and subjectivity in the hiring process.

Therefore, with the second type of anonymisation, there are breaks within the frame - inconsistencies between the vendors' understanding of gender bias and their ability to execute the removal of this in the design. Firstly, the 'optionality break' - as seen in the

demos, while anonymisation of candidates to recruiters is possible when shortlisting or selecting, it is sometimes an optional design feature. This ability to de-anonymise reinstates room for the tendencies of human perception to emerge. Secondly, the 'hybrid break' - labelling the systems as 'hybrid' or a 'support mechanism' seems at odds with vendors previously outlined understanding of gender bias as linked to human-based hiring.

This means that there are many ways in which human perception can, and likely does, infiltrate the hiring process when using AI-rec-tech. As I mentioned earlier, this is not something which every vendor is unaware of; not all vendors claim that their systems can remove human perceptual biases completely (although, crucially, some do). But in terms of the frame, this incongruity is important to note.

4.3.2 The improver?

The other way in which vendors attempted to address human perceptual biases in their system design was by improving (and thereby reducing) these biases, encouraging the system to be less subjective and less inclined towards stereotypes or homophily. This was done through two design features of the technology: (1) bias reminders, and (2) re-defining fit.

Firstly, bias reminders.

As discussed above, in many cases, these systems are not used for the full process of decision-making (or rather, they don't make the actual decision) but they are often used

to support the human decision-making process. As the Company Blue CPO said, their system offers "an objective datapoint that should be an input to the decision made".

Some systems therefore aim to improve human perceptual biases by providing the recruiters with reminders, through pop-up messages or similar, of their own conscious or unconscious biases. For example, in one of the company demos, they showed me that when the recruiter goes to de-anonymise the candidates, a message arises which reminds them that unconscious bias is one of the reasons for bad hires and asks them if they're sure they want to de-anonymise the candidates.

The Company Blue HOS also told me that they are building in this sort of messaging:

> "[A] lot of that coaching in the product like, alerts, like, they make sure that this isn't your only decision point […] you have to think about the whole person, not just one part. And we always tell people on demos, this is not a red light, green light solution. You should not just assume I see a five star and it means I should hire them, or you shouldn't assume that a one star can't do the job right."

This is an operationalisation which aims to improve subjectivity in the selection process. It prompts the human to take more time on their considerations, to reflect upon their processes, and to notice the patterns and assumptions in their thinking. However, considering bias reminders in practice sheds light on the **third break in the frame - the 'ignorable break':** the ability to ignore design features in favour of human judgement. This differs from the optionality break because it is not optional to get these reminders, but employers are able to ignore the alerts. This means that there is no guarantee that this design feature would change the employers' decisions in practice.

Secondly, re-defining fit.

Some of the systems were designed so that clients can search for candidates who are the opposite of their existing team, leadership, and/or high performers. This design attempts to re-define how their clients perceive the notion of 'fit' or 'success' in a particular role, in a bid to help them select more diverse candidates. Therefore, it aims to avoid the tendency of 'homophily' as well as subjectivity. As it said on one of the company websites: "fit does not mean hiring who you or the team likes but rather being open to hiring people unlike you or any others in your organization".

The Company Red system was the main proponent of this design. In the system - which measures people in terms of the 'culture' they bring to the company - there was the option to filter candidates by who is the most like their existing top performers, or by who is the least similar. In other words, client companies had the option to 'clone' their top performers, ideal employee, team, or manager, but they also had the option to see the most 'diverse' candidates, meaning those who were least like their high performing employees, ideal employee, team, or manager.

The Company Red CEO told me that this means if companies want diversity, if "this person is the most different from the team, regardless of gender, then go with this person, and so that's another way that it avoids gender bias". This is also suggested by Kleinberg et al. (2020), who describe that because algorithms are built to predict past hiring decisions of the firm's human HR managers, the solution is to swap in different outcomes to be predicted that are less tainted by human discrimination.

The Company Red CTO continues by explaining to me that this section of the interface also highlights biases to their clients:

> "[I]n some ways it actually, um, highlights bias, um, you know when a manager is just hiring everybody that looks like them, right, and sounds like them and thinks like them it shows where that difference is." [Company Red CTO]

Ultimately, the Company Red argued that they're altering the idea of 'fit' within the workplace, because although "everyone for the last 50 years has assumed that best fit means similar" [Company Red CEO], they emphasise that while 'fit' could mean similar, it could (and should) mean completely different:

> "Perfect fit, best fit could be similar, like identical […] or it could be someone who complements or adds culture, right? So, our platform takes a more global concept of best fit, right, and it allows a company to dial, to codify, to literally programme the system, push a couple buttons, and literally a recruiter doesn't have to think and doesn't put their subjectivity into what they think is a good fit with the hiring manager, the team […]" [Company Red CEO]

The company also emphasise that everyone can adapt, meaning that your score is only an indicator, and doesn't necessarily show your potential:

> "We have what we call our four principles, right. One is, everyone can adapt. Ok? Which means that it doesn't… like you might say that we do have a preference for people who are high on altruism for a job where customer service isn't involved […] However, that does not ever mean that somebody who's low on altruism couldn't do the job effectively, ok? Because people adapt, right?" [Company Red CSM]

A similar theme of people being adaptable also came up with the Company Blue CSO, who said:

> "But, saying like, no we're not just saying find a talkative person, we're saying, we're helping you figure out what are the requirements for this role and how does this person align to them and if they don't, what else you should consider, and interview questions you can ask them to see, can they adjust to this? Just because they don't appear that way on an assessment, doesn't mean it's not something that they're aware of as a gap in themselves and have worked really hard to overcome."

These companies, therefore, argue that people are more valuable, in certain circumstances, if they bring different qualities to the workplace. They emphasise, however, that this depends on the role and the company:

> "Every job at a corporate headquarters, every job in a functional role like, ur, finance, accounting, ur, marketing, um, supply chain, you're gunna be working cross-functionally, you're gunna be working on teams, consulting firms, you're putting teams together to solve complex problems, for all those other 50% of jobs, you actually want people to add to the culture, to not be like everybody on the team, to be completely different, so that you're getting the most different perspective on a problem or solution for your client or for the market […] so, our platform is allowing companies to do either/or, or anything in the middle." [Company Red CEO]

While Company Red were the main proponents of this idea of re-defining fit, other companies incorporated similar ideas into their design. For example, I questioned the Company Blue HOS about how they get benchmarks for certain jobs. I asked about whether they get their high performers to take the test to see their behavioural profile and then replicate it in the benchmarks, and they responded to me by saying that it's not this clear cut:

> "So, don't just show me your top performers, also show me your bottom performers. Also tell me what their tenure is, because that does matter right, like if I'm top performing cos I've been there for a very long time and you know the other two people in the top performing category have only been there six months, the two that have been there six months are very similar and the one that's been here for, you know, three years is different, you might be able to understand that this person got to where they are because of the time in the role." [Company Blue HOS]

These systems are designed so that they might be able to encourage companies to select those candidates who are not like their current workforce. In doing so, they aim to improve human perceptual biases by countering our longing for reflection, and, through this, reduce gender bias. Company Red explicitly said during the demo that this technique can be used to avoid subjectivity in hiring and that it aims to work against this idea of cloning employees:

> "[N]ot creating a system which allows people to then use that to, kind of, consistently clone themselves or… or use it in a way that they can discriminate against any other class." [Company Red Demo]

Assuming for now that the Company Red system does accurately measure the factors which it claims to measure (culture and behaviours), in theory, this could be a mechanism for improving human perceptual biases and therefore create a strong frame. It could allow clients to counter their expectations on who would 'fit' or 'reflect' their culture, and, instead, to invite people who differ from their existing workforce into consideration. In doing so, this operationalisation would counter the conceptualisation of gender bias as intrapersonal (e.g. stereotypes, homophily, subjectivity), so long as this also encouraged demographic diversity.

This does show - along with all the evidence above - that vendor companies are considering the way these human biases might materialise. They attempt to design features which address gender bias, but at the same time they are also trying to cater to a diverse range of client incentives and needs. At the IHR Recruitment Technology Virtual Conference (2021) which I attended, one speaker outlined how they know that, for their clients, it is important that the technology matches people who fit their culture, but another speaker emphasised that diverse and inclusive workforces are 21-33% more profitable and deliver 19% higher innovation revenue. These companies are clearly aware of issues surrounding diversity in hiring along with the benefits it brings, while also trying to create saleable products which match clients' desires.

However, as with the anonymisation, this design feature surrounding re-defining fit is vulnerable to the 'optionality break'; it is not mandatory for clients to use this feature and, therefore, depends on the variable incentives of employers.

Additionally, this design prompts recognition of the **fourth break in the frame** which emerges considering this optionality: **the 'profile-shaping break'**: humans play a role in shaping the skills that are desirable for their employees.

During a demo for one company, I was shown the back end of the system where, as a fake client, I could design a profile for my ideal candidate. I could do this in relation to the average for the team, or the current high performers in that role, and I could shift the dials accordingly. Either I could match it to these targets or differentiate it. I could also select to have high similarity to employees or groups, or low similarity. There is a similar design in the mock-up simulator built on the FairTargetSim project, where it attempts to exhibit to recruiters the implications on the male/female split of selected candidates when shifting the dials on a set of targets, including attention, reasoning, memory, behavioural restraint, and information processing speed (Gala & Philips-Brown, 2024).

Therefore, the feature of re-defining fit is not only optional, but it happens to the extent which the employers choose. For example, the Company Blue CSO told me about how it is the clients who set the job target:

> "[Y]ou have to do something called setting a job target which is: here's the behavioural demands of the role, or here are the cognitive demands of the role."

There is room for subjective views to be implanted into the required skills or behaviours which candidates must display, or into the profile of a 'high performer' which candidates must match. Moreover, these desired skills or job targets can seek reflection of the current workforce or specific high performers. As Raghavan et al. (2020) discuss, many AI-rec-

tech vendors offer customisable assessments which can be adapted by the clients to suit their wishes or job requirements.

In this setting, employers can choose which skills or behaviours candidates must exhibit, often in relation to top performers. This flexibility and control enable clients to do whatever aligns with their motivations when it comes to shaping the profile of desirable candidates, and so has no guarantee of reducing the tendency for homophily, or the influence of subjective opinions. The Company Red HOS told me that, of course, it does happen that clients sometimes want to clone their best performers:

> "[W]e'll look at the high and low performers and we'll be able to, um, pull out together with the client the characteristics that make that sales team, you know, what are the high quality sales team, well they have these types of qualities and the low performers have these types of qualities, so you can isolate those characteristics and then, like, try to hire for those."

In other systems, these skills might have been determined by a profile which is customised in-house by the vendors, usually by an IO psychologist, and might have used data across one or several clients. Data provided by clients of their high performers is also likely to contain human subjective judgements around high performance. Alex Vaughan, Former CSO of Pymetrics told me that using client data to measure high performance could maintain the human standards within the company, because "that training data is not random"; it comes from historical candidates who have been hired or highly ranked, and while this could reflect "historical tendencies to hire and retain good people" it could also "retain historical biases in the hiring process".

The Company Red CEO told me how this might play out in terms of subjective judgement surrounding selection of high performers based on homophily:

> "[L]et's say there's a VP of sales, and there's 400 sales… there's 300 salespeople. And we use the tool to pick out the top performers in the first top 100. Well, if that head of VP of sales somehow is identifying, that, they're somebody who's identifying high performers. […] you could always have gender bias in the identification of who's a high performer. Even if that top 100 are the best performers, maybe it's because the Head of Sales gives them the best leads, cos he's some buddy who plays golf every weekend and the women don't play golf. I'm doing extreme example. So, women are always at the middle or the bottom. So, at the end of the day then the tool isolates criteria that are you know, for high performers, but it might actually be correlated with men because the person who's identifying high performers are giving the best leads to the high performers is a man, he's looking to help other men. That has happened, it could happen, it's an extreme example, but there is, with any system, any tool, there always is a chance for bias."

In all these cases, whether it is the client designing their desired profiles or the vendors using client(s) data or subject expertise, humans are still used to determine which competencies are most relevant to the job (Raghavan et al., 2020). Tambe et al. (2019) discuss how this raises questions surrounding why certain skills are chosen, including what these skills might reflect, who they might favour, who is being used to identify the high performers that these candidates are measured against, and how this might perpetuate existing gendered trends around job fit and performance.

Some vendors say that they encourage their clients not to perpetuate existing trends, but it is also acknowledged that there is no way of telling whether they do this or not:

> "We can't tell them what to do, they can ultimately do whatever they want, but we, we express as strongly… it as strongly as we possibly can." [Company Blue CSO]

The Company Red CTO also recognised that these systems could be weaponised depending on the clients' ideas of what they want from the systems:

> "[T]hey [the client] say look this is proven success of a salesperson is someone who looks like this so I'm just hiring people like that, well… you don't know that you wouldn't get someone better if you tried something different, but you're not trying something different."

Indeed, while it is impossible to know from my own data how clients might be using these systems, it is possible to see how vendors are marketing, framing, and understanding this design in relation to their clients. The optionality is a data point itself to exhibit that vendors understand that clients want to use these systems in a variety of ways and as seen in Chapter 2, there has been much work on the motivations of clients to use these systems for efficiency and finding productive employees.

Many vendors certainly market one of the big advantages of their systems as the ability it has to replicate high performers in hiring decisions. In one of the company's reports, it said that because the "client can now easily identify those candidates that have the traits of a high performer without subjective judgement", clients can now reduce turnover by 40%, save time, effort, and around $10m in combined cost savings and production efficiencies. Many of the other websites also have statistics on how to 'find your perfect fit' to increase retention and costs saved and decrease time to hire (Pymetrics, 2022e). Another report advocates that "[b]usiness success isn't up to chance. It's about having the right people in the right roles, for the right business needs". Indeed, there is much emphasis from vendors on how their tools optimise for organisational 'fit' which is often based on replicating top performers (Dencik & Stevens, 2021; Sanchez-Monedero et al., 2020). This is often presented as an advantage for employers as it can help to find employees who match their mission (Allal-Chérif et al., 2021; Hmoud & Varallyai, 2019). Additionally, recent research into 12 AI-rec-tech vendor and client companies in Italy has found that clients are motivated by talent capture, increased productivity, and saving time and money, and vendors are therefore incentivised to provide this to their clients (Sartori & Collett, in revision). While these findings are contextual and not representative of all

employers, it does demonstrate that these systems might be used in multiple ways to support a client's business values or needs.

Moreover, the vendors in this study also consistently communicated to me that their clients usually purchase these systems to drive revenue and productivity:

> Collett: "So, what do you tend to see the trends being in terms of the motivations for why clients want to use the recruitment tech?"
> Company Red CSM: "The reason that they're motivated to do…to use it…is um…for money"
> [Pause]
> Collett: "[Laughs]. And that's it?"
> Company Red CSM: "That's it. Ok?"

Therefore, they argue, if there is a perceived decision between reducing bias and increasing revenue, the clients tend to go with the latter:

> "[When] business leaders are making that comparison, they're always going to, for the most part, opt for the revenue side […] So, unless there is a leader in the organisation who says no, this is also important, it's very difficult to move something forward." [Head of Impact at Pymetrics]

One of the participants who worked in Sales, and therefore regularly interacted with potential clients, told me that:

> "I have to say that um you know we… that we don't… there's not a lot of talk specifically about gender bias […] I don't have a lot of deep discussions with my clients about gender bias […] my sales pitch is, is, um you know you can reduce subconscious bias whether that's ethnicity or gender, and they all kind of nod..." [Company Red HOS]

In addition, when I asked another Sales leader about where bias featured on the list of priorities as to why clients purchase the technologies, they told me:

> "I don't often hear it if I'm being honest. I don't hear people saying I'm coming to [Company Blue] because I wanna be more thoughtful in who I bring on and I wanna promote diversity. It's sad but it's true, it's not how they're thinking." [Company Blue HOS]

Therefore, it is certainly plausible that clients will not be motivated to use the re-defining fit feature and will instead opt to design their own profile for ideal candidates or use data to replicate their high performers. This optionality and profile design flexibility exhibit 'breaks in the frame', where human perceptual biases can influence decisions around candidate suitability.

Having discussed how these breaks allow room for human biases to infiltrate, it is also important to note that there is no evidence to suggest that re-defining fit does change demographic diversity. While it might be a counteraction of homophily and stereotypes if it fulfils its purpose, it is difficult to say this creates a strong frame in design without seeing how it impacts demographic diversity.

Some vendors did claim that their tests make a difference to numbers around demographics:

> "[F]rom what we see at Pymetrics every single one of our clients sees improvement in diversity." [Frida Polli, CEO of Pymetrics]

However, they did not provide me with any further information about how they know this, and most vendors seemed unaware of how their systems impact demographic diversity of company hires. Therefore, there is no evidence to know how setting the system to 'clone' vs. 'diversify' impacts demographic make-up of hires:

> "Yeah, we have not focused, and in some way's purposefully, on the demographic side. Um, does it… result in it [demographic diversity]? Um… I don't have the data to really answer that […] I'd have a guess it would result in some diversity of, you know, gender or race and demographics in general, um I just don't have the data to say that it does." [Company Red CTO]

> "We don't get to see what decisions they make […] we don't have, we're not, like, we don't track that. We'd love to get to a point where we could say like oh you've decided to hire Clemi, great, like we now know that person is hired, we can track

> what happens from there, but people don't come into our system and input data that way." [Company Blue HOS]

This speaks more to the framing of gender bias as 'statistical', which I discuss in Chapter 6, and is also an area which needs to be further researched. But this is important to note because it emphasises that, whilst the vendor companies claim to be motivated by concerns for gender bias, the optionality built into their design allows clients to bypass mechanisms for reducing human perceptual biases altogether and, moreover, they aren't sure how their clients use of these systems are impacting diversity.

Overall, vendors do attempt in the design to improve – or in other words, reduce - elements of human perceptual biases through bias reminders and re-defining fit. However, there are 'breaks in the frame' - inconsistencies with their understanding of gender bias which emerge in relation to these features. Firstly, there is the 'ignorable break', where bias reminders or alerts can be ignored by employers, and secondly, there is the 'profile-shaping break', where humans continue to model the desired skills and behaviours of candidates. Both breaks show that there continues to be room in this design for human perceptual biases to persist at various points of up and down stream of system design.

4.3.3 A note on the presentation of gender

On another note, these operationalisations of improving human perception surrounding biases rooted in subjectivities, stereotypes, and homophily, are somewhat counteracted by the stereotypical presentations of gender within the assessments themselves. Here, I outline the **fifth break in the frame - the 'contradictory break'**: despite the vendors'

conceptualisation of gender bias in relation to stereotypes, there continue to be stereotypical presentations of gender in some of the systems.

When asking the vendors about the options for gender within the test, or when filling out the test myself, I found that while some of the tests had the options 'female', 'male', 'non-binary', 'prefer not to say', or in some cases 'female', 'male', 'I don't identify with the above', a surprising number of companies only had 'male' and 'female' options for gender. Sometimes, companies would tell me this was because there was not enough data on non-binary people to have this as an option, or because by adding options they are automatically excluding people. While many assured me that they'd like to go further with gender options eventually, it was clear from the interviews and autoethnography that many of these tests are still designed mostly based on the gender binary, therefore "mapping static categories onto diverse bodies" (West et al., 2019, p. 17). This risks categorising gender along the biologically essentialist and binary lines of male and female (Sinead O'Connor & Liu, 2023; Scheuerman et al., 2019), excluding people (Keyes, 2018), and baking this gender binary in to the system design and the data it collects in order to continually train the model (Bivens, 2017; Bivens & Haimson, 2016).

Kevin Guyan's book *Queer Data* provides an overview of the tensions of gathering data which is queer, which aims to disrupt the binaries of male/female, heterosexual/homosexual, and cis/trans: "decisions made about who to count, what to count, and how to count are not value-neutral but bring to life a particular vision of the social world" (Guyan, 2022, p. 1). While queer theory disavows categories, there is a tension between being counted and being beyond counting. But simultaneously, it is

important for people to see themselves reflected in the data, even if it's not always a truly accurate representation (Guyan, 2022).

On top of this, I found it striking that so many of the tests presented stereotypical versions of gender throughout their assessment in visual form. In my demo journal I wrote: "I am noticing there are some stereotypical traits in the appearance of gender in these simulations". As noted in the literature review, gender stereotypes were one of the main factors noted in relation to bias (Fabris et al., 2023; Nadeem et al., 2022), and as seen above, this is something which the vendors understand to be an integral part of human perceptual bias.

These gender stereotypes, however, also emerge through stereotypical imagery (Sinead O'Connor & Liu, 2023). During one demo, I was shown how the pre-assessment chat would look to a candidate. It presented me with a chatbot with a female name. She much resembled a doll: she had straight, dark hair, perfectly straight white teeth, rosy cheeks, shaped eyebrows, and she wore dark make-up. Another demo used a similar avatar in a simulation where I was tested for my skills on the job during a free assessment. In figure 6, I give an example of a similar image which, while not taken from the demos/tests themselves (due to anonymity and copyright reasons), is similar to the images I encountered during the assessments.

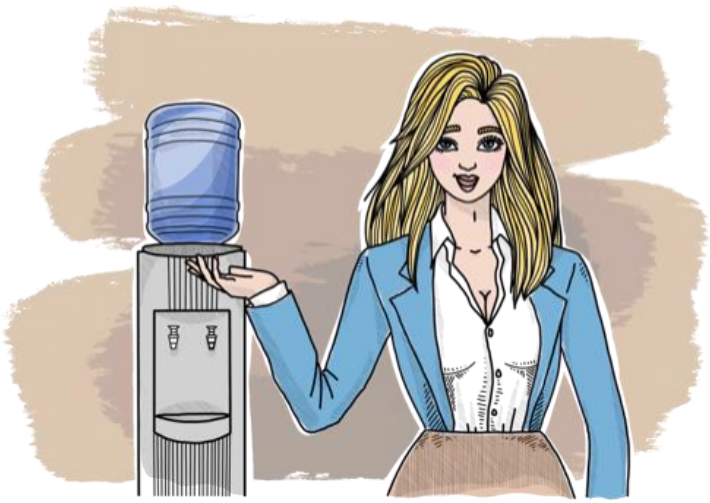


Figure 5: Replication of demo avatar, credit: Joanna Rivera-Carlisle

During my autoethnographic work, with one test a chatbot asked me questions. She had brown hair, a ponytail, and perfect teeth and she really took on the role of the supporter, cheerleader, encourager. Every time I got a question wrong, she would ask me things such as ‘why not try again?’ followed by a winking face emoji or ‘let’s take another shot’ followed by a devil face emoji.

This made me question whether certain aspects of these systems improve human perceptual biases when, often, the chatbots and ‘helpers’ within these simulations were gendered as female. Many of these chatbots presented a stereotypical and conventional feminine image of thinness (Faragó et al., 2022), emotional expressiveness (R. Adams et al., 2015), straight and contained hair (Donahoo, 2023; Toerien & Wilkinson, 2003; Weitz, 2001), and even elements of flirtatiousness. This also raises questions about consistently personifying the female chatbot as the ‘helper’, something which has been broadly discussed in the literature about virtual personal assistants (VPAs) such as Siri, Alexa, and Cortana, all of which are gendered as female, which risks promoting ideas of the female as subservient (Bergen, 2016; Collett & Dillon, 2019; UNESCO, 2019;

Woods, 2018) and “transfer to the digital realm the gendering and stratification of labour found in the real world, with working women principally confined to jobs of lower power, status and pay” (Dillon, 2020, p. 10).

While stereotypes are questioned by the informants as a constituent part of biases in human perception, and this is a crucial part of framing gender bias as intrapersonal, we also see that these stereotypes are arguably embedded into the technology for their clients and candidates alike to see. As noted, then, this aspect of the design – while not intended to address human perceptual biases – in some ways presents and reinforces these stereotypes. In this sense, I label this as the ‘contradictory break’, as these binary gender options and stereotypical gendered images encourage exactly the tendencies vendors are expressly trying to reduce.

4.4 Discussion

Over the last decade, AI has increasingly been used by HR practitioners to improve efficiency and productivity (Nocker & Sena, 2019). These systems are often seen to save time and cost, improve accuracy, remove bias, reduce workload, and produce a better candidate experience (Hemalatha et al., 2021). Indeed, in recent years, some have argued that AI-rec-tech is able to reduce or remove human biases (Black & van Esch, 2020; Upadhyay & Khandelwal, 2018). However, there has also been much pushback, as some advocate that these systems are really a mechanism for perpetuating human biases (Barocas & Selbst, 2016; Köchling & Wehner, 2020; Raub, 2018).

In this chapter, I have spoken to this debate by providing the first of three answers to my research question: how do AI-rec-tech vendors frame gender bias in hiring?

I have explained and analysed the 'intrapersonal frame' which interprets tendencies of the human mind to be a source of gender bias in hiring.

Firstly, I outlined the vendors' conceptualisation of gender bias in hiring within this frame (sub-RQ#1). Vendors often conceptualised gender bias as intrapersonal – linked to the human mind. This was evident in their discussion of three tendencies of human perception which they linked to gender bias: pattern recognition (stereotypes), homophily, and subjectivity. I showed that the first two of these tendencies (stereotypes and homophily) resonate with the two main strands of literature on gender bias in hiring and the workplace, as demonstrated in Chapter 2. However, research in this area focusing on subjectivity is less common. I argued that this showed how vendors place new emphasis on subjectivity as a source of gender bias in hiring, and I hypothesise that this might be a way of highlighting that their systems, in contrast, are 'objective', a quality which is often perceived to be beyond the ability of humans. Indeed, these three tendencies all seem to be designed to place their systems in a preferential position to the alternative of human-based hiring when it comes to bias.

While I am not claiming that literature and theory on these topics has caused vendors to develop this frame (although this is not impossible, considering that many of them have a background in academic studies), previous research supports, reinforces, and gives weight to their conceptualisation of gender bias in the intrapersonal frame. It illustrates the importance of considering human perceptual biases in the context of hiring decisions.

Clearly, this frame resounds with a broader understanding of bias and is vital for us to consider when it comes to recruitment practices. This can be seen in the discussion of topics such as stereotypes and hiring in our image in the mainstream media (Carnegie, 2023; Grant, 2018; S. Iqbal & Kar, 2021; Kramer, 2022). This is likely to be another source for the opinions and knowledge of the vendors. Indeed, I found that there was engagement with popularised academic works such as *Thinking Fast and Slow* by Daniel Kahneman in my interview with Frida Polli, CEO of Pymetrics.

Secondly, I explained and analysed the vendors' operationalisation of this understanding and their attempt to address gender bias rooted in human perceptual biases within their system design (sub-RQ#2). I argue they attempted to 'design out' this form of bias through 'removing' or 'improving' human perceptual tendencies. They aim to do this through design features such as anonymisation of candidates from recruiters, bias reminders, and re-defining fit. These operationalisations demonstrate that vendors do attempt to address this bias and do consider how human perceptual biases might alter hiring processes and decisions. Indeed, the operationalisation of re-defining fit – if able to counter expectations and desires of recruiters - could certainly help to reduce human perceptual biases such as homophily or stereotypes. However, there is no indication of how much employers use this feature or what impact it has on demographic diversity of accepted candidates, something which is crucial to addressing bias as outlined in Chapter 1.

While the conceptualisation of this frame was strong, therefore, there were vulnerabilities when it came to the operationalisation of this frame.

I showed how these vulnerabilities took the form of 'breaks in the frame', which I defined as inconsistencies between the conceptualisation of gender bias and the ability of the operationalisation to address this conceptualisation. These breaks therefore represent places where human perceptual biases (and therefore stereotypes, homophily, and/or subjectivity) have the potential to influence rankings or decisions surrounding candidates, and therefore impact the design being able to function to remove or reduce these biases with any certainty.

While exploring the design features which attempted to 'remove' or 'improve' human perceptual biases, I argued that there were five 'breaks in the frame', the common feature of these breaks being that they leave room for human perception to infiltrate hiring decisions:

- **Break 1: the 'optionality break' –** when discussing the anonymisation feature, I outlined how the vulnerability in this operationalisation lay in the fact that this was optional. Therefore, human perceptual biases might still infiltrate if clients decide to disable the feature. Just as easily as tendencies of perception might be sequestered, they might be re-introduced.
- **Break 2: the 'hybrid break' –** the anonymisation feature and the optionality break prompt recognition of how vendors often communicate and encourage the 'hybrid' nature of the technology. This perceives the system as a support for human decision-making, leaving much room for human interpretation and judgement to occur.
- **Break 3: the 'ignorable break' –** in discussion of bias reminders, I highlighted how these reminders and alerts are ignorable, and therefore leave space for human perceptual biases to occur.

- **Break 4: the 'profile-shaping break' –** when outlining the design of re-defining fit, I explained the profile-shaping break, which illustrates how the desired skills and behaviours are often decided by humans. This might occur in customisable tests which gives clients the control, or custom tests designed by the vendors for specific jobs using client's data or internal expertise.
- **Break 5: the 'contradictory break' –** gender stereotypes and binary conceptualisations of gender which are sometimes interwoven into system design sit in tension with the vendors understanding of gender bias as linked to stereotypes.

Through outlining these 'breaks in the frame', I illustrated that optionality and flexibility within these systems is likely to appeal to their clients, who might want to use systems in a variety of ways. I also noted that design can be subject to more than one break in the frame. For example, the re-defining fit operationalisation is vulnerable to optionality and to profile-shaping. This illustrates that there are multiple ways in which human perceptual biases might still influence rankings and decisions when AI-rec-tech systems are used.

This assertion about 'breaks in the frame' poses a question: is it possible for this frame to be intact, and therefore to have the removal, or even a definite reduction, of human perceptual biases?

It is inevitable that human perceptual biases will infiltrate the use of AI systems within hiring due to the necessity of human choices up and down stream. While some of the breaks above might be altered, for example, reducing stereotypical images in chatbots and simulations, there will continue to be elements of human perceptual biases in the

construction of desirable skills, the labelling and selection of data, and the final shortlisting and decision processes. Therefore, given that some breaks can be avoided, and some can't, this frame is inevitably broken; we can never have a situation where there is no human involved in at least some part of this process. As Ajunwa argues, to argue for or against automated decision-making versus human decision-making is a false binary – the two can never be wholly disentangled (Ajunwa, 2023a).

This thesis is not here to judge whether humans should be removed entirely from the hiring process or not, but rather to assess and analyse the way(s) that vendors are framing gender bias, and to measure their understanding and claims against the complexities of how the systems are designed to work. This chapter has demonstrated that the relationship between humans (and their biases) and AI-rec-tech systems is complex, and while there might be the potential for improving the influence of some human perceptual biases through using these systems, there is no guarantee and no evidence either way, largely because it is still opaque how these systems are used by clients. However, it is clear from this design that there are many ways in which human judgement can be, and still is, used to shape the processes and outcomes of these technologies.

The inevitability of the influence of human perceptual biases, even when using these systems, was recognised by some, although not all, of the vendors. As we saw in this chapter, certain vendors did acknowledge that while their systems might be able to reduce human biases, there is no guarantee. However, some of the vendors did claim that their systems can remove human biases, and for these vendors especially, these 'breaks in the frame' are crucial to acknowledge.

Discussing ‘breaks in the frame’ is valuable for two main reasons.

Firstly, it highlights inconsistencies regarding what the vendors assert about these technologies and how they tell me their design works. This prompts us to draw boundaries around what claims can be made regarding this technology. While not all vendors did claim that their systems ‘remove’ human bias completely, all of them did claim that to some level, they do ‘reduce’ human bias. The ‘breaks in the frame’ demonstrates that this claim depends on design and usage. Therefore, it highlights the inability for vendor companies to claim on certain terms that these systems remove, or even reduce, human bias in hiring decisions. Their systems can only claim to provide design features which might help to mitigate elements of human perceptual biases in hiring decisions. They should note that there are ways human perceptual biases can still infiltrate hiring standards and decisions.

Secondly, the breaks help to signify points where either action or more research is needed.

In terms of action, as noted above, vendors should stop claiming that their systems can remove – or even reduce with certainty – human biases. In addition, there are certain elements of design which can be changed with relative ease. For example, the inclusion of non-binary gender options in pre-assessment forms, and the re-conceptualisation of the stereotypical images of women and chatbots within the systems.

In terms of future research, there should be investigation into the extent to which these breaks occur when these systems are used by clients. It is likely that clients are using these systems in a multitude of different ways, as “[s]ome might take hiring ADSs as

suggestions for their own decision-making, others might implement a hiring ADS's outputs directly" (Sloane et al., 2022, p. 7).

In addition, research should endeavour to understand the impact of these design features (anonymisation, re-defining fit, bias reminders) on demographic diversity in practice. Ideally, this could also be done in a comparative setting to see how and in what way different approaches at these stages might 'reduce' human perceptual biases, even if it cannot remove them fully. As Raghavan et al. (2020) point out, while it is easy to criticise vendors for their choices, it's not clear if there are better alternatives, and this is something which needs to be explored, especially in large-scale hiring scenarios. Indeed, recent research has shown that the use of automated systems lowered diversity compared to traditional hiring with humans. However, this was generally due to the human-based elements of the system, the 'meta-algorithmic judgement', which gave the recruiting manager discretionary power. Usually, the humans ignored the algorithms recommendations, and therefore there were unintended human-algorithmic interaction effects. This relation between humans and AI-rec-tech systems will therefore be crucial to explore more widely (Bursell & Roumbanis, 2024).

Overall, regarding the social shaping of technology (MacKenzie & Wajcman, 1999), this frame has been shaped by prevalent beliefs and realities in society. Indeed, it has attempted to counteract the social reality of human perceptual biases (stereotypes, homophily, subjectivity) which – as we see – there is evidence to support its occurrence in the workplace. However, as demonstrated, the AI-rec-tech systems also continue to reflect stereotypical presentations of gender through avatars and chatbots, and limited gender options in forms.

On the other side of this bi-directional relationship, it is less clear how this frame in design might work to shape society. The 'breaks in the frame' show there is no guarantee that the design features in this chapter will change human perceptual biases in society. If anything, the presentation of gender in these systems continue to perpetuate gender stereotypes of women as helpful, ever positive, and flirtatious. While re-defining fit and anonymity might enable a subversion of these stereotypes in the context of hiring in our society, there is no guarantee.

4.5 Conclusion

In this chapter, I presented the first of three synchronised answers to my overarching research question: how do AI-rec-tech vendors frame gender bias in hiring?

I answered this by outlining the 'intrapersonal frame'. My analysis in this chapter has shown that vendors conceptualise gender bias to be intrapersonal – interlinked with human perception. However, I have argued that there are 'breaks in the frame': inconsistencies between the conceptualisation of gender bias and the ability to fully address this bias in the technological design. Therefore, while there are ways in which these systems might exclude human perceptual biases, there are multiple avenues through which human perceptions are able to creep back into the frame.

In the next chapter, I discuss in more detail what exactly these systems claim to measure and will scrutinise how this might relate to factors 'out-of-frame'. Next, the 'structural frame'.

CHAPTER 5: THE STRUCTURAL FRAME

5.1 Introduction: framing biases as structural

After my interview with the co-founder and CEO of Code Signal, Tigran Sloyan, I sat in silence for a while. This wasn't unusual; I tried to do this after each interview. Often, it was helpful to take some time afterwards. It gave me a moment to write in my interview journal and mull things over. But there was something different about this interview. As I sat there, I couldn't stop thinking about something Tigran had said to me early in the conversation.

It had started off as a story.

When Tigran had been young, he'd been a mathlete. He told me, "My parents had this theory that if you get really good at something, you'll find your place no matter what, even if, you know, all things are stacked up against you" [Tigran Sloyan, CEO and co-founder of Code Signal]. He continued, telling me he had known someone on the circuit for the computer programming equivalent of these competitions. A young boy from a small village in a poor country. The boy had such raw skill and talent from an early age, he was a legend, and he won everything. While they knew of each other, they didn't run in the same circles, and so they'd never met. Tigran told me the two them were like the Michael Jordan and the Lionel Messi of their world. But a few years down the line, while Tigran went on to attend a top university and was scouted to work at a big tech company, the boy he'd known of went on to an obscure university, and after he graduated, was unable to get a job. But this was impossible? How had this happened? This boy was the

best engineer Tigran had ever met. Bringing the story to a close, he told me: “I was like yeah, we gotta fix this. And this was kind of where the initial idea for the company was born of, like, can we create a system where you can directly measure ability” [Tigran Sloyan, CEO and co-founder of Code Signal]. This boy from the story is a co-founder of Code Signal.

This anecdote hints at the second way in which these companies framed gender bias in hiring: as structural injustice derived from inequalities in education and opportunity. Some AI-rec-tech vendors aimed for their system to address this through measuring skill directly, instead of qualifications. This, they propose, gives people a chance who might otherwise be overlooked due to their position within the socio-economic structure. People just like the boy in the story: an exceptional talent who exhibited skills without formal education or ‘prestigious’ credentials. In this sense, the vendors merged ideas of injustice with those of bias. Biases, to them, emerged from these structural injustices.

This interpretation of gender bias was often spoken about in relation to race and class, as well as gender. As I noted in Chapter 1, I treat the term gender intersectionally – gender intersects with other protected characteristics such as race, ethnicity, class, and so on, all of which are also subject to power dynamics. Therefore, conversations about other protected characteristics are also relevant to gender – they all signify patterns surrounding unequal power dynamics (Crenshaw, 1989; Hill Collins, 1990). Examples throughout this chapter therefore refer to multiple protected characteristics, with a focus on gender.

AI systems often interact with social structures, and are frequently seen as reproducing inequalities in these structures: “[w]hen applied in social contexts such as policing, the

court system, health care, and education, they [AI systems] can reproduce, optimize, and amplify existing structural inequalities" (Crawford, 2021, p. 211, also see Benjamin, 2019). For example, when used by courts and parole boards to forecast future criminal behaviours, an AI system was found to be inaccurately identifying black defendants as future criminals more often than their white counterparts (Angwin & Larson, 2016). These trends in AI outcomes might be referred to as the 'automation of inequality' (Zajko, 2022) or 'algorithmic oppression' (Noble, 2018). Often, this is because AI systems are trained on data which reflects our societal biases (Vlasceanu & Amodio, 2022) such as structurally embedded racism and sexism (Noble, 2018).

When it comes to AI-rec-tech specifically, these systems have also been theorised as reproducing and amplifying inequalities (Ajunwa & Greene, 2019). The exclusion of training data on some vulnerable groups in these systems, for example people with disabilities, has been seen to reproduce and even worsen outcome inequalities for these groups in hiring contexts (Tilmes, 2022). Certainly, it is extremely important to consider the reproduction of structural inequalities in the context of AI-rec-tech. Employment affects people's livelihoods and opportunities, and unequal access to work frustrates efforts to reduce poverty (UNRISD, 2010). This interaction of AI with structural inequality therefore deserves our attention (Gehl Sampath, 2021).

Indeed, some of the vendor companies consider these issues. While structural injustice was not spoken or written about by all the companies, several vendors acknowledged issues of inequality in our societal structures. For some vendors, it was their core mission to reduce biases tied to structural injustices. As this chapter demonstrates, some of these vendors didn't just acknowledge structural injustice, but attempted to account for these

injustices in the design of their systems. This technology was their bid to act against these socio-economic structures rather than reproduce and reinforce them.

In this chapter, I use interviews with vendors, analysis of websites and reports, along with autoethnography and walkthrough data to provide the second of three concurrent answers to my overarching research question: how do AI-rec-tech vendors frame gender bias in hiring?

Here, I outline and analyse the second of these frames: the structural frame. This follows naturally from the previous chapter, as it indicates the accumulation of intrapersonal biases; these social structures are the manifestation of a history of human perceptual biases mobilised in human action. In turn, these structures often continue to perpetuate human perceptual biases. However, this chapter is not about how these biases are linked to individuals as it was in the previous chapter, but rather, how "bias […] is embedded in obdurate social institutions" (Joyce et al., 2021, p. 6).

Overall, in this chapter, I show that several vendors conceptualised gender bias in relation to structural injustice, particularly within education. I demonstrate how they attempt to address this in system design, for example through making questions understandable for a broad audience (question neutralisation) and through skill/behaviour measurement instead of reliance on credentials. However, I argue that this leaves unanswered questions, particularly surrounding two areas. First, are skills and behaviours really such a 'neutral' measurement, or are these factors also impacted by education and socialisation? Second, how are these skills and behaviours being conceptualised and measured in the first place?

Before beginning, it is important to note that, due to the context of these interviews and the geographical placement of the vendors, this chapter is mainly relevant to the Global North and particularly the US and UK, although at times I do discuss the broader global context. While many of these issues might apply to countries in the Global South, the structural barriers and issues which would arise are likely to be different in many ways.

5.1.1 Chapter outline

The chapter proceeds as follows:

First, I answer sub-RQ#1 for this frame. I explain that this frame conceptualises gender bias through harnessing narratives on inequality in education and opportunity. As I show, this embodies a definition of structural injustice which chimes with that of Iris Marion Young (I. M. Young, 2011). I show that this is an important and strong conceptualisation of gender bias which must be considered within the hiring context. In addition, I argue that the vendors' understanding of how their systems relate to structural injustices exhibit parallels with the ideology of 'the American dream'; they see their systems as enabling people to break free of inequality due to hard work, ambition, and innate skill.

Second, I answer sub-RQ#2 for this frame. I demonstrate that vendors operationalise addressing this type of bias through attempting to account for structural injustice in design. They do this in two ways in particular: neutralising questions (making them accessible and understandable for all groups) and assessing soft skills, behaviours, and hard skills (rather than qualifications). This is seen to equalise the playing field for those

who have skills which are right for the job but can't demonstrate that through educational credentials.

Third, I make the case that although the vendors' conceptualisation of bias in relation to structural injustice is a strong one, and crucial to consider in the hiring context, there are two vulnerabilities in this framing of gender bias regarding the design. I use Goffman's idea on 'out-of-frame' activity (Goffman, 1974) to outline these. 'Out-of-frame' activities refer to factors outside of the vendors' main 'track' or 'channel', and therefore outside of their control or outside of our discussion. However, these external or unknown factors impact the function of the design, and therefore have the potential to disrupt the integrity or function of the frame.

The first 'out-of-frame' vulnerability I label as 'out-of-frame permeability': realities out-of-frame, or out of focus, which might permeate and therefore impact the frame in design. I argue that it is possible that structural injustices continue to impact soft skills and behaviours in ways unforeseen to the vendors. I also point out that there is still a certain level of hard skills needed for certain professions which is likely tied to education and opportunity. Therefore, while outside of the vendors' control (and therefore out-of-frame), it is likely that education and opportunity still 'permeate' this frame. I argue that while this design might therefore account for a minority who develop certain skills in the face of lacking education and opportunity, their design is still vulnerable to structural injustices.

The second 'out-of-frame' vulnerability I label as 'invisible out-of-frame': crucial factors of the design which are invisible to me as the researcher, but which impact the function

of the system. I argue that while the factors measured by vendors are seen to be neutral, we as researchers cannot be certain what metrics they are using to choose and measure skills, behaviours, and soft skills; these crucial decisions are therefore 'invisible out-of-frame'.

Fourthly, in the discussion section, I return to my main research question and summarise the structural frame. While recognising the awareness and efforts of these vendor companies to account for structural inequality in their system design, I argue that this design is vulnerable to the two types of 'out-of-frame' activity outlined in the previous section (Goffman, 1974). I relate these vulnerabilities back to the ideology of the American dream. This design does not – and could not be expected to - solve for the root issues of structural injustices. It might enable certain candidates who have developed skills in the face of inequality to have a better chance of getting a job, and the neutralisation of questions in theory should make these tests more accessible to multiple groups. However, structural injustices will likely continue to impact candidates, regardless of the measures to account for these injustices integrated into the design. I discuss why this frame is valuable to study and how the questions raised by the vulnerabilities indicate areas for future research, particularly around the 'underlying science'.

I conclude in the fifth section by summarising the structural frame and the chapter overall.

5.2 Structural injustice and the American dream

In 1931, the historian James Truslow Adams' book, *The Epic of America,* was published. Adams wrote about American dream, a "dream of a land in which life should be better and richer and fuller for every man, with opportunity for each according to his ability or achievement". This meant that "each man and each woman shall be able to attain to the fullest stature of which they are innately capable, and be recognized by others for what they are, regardless of the fortuitous circumstances of birth or position" (J. T. Adams, 1931, p. 404). In America, people would be able to move forward, "unhampered by the barriers which had slowly been erected" and to be "unrepressed by social orders which had developed for the benefit of classes rather than for the simple human being of any and every class" (J. T. Adams, 1931, p. 405).

The ideology of the American dream, which emerged during the Great Depression, acted as a kernel of hope. It provided hope that, even in the face of inequality and instability, people might have the opportunity to achieve great things if they were capable and worked hard. It is my interpretation that, in a way, the hope of the American dream in the face of inequality is encapsulated in the structural framing of gender bias evident in some of the vendor companies.

In this section, I discuss how some vendors conceptualised gender bias as linked to structural injustice. In turn, I interpret that they saw their systems as able to facilitate equal opportunity for success, regardless of background. They recognised their assessments as allowing people to achieve their dreams if they worked hard, harboured ambition, and exhibited innate skill. In my eyes, these companies – in their own way –

saw their systems as facilitators of the American dream. Therefore, in this section I answer sub-RQ#1: how do AI-rec-tech vendors conceptualise gender bias in hiring?[15]

First, though, I must define structural injustice. I do this in relation to the understandings of the vendors, which I found to mainly hinge on structural injustices resulting from inequalities in education and opportunity for social mobility, which directly have an impact on peoples employment, their income, and their standard of living.[16] Due to this focus, I draw on Young's theory of structural injustice (I. M. Young, 2006, 2011) to help define the concept.

Young argues that "[s]tructural injustice, then, exists when social processes put large groups of persons under systematic threat of domination or deprivation of the means to develop and exercise their capacities, at the same time that these processes enable others to dominate or to have a wide range of opportunities for developing and exercising capacities available to them" (I. M. Young, 2011, p. 52). Crucially, Young highlights how social and economic structures including the housing system and the service economy - an umbrella term which includes education, along with financial services and health - encompass resource differentials, and therefore power differentials, between groups. These differentials create inequality in opportunity for some people to develop resources and capacities, and these inequalities, to Young, are structural injustices (Jenkins, 2021).[17]

[15] Unlike the previous chapter, this conceptualisation was often spoken about in relation to how the vendors might address the issue. Therefore, the data exhibiting this understanding of gender bias continues to unfold throughout the chapter.

[16] This definition enables me to include material where 'structural injustice' is explicitly and implicitly mentioned.

[17] I do not attempt to go into the deeper philosophical roots of how responsibility is linked to injustice; I simply use Young's work to define the concept of structural injustice broadly speaking.

Regarding the structure of education, the main focus in this chapter, Young points out that in most of the world the amount and quality of education a child receives is largely determined by their parent's wealth. This builds a power and resource differential into the structure. She gives the example of the US, where ballots or legislative bills calling for increased public funding are constantly defeated, maintaining inequality for some people to access better education than others. This inequality is not just about wealth, but gender, too. Girls are socialised primarily for domestic duties in the home, and perhaps, when girls do go to school, they may experience prejudice regarding subject selection and capability (I. M. Young, 2006).

From Young, I adopt the basis for a definition of structural injustice, which I define as: the more limited opportunities for upward social mobility[18] resulting from inequalities in power and resources within socio-economic structures.[19]

Some vendors saw structural injustice as intrinsically linked to gender bias. One report from Pymetrics made this particularly explicit:

> "Structural injustices and unbalanced representation of the population can be reflected in your dataset, causing discrimination against certain protected classes […] oftentimes seemingly unrelated data points can serve as proxies for determining other factors such as gender or race. For example, the quality of an applicants' high school may seem like an innocuous data point, but school districts are often tightly interwoven with historic patterns of bias." (Pymetrics, n.d.-b)

[18] The OECD defines social mobility as a change in a person's socio-economic situation, either in relation to their parents (inter-generational mobility) or throughout their lifetime (intra-generational mobility). It can be measured in terms of earnings, income, or social class, but can be understood to encompass other well-being dimensions such as health and education (OECD, 2024).

[19] Throughout the chapter, I refer to inequalities in reference to power or resource differentials, and injustices in reference to this definition of structural injustice as a whole.

This quote summarises how structural injustices in the real world were seen as possibly being captured within the data (even via proxy) and therefore, as the vendors note, these biases (structural injustices) might be reproduced using the system.

For participants in the US, these structural injustices were largely linked to education. The Head of Impact at Pymetrics told me that, "one of the biggest challenges we have in the US which becomes, you know, a challenge for us as a country, frankly, is - we have not done a good job around education". They referred to the way in which US public schools are funded locally, rather than by child. This means that districts with expensive houses and correspondingly high rates of taxation have higher amounts of funding than poorer districts (Hochschild & Scovronick, 2003). The Head of Impact at Pymetrics argued that this inequality in education is reinforcing wealth inequalities. If all those in a particular zip code are poor, they won't have the same opportunities as those children in wealthier zip codes:

> "We have an education issue, and we have an opportunity issue, right? And so, if the talent's not getting exposure and educated, then they're not even being considered." [Head of Impact at Pymetrics]

Here, we see the link between education and opportunity for upward social mobility – education is seen as the key to securing a job, which itself enables people to alter their social status and wealth.

Within education, some participants discussed the inequalities between genders. Sophia Baik, COO and co-founder of Code Signal, spoke about structural injustices exhibited by inequalities between men and women's education, both regarding subject take-up and the type of learning they receive. In terms of who takes STEM subjects, they told me that, "the men start programming much younger, um, than women in general". They continued

by explaining that even if women get a computer science degree, the knowledge they gain could be different. For example, they might have varying practices outside of the curriculum which might help or hinder them. This inequality in STEM is certainly something that is reflected in the global data. Overall, women are underrepresented in STEM education, currently holding 36% of STEM degrees and making up only 25% of the STEM workforce globally (Gallego et al., 2019).

Just about as soon as these structural injustices were acknowledged by vendors, they would tell me about how it was their mission to try to account for them, and thereby, to help change the outcome of employment decisions.

During our conversation about tackling gender biases, the Company Red CSM, for example, told me:

> "Our mission is: democratise professional development. Ok? And what that means is anybody in the world with any level of education or any background has an equal chance of getting ahead in their life and in their career in a way that they can positively impact their family economically or achieve whatever goals they have […] because we pay attention, we help organisations pay more attention to who people are instead of who people, um, what people have done or where they were educated or those things that show up on the resume […] everyone can learn a job, right, it's the opportunities to, you know, have a chance to, you know, learn those things that's the problem."

The Company Red CSM focuses on unequal access to qualifications or institutions, and how this might 'show up' or signal their suitability on a CV or resume. However, just as with the American dream, they emphasise that this should not be the case; people who work hard and demonstrate the right skills should be able to achieve their goals, regardless of their educational credentials. It was clear to me that, in discussions of bias, vendors were acknowledging inequalities in education, wealth, background, class, and

recognising that these cause disparity in opportunity. They were framing biases, including gender biases, in relation to structural injustice.

As seen above with Company Red, some vendors saw it as their mission, considering these inequalities, to equalise opportunity through their systems. As the Company Green CEO noted, "if there are things that people think are keeping certain sections of society from having access to the same opportunities, those are the things that need to be rooted out or solved for" and this was something which their technology attempted to work towards. Many of the companies believed that these structural injustices should not hold people back, and that their systems could help to achieve this:

> "It doesn't matter what school you went to, it doesn't even matter if you have a degree or not, if you're good, you get recognised, you'll make a lot of money, you'll change your life and your career." [Tigran Sloyan, CEO and co-founder of Code Signal]

> "[T]hey shouldn't be held back if um they don't have the right experience. If they're able to develop the right skills for the role they should be able to show it openly." [Company Green CEO]

Underlying the understanding of gender bias in relation to structural injustice, therefore, there was an interesting commonality between companies: some of these companies saw their technology as enabling social change and upward social mobility. At a conference, one company told the audience about the "social mobility benefits" of their systems, and the Pymetrics website asserted, "we're fixing the front door to hiring" (Pymetrics, 2022b). The Company Green CEO also expressed this, talking about how their main mission as a company when reducing bias was "removing barriers to upward social mobility".

The vendors' conceptualisation of gender bias in relation to structural injustice is a strong and important aspect of the frame – there must be recognition of the role of education,

wealth, and opportunity when it comes to inequalities in hiring. However, the vulnerabilities emerge regarding the frame in design.

It was my interpretation that many of these views of the vendors on structural injustices utilise narratives which parallel with the ideology of the American dream. For example, vendors say that "anybody in the world" should be able to "achieve whatever goals they have" [Company Red CSM], or that "if you're good, you get recognised, you'll make a lot of money, you'll change your life and your career" [Tigran Sloyan, CEO and co-founder of Code Signal]. The American dream encourages people to pursue their dreams, advocating that they have the freedom and opportunity to make what they want of their life (Hochschild & Scovronick, 2003). If you are willing to work, persevere, and play by the rules, the American dream tells you that you'll have a better chance to prosper than anywhere else on earth (Jillson, 2016).

Later in this chapter, I use this parallel with the American dream to show how the vendors' aspirations to provide equal opportunity are vulnerable to the same tangled web as the American dream. The American dream narrative tends to be interpreted as placing successes, and equally, failures, on the individual, instead of jointly focusing on individuals as well as structures in influencing people's lives (Hochschild & Scovronick, 2003; Hoyt et al., 2021). Similarly, AI-rec-tech systems rely on the achievement of the individual, regardless of structures. It is not that these systems – like has sometimes been seen with the American dream (Hoyt et al., 2021) – blame people for their failures. But if a candidate can exhibit the right skills and behaviours, they should be considered to get the job. However, rarely can people break free of the structures which bind them, and often these structures impact them in multiple – sometimes unforeseen – ways. These

structural inequalities constantly impact the lives and opportunities of people from less privileged backgrounds and in many cases, it cannot be expected that they will succeed in the face of these injustices (Hoyt et al., 2021; Putnam, 2015). Therefore, while their design might make a difference to rare and exceptional individuals, mostly, the continued influence of socio-economic structures will remain.

Not all vendors shared this conceptualisation of gender bias, and of those who did, there was a generally realistic attitude about whether their systems were able to completely account for, or indeed change the root causes of, structural injustices. Regardless, this conceptualisation – or frame in understanding - of gender bias as related to inequalities in education and opportunity is valuable and important in the context of hiring. Throughout the chapter, there is much evidence to support the vendors' view that structural injustice is a major factor which impacts hiring, exhibiting that this is altogether a strong conceptualisation of gender bias.

5.3 Accounting for structural injustice in design

Having explained how vendors conceptualise gender bias in relation to structural injustices in education and opportunity, here, I move on to answer sub-RQ#2 for this frame. I demonstrate that vendors operationalise addressing structural injustice in design in two ways: neutralising questions and assessing skills and behaviours (rather than qualifications).[20]

[20] Some vendors implemented these design features even though they didn't refer to gender bias in relation to structural injustice, education, or opportunity. It might be that these vendors did think about issues of structural injustice in relation to bias but didn't note it in the interview, or it might have been to account for a different root cause of bias. Alternatively, it might have been that this design, for some vendors, was not intended to counteract or address bias at all but was rather a way to find the best candidates for their clients.

5.3.1 Neutralising questions

The first way in which vendors attempted to address the bias of structural injustice in system design was through question neutralisation. This involved altering question wording and structure so that questions were understandable and accessible to as many people as possible. The aim was that, regardless of background or education, there was an equal chance of being able to answer the question to reflect ability. Vendors saw this to be a way of reducing the impact of education and opportunity on candidates' ability to answer the questions.

Vendors noted that the way in which questions are phrased might affect who can answer them and to what degree. For example, Sophia Baik, COO and co-founder of Code Signal, explained that a question on chess might be better understood by some than others:

> "It'll be a coding question, but it'll be, like, about chess. If you know how chess works then it'll be a lot easier for you to understand what are they asking, what are you to do, than if you have no idea, like, never played chess, right? Cos then you're spending time trying to understand what is being asked, what does it even mean, right? So those are also another way that we think about like, are we introducing advantages or disadvantages in the way that we're asking questions, so I know we tried a lot to make sure that we are checking for those, whether it's a gender or cultural thing but like are we unintentionally, like are we assuming that everybody knows this?"

Sophia is noting that, perhaps, some people are more likely to have been taught chess than others. This might be people who have grown up in a middle-class or upper-middle class family, and also, generally, chess is male-dominated (Dilmaghani, 2021). Therefore, a question using chess as an example would not be understandable to many and, indeed, might offer an advantage to the group with this pre-existing knowledge – in this case, men from middle or upper-middle class backgrounds.

Similarly, if people have grown up playing video games, they might be more familiar with assessments which use gameplay formats. Again, this is likely to be more common for men than women (Leonhardt & Overå, 2021; Rehbein et al., 2016). However, while success at playing video games might be related to being a drone pilot, it would be hard to prove it is related to other jobs, such as being a store manager (Andrews & Bucher, 2022).

The Company Blue CSO also spoke about ensuring people have an equal chance to be able to answer the question. They talked about the example of a test in the 70s which gave a question to kids about gemstones, asking a question similar to: 'which of the following best describes a ruby?'. They found that children in the African American community associated Ruby with the girl's name, rather than the gemstone, and so tended to give a different answer. It was therefore clear that the question phrasing was important to consider in terms of cultural contexts because people might interpret these things very differently. They ended off by telling me: "we really need to think about these things, you know, the words matter" [Company Blue CSO].

The vendors also noted that questions might - in extreme examples - be answered differently by men and women. They were aware that they needed to particularly alter questions where these differences occurred:

> "[B]ecause of the way a question's asked, subconsciously, all women might answer a certain way." [Company Red CEO]

> "[I]f you make your questions biased, right, obviously the data that those questions generate is that you've given a disadvantage to the female population because of the questions, which means the output of that data will be biased." [Tigran Sloyan, CEO and co-founder of Code Signal]

However, as Tigran told me, this can be changed by ensuring that you "don't use gender biased language, do not use things that are more sort of… give any sort of advantaged, like, I dunno, sports-based questions and sports that are way more popular with like, male than female, or any sort of thing" [Tigran Sloyan, CEO and co-founder of Code Signal].

Therefore, many of the vendors tried to neutralise questions so that they were equally answerable by people from all backgrounds and genders. This was a mechanism for reducing the likelihood of structural injustices – differences in education and opportunity - affecting the outcome of the assessment.

In terms of testing the interpretation of these questions, some of these vendors did investigate how the questions were received by different populations of people. For example, the Company Green CEO told me:

> "With one of our American clients what we did was, ur, we ran our questions, ur, with you know with basically a set of candidates, with a set of 100+ candidates, and we could find a tiny bias for, ur, you know, how some more men were answering that correctly than women […] so we had to rephrase the questions to be more gender neutral."

This was something they continued to do; there was "continuous tracking" [Company Green CEO].

The Company Red Partner also noted that question tweaking was an iterative process, telling me they would:

> "[G]o back and you might have to change the language because, to different people, you know, and that's often the case for example, or it used to be the case that particularly sense/feeling type questions […] because women see them softer than men see them, and, you know, so you have to kind of like find that halfway house."

In addition, the Company Blue CSO told me that they test the interpretation of questions through analysing every individual question and whether it's answered differently by demographic group:

> "[Y]ou could see if the question [is] favourable to one group versus the other. Um so for whatever reason is one group answering it […] are they responding a different way? And it's looking for those systematic differences in responding."

This stresses that while these differences could be from the phrasing of the questions (which vendors attempt to address), it could also be from differences in socialisation and therefore might reflect inequalities in our society.

In addition, Company Blue have cultural bias reviewers - people from all over the world who represent different ethnic or religious groups, for example, and who assess their questions to see if there is anything that could be differently understood. They tell me about one question which was picked up on:

> "There was a question, like an analogy or something with apple in it, and one of our reviewers in the Caribbean said, like, we don't really eat apples here." [Company Blue CSO]

In addition to thinking about the interpretation of questions by men and women, some of the vendors also considered different education levels:

> "[W]e pitched it at the top of the bottom quartile of 15 years old. So, basically, we're confident that the test works and can be understood by […] most kids aged 15 upwards." [Company Red Partner]

This means that their use of words is less complex, and it removes the ambiguity for anyone who might not have been educated beyond this point.

This question neutralisation applied to interview questions after assessments too. The Company Blue HOS told me that even if someone is not a good fit for the job, "maybe

they didn't have the same opportunities, but they do bring valuable skills, valuable knowledge, like, experience, to our team". Because of this, they tell me, they've tried to neutralise their interview questions, too:

> "I'm gunna ask questions that are a bit more cognitive based so I can see, can you pick up what I need you to pick up at the rate that I need, right, can you solve complex problems to the rate that we would need you to do in this role […] so we make sure that we're making the right decisions, right, that are inclusive, that promote diversity, that aren't excluding people because, you know, they might not have had the same opportunities."

Therefore, one way in which these companies operationalised addressing gender bias surrounding structural inequalities was through altering questions to neutralise their interpretation across different ethnicities, backgrounds, genders, and cultures.

5.3.2 Assessing skills and behaviours

The second way in which vendors attempted to address the impact of structural injustices was through measuring skills and/or behaviours in their assessments, instead of using qualifications or credentials as a direct indicator of skill. As Tigran Sloyan, CEO and co-founder of Code Signal told me, "[o]ur job is to measure skills and then show, prove that someone is qualified in a data-driven way".

Many of the company websites advertised their systems to be doing exactly this, stating that:

- They are the "first company to develop an objective skills-based assessment platform." (Code Signal, 2022)[21]

[21] Interestingly, some vendors did discuss measuring of skills and behaviours as 'objective'. This resonates with the discussion about subjectivity vs. objectivity in the previous chapter and so, arguably, this operationalisation could also be seen as way of addressing human subjectivity through system design.

- They offer assessments which look at specific skills and competencies such as emotional intelligence, past behaviours, personality, and that this will "make it easy to get the information you need about a candidate."
- Their "gamified soft skills assessments are the new standard for understanding talent" and that these assessments understand "inherent fit to role rather than using a one-size-fits-all approach or restrictive credentials." (Pymetrics, 2022a, 2022g)

Crucially, measuring skills and/or behaviours was seen to reduce biases, particularly those related to differences in background. Again, the websites of various companies advertised this aspect of the technology, saying that:

- They provide "tools to help you make hiring decisions based on merit, not on background".
- Their "platform digs deep into candidates' skills and experiences against their fit, without bias".
- Their "soft skills assessment carries no bias" (Flores, 2021) on the Pymetrics blog.
- CVs or resumes have "an inherent bias towards candidates with good credentials (education and work history)" but assuring that "good candidates can come from anywhere" which is why they test for "on-the-job skills".

Measuring skills/behaviours was seen by most companies as a mechanism to reduce bias which emerges through inequalities in opportunity surrounding education and background. Conversely, the disparity in qualifications which these inequalities produce was seen to be an unfair way of measuring whether a candidate was suited to the job. This was perceived to display the biases deeply rooted in our societal structures, whereas skills

and behaviours were seen to be a neutral way of measuring ability and suitability. As one Code Signal company report said:

> "[F]ocusing on skills and abilities, you don't limit your top of funnel based on keywords on a LinkedIn profile, university attended, or previous employers. This opens up your top of funnel to anyone who can get the job done – regardless of their background." (Code Signal, 2020)

This was even the case in jobs which required particular 'hard skills' such as coding or programming. Code Signal, which looks specifically at finding engineering or coding professionals, argued that it should not be using proxies like educational institution to access talent and skills, and one of their reports talked about how more than 50% of engineers are self-taught today:

> "[D]emocratization of education, especially around programming, has destroyed the notion that high-quality learning experiences are only accessible via educational institutions." (Code Signal, 2020)

They called this 'data-driven recruiting': not relying on education or background as an indication, but rather, measuring skills directly.

Almost all companies operationalised bias reduction through assessing skills and/or behaviours rather than qualifications. This came under three broad umbrellas: soft skills, behaviours, and skills. While some vendors measured a combination of these, at times they would focus their assessments on just one.

Firstly, some of the systems assess *soft skills*.

Many have argued that soft skills such as empathy, teamwork, critical thinking, should now be termed 'durable skills', because the term 'soft skills' is seen to 'feminise' the humanities disciplines (where these are often thought to come from) and undermine their

importance when compared to ‘hard skills’ provided in other fields such as STEM (Hutson et al., 2023). However, the vendors referred to these, generally, as ‘soft skills’, and the main proponent of this approach was Pymetrics.

As the Pymetrics company website states: “[i]t’s time to reinvent the status quo by leveraging soft skills, capabilities and unbiased algorithms to find the best person for any job”, continuing that they leverage “AI behavioral soft skill assessments to help enterprise companies build diverse teams of top performers” (Pymetrics, 2022h, 2022f). One Pymetrics white paper also says that “‘soft skills’ are the capabilities critical to delivering business values and adapting new hard skills as workforce needs evolve in the face of changing consumer and business landscape” and continues by explaining that soft skill intensive occupations in the job market are increasing, so employers should be considering how to attract top talent regarding soft skills (Pymetrics, 2020). They advertise that their platform gathers data from candidates to show these soft skills such as focus, effort, risk, attention, and decision-making. These were measured through gamified assessments, multiple choice assessments, and/or self-report questionnaires. Pymetrics advocate that if you base your hiring on soft skills, you are “identifying talent most likely to stay longer and perform better” (Pymetrics, 2022h).

With the rise of AI, soft skills have gained new attention. AI is seen to be increasingly replacing repetitive or routine tasks rather than jobs which require particular soft skills (UNESCO et al., 2022). It has been argued that this shift is potentially to women’s advantage (Sarah O’Connor, 2019) as soft skills have long been associated with women more than with men (Hong, 2016). In general, companies are placing renewed importance on soft skills, not only as crucial for jobs and working within teams (Grugulis & Vincent,

2009), but as a neutral measurement which doesn't place consideration on background or educational attainment.

So, how are soft skills defined? A Pymetrics report defines soft skills as, "the cognitive, social, and emotional characteristics inherent in people that are indicators of behavior and potential to succeed in different roles". They contrast these skills to hard skills, which are learned abilities acquired through training and experience, saying therefore that "soft skills hinge less on access to opportunities and are hence more evenly distributed across the population" (Pymetrics, 2021). Therefore, soft skills are, to Pymetrics at least, an element to overcoming the bias related to structural injustices, including disparity in opportunity. A post on their company blog also defines soft skills includes "interpersonal skills that include teamwork, collaboration, and communication. It is difficult to teach these skills to individuals" (Flores, 2021). Again, this echoes that soft skills are perceived to stand outside of pedagogical settings, and therefore outside of the realm of structural injustice.

In my interview with Alex Vaughan, Former CSO of Pymetrics, he told me that:

> "Importantly, the tools that we use are focused on behavioural or soft skills, rather than hard skills, um, and they are not um, they're not performance-based in the sense that there's no right answer for a lot of the assessments that we do."

Alex continued this point by saying that they "measure a whole suite of these, ur, soft skills, that really influence the kinds of jobs that people enjoy and are good at in their day-to-day life, and then what we try to do is we try to build machine learning models that match people to certain profiles of what a successful employee looks like in these jobs" [Alex Vaughan, Former CSO of Pymetrics].

He then gave me the example of comparing a stockbroker and an accountant in terms of risk-taking skills, explaining that while they're both in finance, stockbrokers deal with risk every day whereas an accountant might need to be more risk averse. Once someone has played their gamified assessment, this data from the assessment would show their approach to risk-taking, and shed light on whether they match the optimal risk-taking profile for each job. He went on to tell me that this is more of a neutral measure, because it avoids measuring based on qualification or institution, and therefore doesn't base the decision on the biases emerging from structural inequalities in education:

> "They [clients] were optimising for a very particular definition of what good looks like. So good looks like you went to Harvard, or good looks like you did a math undergrad with a business minor. You know, this very narrow range of experiences that comes with the definition of hard skills. Pymetrics doesn't do that at all. Pymetrics focuses on soft skills and so if you select on soft skills, you will automatically get a different kind of diversity than selecting on hard skills." [Alex Vaughan, Former CSO of Pymetrics]

Secondly, some companies focused on *behaviours*.

Company Red emphasised that they try to avoid a selection assessment which will be measuring people based on education or qualifications, but instead measures behaviours and motivators. They argue that behaviours are what initiates and sustains culture of a workplace, and so should be a crucial starting point for these companies; creating an inclusive culture, they say, will also help to lower turnover of employees.

The Company Red CTO told me that this technique of measuring behaviours allows it not to be weaponised by people, because it gives 'social desirability neutrality':

> "Nothing says high or low, it's deliberate or decisive, you know, outgoing or reflective, those are all good things or bad things depending upon the situation. […] So that was kind of the start where we tried to eliminate bias was not make it, you know, where people could weaponize it in any way by saying this is good and this is bad."

Company Blue also analysed behaviours or cognitive traits rather than skills, using a self-assessment questionnaire. They do this by recommending to their clients that:

> "[Y]ou have to do something called setting a job target which is: here's the behavioural demands of the role, or here are the cognitive demands of the role, and every person that takes the assessment, what you're encouraged to look at is how they match up to that, what their relevant fit is […] it's not just like automating a decision that could adversely impact certain groups."

This means that "you're not even thinking about a person, you're thinking about the job" [Company Blue CSO]. The Company Blue CPO also expands on this, telling me that this allows the reduction of bias in these systems, as they're not measuring qualifications, education, background, but instead they're measuring behaviours:

> "[I]t's all about behaviour, behaviour patterns are neutral and therefore they're objective, right." [Company Blue CPO]

Therefore, similarly to soft skills, measuring behaviours is seen as a way of levelling the playing field for candidates; this focuses on the requirements of the job, rather than their qualifications or credentials, and in addition it might help to create an inclusive culture by encouraging a variety of behaviours rather than good behaviours and bad ones.

Thirdly, some companies look to measure what would be referred to as just *skills*.

These weren't referred to as soft skills; these were skills related to the job their clients were aiming to fill. Really, this would usually measure what would traditionally be thought of as 'hard skills' such as coding, programming, or maths. Some vendors referred to this as hard skills. For example, in one report they say: "being able to actually assess hard skills is critical to understanding the potential value add they bring to your organization" (Code Signal, 2020). On the Pymetrics website, they say they are

"prioritizing roles based on soft skills as well as hard skills matches" (Pymetrics, 2022d), with their blog expanding on this, saying that "[t]he right combination of hard and soft skills amounts to the perfect candidate for a job" (Flores, 2021).

However, some vendors just referred to them generally as 'skills'. These aimed to be a direct measurement of ability, the idea being that this would not discriminate against anyone without formal qualifications.

Code Signal told me about their technique to reduce gender bias from structural inequalities. Measuring the candidates' skills this way was seen to avoid some of the biases that could otherwise occur:

> "Our job is to measure skills and then show, prove that someone is qualified in a data-driven way so that it's not at the mercy of the interview." [Tigran Sloyan, CEO and co-founder of Code Signal]

From the same company, Sophia Baik also explained to me that this helped women, because prior to this skill measurement, women would be perceived according to their stereotype (as discussed in Chapter 4) rather than their ability. She told me that "people will say oh we hired this woman because she's a woman, and everybody's talking about hiring women, so like we hired her but oh there is no way this person is any good. Right? So that's the kind of conversation we wanna change" so that everyone "saw the score and there is no question about whether this person is capable or not" [Sophia Baik, COO and co-founder of Code Signal].

To many of these companies, therefore, measuring skills helps to mitigate biases related to structural injustices. The Company Yellow IOP1 also emphasised that they, "try to make sure that there is a focus on the actual job, um, that we create the assessment for

those particular roles, so not the person. So that really helps to create a minimal bias there". They continue by saying to me that there can't be bias in these questions: "I'm gunna ask you, let's say, a math question because you're gunna apply for a finance role, well, it's a right or wrong answer, there's no bias there at all". The Company Yellow IOP2 also tells me that because it is focused mainly on the job, it really "does a good job at already not being biased", mainly because this is "all focused on job role really and work environment" which helps to be less biased and to focus on how they're going to perform on the job.

Chanukya Patnaik, CEO of Skillspace.ai also explained their assessment of skills for coding-based jobs:

> "Ur, in our case we are focused on skill-based hiring, rather than […] gender-based, so it's purely skill-based. If someone has the ability to code and able to solve some problems, then yes, I mean, you're ahead of others in the pool of candidates."

He explained how they do this based on a standard software development process; every developer is expected to be good with logics, data structures, algorithms, so they use questions based on these to assess candidate capabilities and their coding aptitude. This technique, Chanukya said, allows them to remove some of the gender bias, which emerges when they're screened based on the qualifications which they have, but instead, "we see that ok irrespective of what gender they are, what, ur, diversity they come from, I mean, like […] their ability to code or ability to provide real world solutions" [Chanukya Patnaik, CEO of Skillspace.ai].

5.4 Two 'out-of-frame' vulnerabilities

I have shown that there are two main ways in which these vendors attempt to address structural forms of gender bias.

In this section, I now consider how these design mechanisms are vulnerable to 'out-of-frame activity'. I outline two types of vulnerability to which the structural frame is subject. These out-of-frame elements are external or unknown factors which impact the function of the design.

As mentioned earlier, this is not to say that the vendors always believe their systems to fully account for these structural injustices. Generally, vendors were realistic. None of them claimed to be able to solve the problem of structural injustice or to account for it entirely. This is an important distinction – while I am analysing the vulnerabilities of being able to account for structural injustices, I am aware that vendors weren't attempting to solve for these injustices entirely. Many of the companies were blunt about their intention to make a positive difference in a system much broader and wider, and therefore out of their reach:

> "What we're looking to do is have a… you know even a small positive impact on these outcomes that we're talking about. And, to me, that has to be enough […] you know, have the positive impact and then be able to walk away, you know, from, you know, what ultimately, you know, what is a broad systematic problem that's going to take decades to resolve." [Company Red CSM]

> "Some people just have greater access to resources, they can pay for additional tuition, they have better mentors because they have access to society and schools, you can't solve for those." [Company Green CEO]

> "You cannot fix everything, especially, like, one company cannot fix everything […] Cos it's one thing to say we're gunna go change, you know, create more gender equality in education so that there's just as many or more women in computer science and programming than men. It's another thing to say at least for those that have gotten there already, through all the biases, through all the pushback, through all the probably their own parents telling them they can't do it, at least recognising what they've been through at that point." [Tigran Sloyan, CEO and co-founder of Code Signal]

> "I think what we can do is, like, how we can reduce bias and also start to look further into what are some of the driving factors that can be… and how do we actually make changes in that regard." [Sophia Baik, COO and co-founder of Code Signal]

In addition, Sophia told me that they "cannot actually normalise" for the fact that men start programming much younger than women. And the Company Red CSM told me that their aim as a company to be an "influential and impactful contributor to solving that problem", but they also recognised that this is within "a complex ecosystem of multiple, um, angles and factors and influences".

This sheds light on the frame in design: the operationalisation of attempting to address gender bias in structural injustice seeks to make the best of the current situation, rather than to change the root cause of these inequalities.

Therefore, I acknowledge that the design is not seeking to solve structural injustice itself, and I acknowledge that it might have advantages for those rare applicants who – in the face of poor education and lack of opportunity – have managed to obtain skills without credentials or formal training. However, I nonetheless want to consider the vulnerabilities of a design which attempts to account for these structural injustices.

Goffman's notion of 'out-of-frame' activity details how framed activity provides a focus of attention but encourages consideration of the things which occur apart from this dominating line. While people have a main line of activity, there are things which are 'out-of-frame'. These are "subordinated in this particular way to what has come to be defined as the main action" (Goffman, 1974, p. 201). Individuals have a main 'track' or 'channel', but we can also look at the channels of subordinated activity – events or deeds

that appear to be dissociated. This means that "during the occurrence of any activity framed in a particular way, one is likely to find another flow of other activity that is systematically disattended and treated as out of frame, something not to be given any concern or attention" (Goffman, 1974, p. 210).

Therefore, a focus on 'out-of-frame activity' prompts exploration of areas which occur outside of the vendors' main 'track' or 'channel' regarding their understanding and design. This might be factors which are outside of their control, or factors outside of our conversations. Illustrating these as vulnerabilities allows me to show how factors which are 'out-of-frame' - or out of focus - do have the potential to disrupt the integrity or function of the frame in design.[22]

Goffman discusses that there is indeed a stream of 'out-of-frame' activity which is consequential for the main activity: "there is to be found a stream of signs which is itself excluded from the content of the activity, but which serves as a means of regulating it, bounding, articulating, and qualifying its various components and phases" (Goffman, 1974, p. 210). There is some 'out-of-frame' activity which doesn't affect the frame, and there are some activities which are consequential.

In this section, I argue that there are two types of 'out-of-frame' activity which could be consequential for the integrity of this frame in design. There are, as it were, two 'out-of-frame' vulnerabilities to the structural frame. The first out-of-frame vulnerability is 'out-of-frame permeability', the second is 'invisible out-of-frame'.

[22] As mentioned previously, I argue in this chapter that the vendors' conceptualisation of gender bias as related to structural injustice is well-founded, and so my analysis applies more to the operationalisation – the frame in design.

Largely, both focus on analysing the operationalisation of assessing skills/behaviours more than question neutralisation. Question neutralisation is, in theory, an effective way of ensuring that people can answer the question, regardless of their background, gender, class, and so on, although, similarly to those design features in Chapter 4, there is no evidence either way.

5.4.1 Permeability: out-of-frame influences

Permeability - in scientific terms - indicates the crossing of liquids or gases through a surface. It means a surface is porous and penetrable. Translating this to the structural frame, I define 'out-of-frame'[23] permeability as: realities out-of-frame, or out of focus, which might permeate and therefore impact the frame in design. I discuss how this might happen in two ways:

(1) **Structural injustices might impact soft skills and behaviours:** there are ways in which structural injustices - for example, inequalities in education - might still affect the supposedly neutral measures of soft skills and behaviours.
(2) **Structural injustices impact hard skills:** differences in education impact people's hard skill development, which are measured by companies in the form of 'skills' and will be required for more highly skilled professions.

I deal with each of these permeabilities in turn.

[23] These factors are out-of-frame as they generally go unacknowledged by the vendors, and therefore are not within the vendors' focus.

Firstly, there is no guarantee that soft skills and behaviours are a neutral measure, isolated from structural injustices such as inequalities in education. It is possible that these factors could relate to background, education, gender, class, and so on.

Education type and quality are likely influenced by socio-economic status. This is not just because socio-economic status influences the school which children attend, but because it influences their social surroundings from an early age. Gillani & Eynon (2023), for example, discuss how educational outcomes are influenced by social factors, such as children's connections to schools, parents and families, and the broader neighbourhood, which shape children's education and life from an early age. Lareau (2002) also observed class-based differences in parenting style, finding that parents of more privileged children performed 'concerted cultivation' – interfering more with schooling and attempting to foster more organised leisure activity, whereas children from less privileged, working-class families tended to have their parents practice 'natural growth', which left them to their own devices regarding leisure activity. Overall, this impacted their varying levels of entitlement (Lareau, 2002).

Similarly, children from middle-class backgrounds have been found to be taught, by their parents, to ask for help more at school. As a result, they receive more attention than their working-class peers, including better marks, more recognition for their ideas, and more opportunity for creativity (Calarco, 2018; also see Rafalow, 2021). Putnam (2015) also interviewed and spent time with vulnerable kids from low socio-economic backgrounds, and from this argued that the experiences that children have – the families into which they're born, the parenting and schooling which they receive, the communities they're raised in, these things have a powerful effect on their lives.

These socio-economic factors – children's parents, neighbours, friends, friends' families - influence children socially, as do the educational institutions of which they are a part.

We see from literature within education and sociology that structures impact people in many ways, not just through the hard skills or qualifications which they gain. Educational institutions shape us socially. Eynon (2022) notes that the philosopher Biesta offers a framework for understanding the purpose of education: education is about 'qualification', equipping young people with skills; 'socialization', helping young people to become a part of a collective social, cultural, political order; and 'subjectification', helping individuals become independent and autonomous citizens (Biesta, 2009, 2015). Depending on the setting of education, the teachers, the fellow students, the form and content of the education itself, people will gain different qualifications, but they will also be socialised differently. Emile Durkheim, an educational sociologist, for example, argued that education is a socialisation process which shapes our norms, values, and beliefs – referring to this as 'cultural transmission'. Education has more influence than we give it credit for; it can change the course of people's lives, including their socialisation and behaviours (Durkheim, 1956).

Goyette (2017) also discusses how education can influence our social side, explaining how sociologists have identified two types of curricula:

1. The formal curriculum (explicitly taught in schools – knowledge or skills).
2. The hidden curriculum (subtler than the formal curriculum, we may learn what is expected of us based on our anticipated roles in society, the unspoken rules we should follow to be rewarded).

These curricula are not just dictated by the type of school we attend and how that school educates us, but also by our gender, ethnicity, race, religion, and so on. Schools therefore provide stratification by qualification and grades, and also play a role in socialising us to abide by gender roles (Legewie & DiPrete, 2012), racial roles, and/or class roles. Schools can reinforce the gender binary (Thorne, 1993) and might reproduce the dominant gender norms throughout the curriculum (Evans & Davies, 2000), prompting young people to associate particular things with women and men, for example normalising 'masculine' or 'feminine' sports or subjects (Domina et al., 2017). Teachers also act as role-models and, unfortunately, girls, children of colour, and those in low-income environments are often less likely to be exposed to inventors who share their backgrounds, reflecting longstanding structural inequity in society (Gillani & Eynon, 2023). This might impact how students learn to think, respond, and act. And therefore, these structural differences might affect their skills and behaviours, as well as their qualifications, and it might do this according to class, gender, race, ethnicity, even if the exact patterns and ways in which this happens are not always generalisable.

Vendors themselves tell me that sometimes they find patterns in how different genders are answering questions. The Company Red CSM told me:

> "[If] I set the metric like the target at low helping or low altruism that will have a bias against women, ok? Because women tend to hit higher on altruism. So, it'll tend, ur, to kind of inadvertently score women lower on average, right."

We saw earlier in the chapter, too, that the vendors find socialisation can affect how people answer questions and that this can be the case according to gender. Later, the same participant tells me that "there are cases when we can tell the difference between men and

women. Which means that you have to be extra careful about the downstream effects of that bias and people losing sight of the ripple effect" [Company Red CSM].

It might be, for example, that if a soft skill or behaviour is desirable, and that candidate is not displaying it, they are less likely to get the job. While vendors attempt to account for this through question neutralisation, it might not always be possible if socialisation is playing a role in answering even the most neutral of questions. In these cases, this permeability from 'out-of-frame', educational influences on soft skills and behaviours, is impacting the function of the frame in design.

Education and background are likely to impact people's skills and behaviours, along with their qualifications and credentials. While some vendors might see these qualities as innate or neutral, it is probable that these structural inequalities are impacting the factors being measured, too. This raises questions about how education and socialisation (within educational institutions and outside of them) might impact soft skills and behaviours, particularly along gender lines and particularly as measured by these systems. This must be considered further, along with how these trends might maintain gendered positions in terms of income, job roles, sectors, and seniority (Grugulis & Vincent, 2009).

Secondly, the measurement of 'skills', by which vendors often mean hard skills, will also be influenced by structural inequalities in education. Hard skills generally refer to measurable technical abilities that can be attained through training, education, or experience. These hard skills will still be required for the most highly skilled, and potentially best paid, jobs. Therefore, we might consider permeability in terms of how

structural injustices within education will continue to affect the outcome of these systems as they're currently designed and used.

In some cases, though not all, AI-rec-tech assessments will take place after there has been an initial screening of candidates, during which there will be measurement for qualifications or experience. In other cases, some of these systems will immediately measure (hard) skills, without relying on qualifications or credentials to indicate them. Still, candidates will need to exhibit the ability to programme or code, and it is likely this will be related to their educational background and the opportunities they've encountered. However, of course this design does enable people who have gained these skills without qualifications or certain credentials to exhibit these skills and be considered.

Even with the 'democratisation of education' through increased free and accessible online educational resources or massive open online courses (MOOCs) (Acemoglu et al., 2014), there are barriers to accessing these settings and developing these skills (Barger, 2020). For example, women are less likely to enrol in STEM MOOCs (Jiang et al., 2018; Urban, 2022). Moreover, some empirical studies have shown that students of MOOCs are primarily attended by those already advantaged in terms of education and socio-economic status, and hypothesise that while these methods of learning might reduce cost barriers, they will do little to reduce global educational gaps until they are able to reach and retain those most lacking education (Pollack Ichou, 2018).

The ability for women or people from less privileged backgrounds to participate and learn to code, for example, depends on whether they have access to a laptop, internet, whether anyone presents this to them as an option in the first place, or whether they are able to

take this up due to unequal caring responsibilities or working other jobs to provide for their family (Shapiro et al., 2017). We know, for example, that in the US, while telephones are concentrated among young, low-income, and Black and Hispanic communities, it is these communities who also rely more heavily on their phones to accomplish more complex tasks such as looking for jobs, filling out applications online, and creating resumes (Ticona, 2022). It is unlikely that these groups will spend time searching for and completing MOOCs in preference of sourcing a job, and it is likely that it would be difficult to conduct a MOOC on a phone because, for example, it might require complex software which can only be downloaded to a laptop.

We must also consider that there is a digital divide between men and women. Digital inequalities can often be understood as emerging from the intersections of gender, disability, ethnicity, and age, as well as economic factors (Eynon, 2023) and education (van Deursen & van Dijk, 2011). Women's use of, and benefits from, digital technologies remains limited owing to economic, social, and cultural obstacles – the gender digital divide is persistent irrespective of a country's digital access or economy (Acilar & Sæbø, 2023; Pawluczuk et al., 2021). Therefore, MOOCs are not necessarily a silver bullet for addressing disparities in educational access and outcomes (Gillani & Eynon, 2023). In addition, much of the democratising power of settings such as MOOCs will be outweighed by the disproportionate advantage of the most privileged, who have access to powerful networks and expert advice, along with private or tailored tuition (Rivera, 2016).

Either way, it is likely that hard skills and/or qualifications are still required for certain jobs. Indeed, some jobs will require certain qualifications and credentials to even be

considered. For example, Verma et al. (2022) found that the occupational (hard) skills are ranked at the top for both AI and ML positions, e.g. candidates need to be capable in data mining, statistics, big data, programming. While soft skills were found to be important – e.g. attitude, motivation, independence – they were not as important as hard skills.

Altogether, this makes it likely that structural inequalities in access to education are still influencing – or permeating – a system design which measures skills and behaviours. And again, this is more likely to harm some people than others.

In the US, for example, there is inequality in educational attainment across all social groups: socioeconomic background, race, ethnicity, and gender. The major differences come with socioeconomic background and race (Goyette, 2017). Mountford Zimdars (2017) talks about how a third of students in the 146 most selective US colleges come from the top quarter of the socio-economic status scale, as measured by parents' income, education, and occupation.

It is widely recognised that the education system in the US reproduces these inequalities through its structure. Darling-Hammond (2001) argues that the US educational system is one of the most unequal in the industrialised world, and that students routinely receive dramatically different learning opportunities based on their social status. The wealthiest 10% of school districts in the US spend nearly ten times more than the poorest 10%, meaning that minority students have fewer and lower-quality books, curriculum materials, computers, significantly larger class sizes, less qualified and experienced teachers, and less access to high-quality curriculum (Darling-Hammond, 2001). Public education in the US is deeply local, with education based on residence, which itself is

based on wealth. This means children in affluent (largely white) districts receive better funding and better education than those in poor (largely minority) districts (Hochschild & Scovronick, 2003). Hochschild (2003) gives the example of two schools in the same state, one in the city where the resident of the district per capita income of $17,000 a year, and another in a suburb where the same is $70,000. The first is underfunded, has underqualified and understaffed teaching, 98% are non-Anglo,[24] almost three out of ten need special education. In the second, they have computers in every classroom, most students do extra-curricular activities, there are good teachers, and 95% of the students are Anglo (Hochschild & Scovronick, 2003).

These inequalities in education in the US are therefore largely disadvantaging low-income racial minority groups (Darling-Hammond, 2001; Gillani & Eynon, 2023; Hochschild & Scovronick, 2003) meaning these groups are less likely to perform well in school in comparison to their high-income counterparts (Reardon et al., 2018). It could also mean that students at more under-funded schools are less likely to take certain subjects, such as STEM. As we saw above, there are certainly gendered differences when it comes to the uptake of STEM subjects.

While this research cannot conclude that the direct measurement of hard skills does not neutralise for background, it can point out that there is likely to be a permeation between inequalities in education to the development of hard skills. It is likely that the realities of these structural injustices continue to impact the factors being measured by the vendors, even if it is not being measured through credentials or qualifications. In other words, it is

[24] This is the term used by the researcher. I interpret ‘non-Anglo’ to mean a person who is not White, and ‘Anglo’ to mean a person who is White.

less likely that low-income racial minority groups have had the opportunity to develop these hard skills because of these inequalities in schooling. While some people might choose to pursue online learning or free courses, or might be innately talented at maths, these people are likely in the minority.

I return briefly to the ideology of the American dream. One factor which has been viewed as hindering the dream is that opportunities are often dictated by structures. We see the same here. Individualism doesn't work if structures dictate your options and opportunities (Hochschild & Scovronick, 2003).

We see hint of this in the data, too. We know that, in the US, among graduates from the same university, students from the most elite backgrounds continue to get the highest paid jobs (Rivera, 2016). In addition, education levels directly corresponding with income. In the last quarter of 2023, in the US, full-time workers aged 25 and over without a high school diploma had median weekly earnings of $713, compared with $917 for high school graduates (no college) and $1,608 for those holding at least a bachelor's degree (BLS, 2024).

In the UK, similarly, we see those demographic factors such as class and gender influence grades and qualifications. Regarding class, children who are not enrolled in school are concentrated in the most vulnerable households (Tuñón & Martínez, 2021). In England specifically, young people who are from poorer families (identified by their eligibility to have free school meals) are half as likely to continue into higher education than those not eligible for free school meals (Mountford Zimdars, 2017). In terms of gender, while girls

have recently outperformed boys at GCSE level (Clark, 2023b), in 2023, 8.8% of female students and 9.1% of male students achieved an A* grade at A-level (Clark, 2023a).

I have now explained the second way in which this frame is permeable - and therefore vulnerable to 'out-of-frame' influences – through the continued influence of structural inequalities on qualifications and credentials. This outlines the vulnerability of **'out-of-frame permeability'**. These qualifications and credentials will still be required for highly skilled, highly paid jobs, whether measured directly via testing or by screening a CV. However, it is true that measuring skills directly might capture some candidates who have not had formal education or attended a 'top' university. Some of the vendors do not claim to, or aim to, do more than this; they recognise the deeper structural roots of these inequalities in education and opportunity. However, this vulnerability is crucial to note as it raises questions about the extent to which free, online learning might be affecting individuals' ability to obtain highly skilled jobs.

Moreover, the wealth of theory and evidence here that structural injustice likely impacts soft and hard skills only continues to strengthen the conceptualisation of this framing of gender bias by the vendors.

5.4.2 Invisibility: exclusions from the frame

Invisibility is the second way in which the structural frame could be vulnerable to 'out-of-frame' activity. While vendors present skills and behaviours as metrics which do not have to be influenced by structural factors such as education, it is unclear what these skills and behaviours are measuring and how. Therefore, I define **'invisible out-of-frame'** as:

crucial factors of the design which are invisible to the researcher, but which impact the function of the system.

This frame is also vulnerable to the invisible aspects or the 'underlying science' which, as researchers, we have no certainty about (either purposefully or not) and therefore cannot know if this assessment of skills and behaviours is measuring what it claims to measure. These invisible elements encompass crucial information to understanding the function of the frame in design (Chamorro-Premuzic et al., 2016; Raghavan et al., 2020).

'Science' was a word commonly used by the vendors. For example, some websites told me that their tool was "built on 65 years of science", another that they provided "scientifically backed psychometric tests", and another discussed that "[p]ychometrics is the science of measuring mental capacities and processes. It applies and investigates theories of personality and cognitive ability to build tools that measure people's underlying traits and abilities", or yet another which said that if using their tool, companies are able to "hire the right people for the right role with recruitment tools that leverage the most powerful behavioral science that exists."

The topic of 'science' came up in interviews as well:

> "Yeah, I mean in the assessment it's thirty years of research in measuring, you know, using established scientific, psychometric baselines […] to make sure that our assessment measures what it says it's measuring." [Company Red CTO]

> "[U]p until now you didn't really have anything other than your intuition […] so now it actually… you have some science where you can say well look here's why this person is not right, is not the best person for the role […] They just say, look, based on the science, these are the people that are better." [Company Red HOS]

Here, I outline what I could see from my data collection, in terms of both how the relevant skills are decided upon, and what those skills themselves mean. In doing this, I show the missing parts – the invisible parts – to the puzzle, and I highlight the questions these raise for future research on AI-rec-tech. Particularly, I argue, these invisible, 'out-of-frame' elements, raise questions about what is meant by the underlying 'science' of these systems.

As noted in Chapter 4, there are multiple ways of deciding which skills are necessary for a job. First, vendors can train the model on previous high performers in that role. This might be specific to the client company, or it might accumulate high performers across multiple companies. Second, vendors pre-build the assessment specific to a particular job, for example using IO psychologists (industrial and organisational psychologists, who focus on the psychology and behaviours of people in the workplace). Third, vendors might offer customisable assessments which the client can adapt to their needs. Fourth, they might offer an assessment which, instead of producing target variables, provides a competency report with scores on cognitive and behavioural traits, which clients can interpret as they wish (Raghavan et al., 2020). For the third and fourth options, which have been spoken of previously, the selection or interpretation of skills are dictated by the clients. I dealt with this in Chapter 4. Here, then, I deal with the first two options.

Some companies use the first of these options. At Pymetrics, for example, they take what makes people good at the job, which is built up through accumulated client data, and then measure these high performers through cognitive testing. Alex Vaughan, Former CSO of Pymetrics told me, for any client:

> "We use the profile that is actually there in the data, not some profile made up by an analyst or a psychologist. And so, what we try to do in our assessment is we

> try to measure a wide diversity of things. Under the assumption that if we measure, say, twelve relatively stable, um, behavioural traits like decision-making under risk, like value estimation, like generosity, like emotional context reasoning, um, some of those will be correlated with success in any given job and some actually won't. […] And so, um the broad approach is to measure many things and then let the data tell us what the answer is."

As noted in Chapter 4, this raises questions about the datasets which these systems are being trained on – for example, how are high performers are being chosen?

Other vendors use the second of these options, using IO psychologists to shape the skills required for jobs. IO psychologists help to design the profiles of ideal candidates and implement tests to measure these qualities. The Company Green CEO told me about how they implement knowledge-based and skill-based assessments using IO psychologists:

> "So, we have in-house expertise in terms of IO psychologists who first review sample job descriptions from our client. We gather that experience over multiple clients and then we of course start to build our own internal library, of, ok this job usually requires these particular skills and competencies. So, we continue to improve that internal library, and internal definitions of roles we have, based on whatever we're learning from our client job descriptions." [Company Green CEO]

In other words, they break down the job description and decide on the skills needed for that, and the ways to measure it:

> "Each job can be broken down into competency, skills, knowledge, and maybe even experience required for it. And then, for each of those items, you can decide on various ways to assess the candidate." [Company Green CEO]

Sophia Baik also told me that they use a team of IO psychologists. First, these IO psychologists "help them also think through what are the skills that we're trying to measure in which stage and what's the best way of measuring those". They create the assessment like a recipe, they tell me, based on their framework:

> "Right, like what are the key skills we're trying to measure, why is this skill important for a certain role, um and how we're gunna measure it and how we're

> going to also score the answer." [Sophia Baik, COO and co-founder of Code Signal]

For example, they might create a framework for an entry level software engineer by breaking that down into different skill sets. This might require programming skills, coding skills, and understanding basic data structures, algorithm problem solving, and implementation of a code. It's through this that they define what the question should look like. In Company Yellow, I also spoke directly to two IO psychologists who do exactly this type of work for their company's tests.

Although not referring to them as an 'IO psychologist', the early start up, Company Orange, also said they were planning to do things similarly:

> "[W]hen we've standardised the jobs […] we'll then be drilling down with, we've got a talent product manager, around um core competencies that are required for each of them jobs within them families. And then once you've got the core competencies what are the core assessments that we could do to see whether these individuals would be relevant?" [Company Orange CPO]

Chanukya Patnaik also spoke about how they have a pool of about 100-120 experts, software engineers at different tech companies, who set their questions:

> "[W]e have a pool of ur experts who sort of create these questions, ok, and put it on our library. So, we have a rich library […] the recruiter who is recruiting can pick any question of their choice, or ur, if they want us to automate the process, there's a functionality where, ur, I mean if you select a role, it will automatically generate questions for you." [Chanukya Patnaik, CEO of Skillspace.ai]

Further to this, they explained to me that it's difficult to have bias at this devising stage, because these questions will focus on their abilities to code or solve problems, regardless of their gender or their background.

The processes of devising desired skills or questions, whether by IO psychologists, experts, or product managers, are invisibilities which beg questions about exactly how

these skills are being chosen. Overall, it is important to acknowledge that the way these skills are being chosen is largely invisible to researchers and requires more investigation; this might carry implications for who is being selected for jobs and might correlate with factors related to structural injustice such as education.

Furthermore, another 'invisible out-of-frame' vulnerability is the construction and measurement of the skills themselves. Of course, it might be easier to measure hard skills, such as programming, than so-called 'soft skills', such as empathy. But the point remains: for example, how are vendors scientifically verifying that someone sits in a particular place on the empathy spectrum?

It was not that I didn't ask this question to the vendors. At one point, I asked a participant what the research was based on, and they told me they weren't sure they were allowed to tell me the name of the scientist, but it was based on 30 years of research.

When looking back at my demos and free assessments, clearly, I was thinking about how I was being defined in certain ways or labelled to have certain traits. It has been noted before that these competencies are constructs and attempts to measure them must rely on other psychometric assessments as 'ground truth' (Raghavan et al., 2020), but what are they based on?

In my diary from my free assessment with one vendor company, I wrote: "I don't understand what lies behind this - what is the scientific basis for these questions and for calculating my answers?". For example, one free assessment provided me with a profile which labelled me as words similar to 'sociable' and 'impulsive', generally profiling me

as someone who thrived on change instead of routine.[25] Similarly, in another free assessment I was labelled as a term similar to ‘advocate’, which was someone who took risks, was a team player, and was socially able. In some cases, these profiles even provided me with images or cartoons to represent these qualities. This was interesting; it made realise even more so how I do not necessarily resonate with these labels. Figures 7 and 8 below loosely replicate some of these images, without taking screenshots from the actual free assessments.



Figure 6: Replication of assessment result image 1, credit: Joanna Rivera-Carlisle

[25] I cannot share the exact terminology in this section due to company IP and copyright, but I have selected similar words.

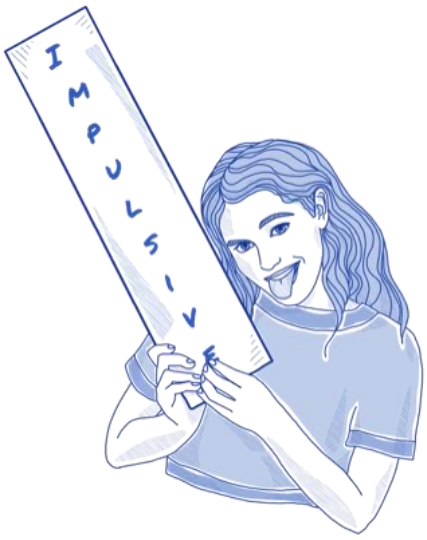


Figure 7: Replication of assessment result image 2, credit: Joanna Rivera-Carlisle

As someone who is largely introverted, not spontaneous, risk averse, and not fond of change, I was slightly surprised at these results, and I was even more surprised that I got similar results across different tests. However, I must have done something to bring them about in both cases, and I was curious about the scientific basis of these tests in terms of how these results are being produced. What metrics and underlying decision-trees were being used to label me in this way? And, for another research topic altogether, how does this feel for candidates who don't get jobs because a test, in their eyes, has mislabelled them?

Having said this, one of the free assessments I completed did resonate more with me, providing me with a profile which labelled my strengths with terms similar to: studious, team-player, fast-paced.

Certainly, though, these free assessments made me question the underlying 'science' of the tests and wish I could see these invisible aspects of the profiling. Of course, this also impacts how effective design features such as re-defining fit might be, as discussed in

Chapter 4. The reversal of expectation and desire relies on these tests accurately estimating these qualities.

In one demo, I wrote about how I was surprised at how some of these tests covered jobs which I imagined would be more related to customer service (e.g. baker, bartender, fast-food cook), or related to meeting someone in person (e.g. personal care aide, childcare worker), or reliant on seeing a person's skills for the job (e.g. carpentry, assembler of electronic equipment, cleaner, cook, firefighter, electrician, construction labourer). I therefore decided to take the test for electrician and received almost 70%, which put me in the moderate to high performance potential. Given that I have no training or experience in this area, I was surprised to receive this score. While I won't dwell on the point, I think it is safe to say that I would not make a good electrician. This, to me, was another reason why I questioned the scientific basis of these tests, and the ability of these assessments to effectively assess certain jobs.

When it comes to measuring these competencies, 'soft skills', behaviours, and even the hard skills required to be an electrician, there must be something underlying these decisions, and it will be crucial in future research for industry and academia to collaborate to explore the underlying 'science' further.

It has been argued that attempts to measure behaviours, for example, must rely on indicators from other psychometric assessments as 'ground truth' (Raghavan et al., 2020). Sloane et al. (2022) discuss how there are possibilities of this being linked to the popular taxonomy of the OCEAN model, which models the Big Five personality traits – openness to experience, consciousness, extraversion, agreeableness, and neuroticism. Indeed,

Drage & Mackereth (2022) conducted their study on video AI tools that claimed to use the Big Five psychological assessments to reduce bias in the hiring process, but found little information about how they map this onto candidates bodies or interview responses. The mitigation of the details on this leads to questions about which psychometrics or behavioural science indicators are being used as the basis for these assessments, and what implications might this have for different genders, along with other minority groups?

Cognitive assessments certainly have a history of imposing adverse impacts on minority populations (Raghavan et al., 2020), and so it is crucial that we understand what assumptions underlie the 'skills' that are being tested, and how this might impact certain groups. Psychological testing from its inception has been found to carry inherent biases against certain ethnic groups, commonly known as the Cultural Test Bias Hypothesis (CTBH) (C. R. Reynolds et al., 2021) and there is a big question mark around how various psychometric and behavioural tests treat different groups – more research on this is required.

Sloane et al. (2022) discuss how these tools are promoted as if they are stable, reliable, objective insights into applicants' suitability for a role, but more research is needed to trace whether these tools actually reveal aptitude, future performance, and cultural fit – we need to use socio-technical audits to look into the underlying data, assumptions, and epistemological roots of the systems e.g. the claims to knowledge that the system is making. By 'epistemological roots', they are "referring to the claims to knowledge that the system is making – specifically, that there is a way to 'know' the interior emotional state of a subject based on externally discernible attributes like facial expression, pupil dilation, or other physiological characteristics" (Sloane et al., 2022, p. 3). Similarly, Rhea

et al. (2022) frame an audit methodology around testing the underlying assumptions made by the vendors of the algorithmic personality tests themselves.

In addition, there have been similar concerns raised with automated interview technologies. The ability of pre-recorded video interviews to capture behaviours and soft skills does raise questions of scientific basis. Many have queried whether people's facial expressions, voices, language, and appearance can indicate their job competence (Barrett et al., 2019) and some have claimed that although these systems assume objectivity, they are pseudoscientific (Vassilopoulou et al., 2022) and that any system which claims to parse human emotion should be banned because the claim that traits and skills can be gleaned from facial expressions and body movements is a pseudoscience akin to the practice of phrenology (Ajunwa, 2021).[26] Some have noted that it is unlikely to be able to reliably infer emotions from facial expressions, for example cross-culturally (Raghavan et al., 2020).

Moreover, it has been pointed out that they might misinterpret facial expressions of people with disabilities such as cleft lip or palate, achondroplasia, Down's syndrome, Parkinson's disease, for example, because people with these disabilities are often omitted from training data sets (Tilmes, 2022). Crawford et al. (2019) argue that the claims of affect recognition to be able to read our inner emotions by interpreting physiological data like micro-expressions and tones of voice wreaks of discredited race science and physiognomy, which has been used to claim that biological differences justify social inequalities. They argue that in this context the use of affect recognition should be banned

[26] There have been moves to adopt this standpoint within the law. For example, in June 2023, the European Parliament adopted its negotiating position on the Act, which included a ban on emotion recognition systems in the workplace (European Parliament, 2023).

by regulators. In contrast, Zuloaga (2020) the Chief Data Scientist at HireVue, discussed (before they supposedly discontinued the facial recognition aspect of their technology) how non-verbal communication information gathered from interview assessments of candidates adds a predictive value to certain contexts, arguing that scholars such as Paul Eckman that there is a small set of basic human emotions that can be accurately interpreted cross-culturally, and that they have a diverse training set.

While only three of the vendors I studied sold interview systems, only one of those used automated video technology. Moreover, they didn't use facial recognition technology, but only speech recognition to provide a transcript and score tone, speed, and so on. However, this also raises concern about the implications of accent, speech impediments, and so on. For example, Tilmes (2022) also argues that speech recognition AI tools also struggle to process the words of people with deaf affect, speech impairments, and so on. More research is needed into this area to understand the actual, rather than just the theoretical, implications.

Another element of these tests - which might explain my high score for electrician – is the fact that many of the tests I took were either self-report or multiple choice. Even if I had no idea about the answer, I had a non-zero chance of selecting the right answer. Again, this raised questions for me about accuracy. In my journal from my free assessment with one company, I wrote: "I was also very struck by the fact that self-reporting also means that's how I see myself, but perhaps not how others might find me in the workplace". More research on the impacts of self-report and multiple choice in these settings are needed, particularly in relation to gender. For example, research has already shown that there is a gender gap in self-promotion, where women tend to self-promote less than men

(Exley & Kessler, 2022; Peng et al., 2022; Tradenta et al., 2019). While some of these systems don't use necessarily 'negative' self-description terms, there might still be a difference in terms of how women choose to self-report.

Overall, the structural frame, I argue, is vulnerable to 'invisible out-of-frame' activity, which is unseen and uncertain to researchers, specifically the 'underlying science'. This applies to how skills and behaviours are chosen and how they are measured. While vendors present skills and behaviours as metrics which do not have to be influenced by structural factors such as education, it is unclear what they are measuring and how. These invisible elements encompass crucial information to understand the function of the frame in design, and therefore exhibit a key vulnerability of the frame.

5.5 Discussion

AI dangers are often discussed in terms of reproducing structural inequalities (Crawford et al., 2019; Zajko, 2022) which cause continued acts of oppression and exclusion (Noble, 2018). Inequalities in our society can shape the data which feed these systems, and this means that the algorithm might replicate similar outcomes.

This chapter contributes to literature on the intersection of AI systems and the reproduction of inequalities. It does this by providing the second of three parallel answers to my research question, outlining the structural frame. In exploring this frame, I have showed how structural injustices are intrinsically linked with gender bias in hiring by some of the vendors, and illustrated how vendors aim to address its impact in system design.

Naturally, this chapter follows from the previous one. These structural biases are a manifestation of the historical build-up of intrapersonal biases which have become engrained in our socio-economic structures, and reciprocally, these structures continue to engrain human perceptual biases in societal structures.

To answer sub-RQ#1 in relation to this frame and explore how AI-rec-tech vendors conceptualise gender bias in hiring, I showed that some of the vendors understand gender bias in relation to structural injustice. I defined structural injustice as the more limited opportunity for upward social mobility resulting from inequalities in power and resources within socio-economic structures. This used Young's (2011) work on structural injustice, whose seminal scholarship chimes with what the vendors said in relation to structural injustices, their focus being largely on inequalities in education and opportunity for social mobility.

Not all the vendors spoke about these issues in relation to bias. However, for the vendors who did (perhaps five companies out of the eight), it was often seen as a fundamental part of their mission. They wanted to play a role in providing opportunity to those who hadn't had access to education or who didn't have the credentials to reflect their skills. In the cases where this did arise, I observed that this was often an issue of personal importance to those in leadership roles. In some cases, they had experienced these inequalities themselves, or seen others experience them, as we saw with the boy in the story right at the beginning of the chapter. These issues were often raised passionately. At times, it was seen to be an aspect of the company's purpose; these systems could be enablers of social change.

Throughout this chapter, I showed that the conceptualisation of gender bias in relation to structural injustice is well-founded. I demonstrate that, in the context of the US and UK, to which this chapter is most relevant, inequalities in schooling, socialisation, and wealth abound. Therefore, again, we see strength in the vendors' conceptualisation of gender bias; this understanding is vital for us to consider within the context of hiring and more broadly.

I then moved on to answer sub-RQ#2: how do AI-rec-tech vendors attempt to operationalise addressing gender bias in their system design? I did this in two stages.

In the first stage, I outlined two ways in which vendors attempt to account for structural injustices through the design of their systems: question neutralisation and measuring skills/behaviours instead of qualifications or credentials. Vendors – as I showed - did not claim to be able to impact the root cause of this issue. Indeed, if they did claim the ability to fix this issue through technology, they would meet with the criticism of many who warn against technological solutionism (Morozov, 2014; Selbst et al., 2019) and those who argue that there are "some problems that cannot be fixed by a technical solution" (West et al., 2019, p. 18). However, vendors do see their systems to be helping, in some ways, to account for these structural injustices and therefore improving the implications of these inequalities, even if in a small way.

In the second stage, I analysed these operationalisations. I saw question neutralisation to be a worthwhile technique for increasing the chances of more people understanding the assessments and being able to exhibit their abilities. However, when analysing

skill/behavioural measurement, I used Goffman's notion of 'out-of-frame activity' (Goffman, 1974) as the key to highlight its vulnerabilities. This highlighted factors which sit out of focus for the vendors, but which might have consequential implications for the function of the frame.

I argued that there are two 'out-of-frame' vulnerabilities which could disrupt the frame in design, specifically the measurement of skills/behaviours.

The first of these was 'out-of-frame permeability': realities out-of-frame, or out of focus, which might permeate and therefore impact the frame in design. This included how sources of structural injustices might affect seemingly more neutral measures such as soft skills and behaviours, and how measuring skills directly will likely still favour those more privileged candidates who have had more resources and investment throughout their education. The second was 'invisible out-of-frame' factors: crucial factors of the design which are invisible to the researcher, but which impact the function of the system. While vendors present skills and behaviour measurement as metrics which are not largely influenced by structural factors – such as education – it is not visible what they are measuring and how. Therefore, it is difficult to know the implications of these measurements. I demonstrate that the two 'out-of-frame' vulnerabilities – permeability and invisibility – show that there are ways in which these systems are still impacted by structural injustices.

Throughout the chapter, I draw parallels with the vendors' understanding of their systems and narratives of the American dream. The American dream advocates that hard work, determination, and innate talent bring success (Hochschild & Scovronick, 2003; Jillson,

2016; Taussig, 2021). Similarly, vendors assurance that their measurement of skills and behaviours can "remove barriers to upward social mobility" [Company Green CEO] because "if you're good, you get recognised, you'll make a lot of money, you'll change your life" [Tigran Sloyan, CEO and co-founder of Code Signal], resound with these narratives.

I recognise that the design choice to directly measure skill, instead of relying on credentials or qualifications, does mean that individuals who have developed skills in the face of inequality out of innate talent, independently driven training, or self-determination, will be able to demonstrate their abilities. Perhaps it will mean they even get the job over someone who has those qualifications. In this case, the boy from the story at the beginning of the chapter might have got the employment he clearly deserved.

However, there are parallels between the vulnerabilities of the structural frame and those of the American dream. Both put the focus on individuals rather than structures. With the American dream, this has led to prejudice and blame against those who don't manage to better their lot (Hoyt et al., 2021). As Young points out, when poverty is discussed, it is often argued that individuals need to take responsibility for their lives, and this assumes that poor people can improve their lot by effort and will. However, we must recognise the crucial role of structures in producing and maintaining injustices (Young, 2011; also see Burke, 2012). While some can battle and win to reach their dreams in the face of injustice – poor education and a lack of opportunity – most are constrained by these structures (Goyette, 2017; Hoyt et al., 2021; Mountford Zimdars, 2017; Putnam, 2015; Reardon et al., 2018; Rivera, 2016; Tuñón & Martínez, 2021). While these systems might change outcomes for the few - and maybe the vendors would agree with this statement - they

likely won't for the many. There needs to be more acknowledgement of the impact that structural injustice has on these systems, and more focus on changing the structures themselves.

These vulnerabilities, therefore, highlight two important points regarding the interaction of the design of these AI-rec-tech systems with structural injustices.

Firstly, the 'out-of-frame permeability' highlights that these systems continue to interact with a wider structural ecosystem of social inequality; it is likely that these structures continue to affect the outcomes of assessments. The fact that these systems can never operate in isolation illustrates the importance for reform of these structures. While educational policy alone cannot break the link between inequality of condition and opportunity (Bukodi & Goldthorpe, 2019) there is still the need for developing policy which equalise opportunities for all children and adults to have access to good teaching, resources, and a thorough and wide curriculum (Darling-Hammond, 2001; Hochschild & Scovronick, 2003).

Secondly, both vulnerabilities highlight areas where future research is needed. Regarding permeability, more research is needed into how education and socialisation (both within and outside of education) impact soft skills and behaviours. Particularly, this should take place along gendered lines. In terms of invisibility, there needs to be more work done to uncover what datasets systems are being trained on and how high performers are being selected. Similarly, research should investigate how IO psychologists are discerning which skills are needed for jobs and how this is gendered in implication and result. In addition, to explore how skills are being measured, there should be further consideration

of which behavioural or psychometric indicators are being used and what impact this has along demographic lines. Although this has already been written about in terms of the racist historical roots of psychometric tests (C. R. Reynolds et al., 2021) this would be an interesting topic for psychologists to explore in relation to these tests specifically. In other words, what are the 'epistemological roots' of the skills and behaviours being measured (Sloane et al., 2022). Moreover, what are the impacts of having these tests as self-assessment or multiple choice? And separately, how are automated interview technologies affecting the chances of people with different protected characteristics? While there has been much theoretical work on the ethical problems with this technology (Crawford et al., 2019; Tilmes, 2022) it would be useful to have a deeper dive into this with vendors and clients.

Concerning the social shaping of technology (MacKenzie & Wajcman, 1999), this frame reflects the negative societal reality of structural injustices regarding education and opportunity – again – through attempting to counter them. In addition, vendors discussed how they see differences in how questions are answered according to demographic group. Again, this reflects gender socialisation, and shapes the technology built by these vendors, who then attempt to account for these differences. Moreover, this frame reflects the social ideology, in my opinion, of the American dream.

Regarding the other stream of this bi-directional relationship between society and technology – again, despite attempts to account for these negative realities through question neutralisation and measuring skills/behaviours instead of qualifications, there is no certainty that this is equalising the playing field in terms of diversity in hiring. Of course, this design will not be changing the structural injustices in education and society

which these vendors are attempting to account for but are only capable of potentially addressing the aftermath of these inequalities.

5.6 Conclusion

In this chapter, I presented the second of three answers to my overarching research question by outlining and exploring the structural frame. We need technology companies to be thinking about how structural injustices interact with their systems, and it is encouraging to see these vendors doing so. This is a crucial conceptualisation of gender bias in the context of hiring.

Vendors operationalised addressing this bias through question neutralisation and assessing skills and behaviours. Through discussing two ways in which this frame is vulnerable to 'out-of-frame' activity, I have highlighted that there are ways in which structural injustices are likely to still impact the outcomes of these assessments, and that future research is needed to get a better grasp of the nature of this impact. In the next chapter, I move on to the 'statistical frame'.

CHAPTER 6: THE STATISTICAL FRAME

Note: this chapter accounts for changes in law and regulation up to February 2024 when the final draft of this chapter was prepared.

6.1 Introduction: framing gender bias as statistical

The BBC documentary *Computer Says No* aired in March 2022 in the UK. One of the storylines covered three make-up artists who had worked at MAC, a subsidiary of Estée Lauder. Due to company redundancies, these three women had been told to re-apply for their jobs via a recorded video interview. This was conducted using software developed by HireVue - an AI-rec-tech company based in Utah, US - which analysed their pre-recorded interviews.

After these interviews, all three women lost their jobs, and all of them were confused about why. Lizzie, one of the artists, said in the documentary: “the questions were a bit odd […] like, how to do a dark smoky eye, I found like a really bizarre question, especially when you can’t physically show someone how to do it”. Onieka, another of the women, said that she was baffled because her sales figures and past performance had been really good: “my track record was like, gleaming, essentially […] so this interview definitely raised some alarm bells to me” (Henry, 2022). The three women brought an unfair dismissal and discrimination charge against Estée Lauder, which they subsequently won (Bargh, 2022).

This is not the first time that HireVue's system has been caught under the legal spotlight. In 2019, the Electronic Privacy Information Center (EPIC) urged the Federal Trade Commission (FTC) to investigate HireVue's business practices, saying that its face-scanning technology threatened candidate's privacy rights (Harwell, 2019b). Subsequently, HireVue supposedly discontinued the facial analysis component from its assessment in March 2020 (Maurer, 2021). But in 2022, a US resident Kristen Deyeler filed a class action complaint in Illinois state court against HireVue for a job she applied for in 2019. Deyeler alleged violation of the Biometric Information Privacy Act (BIPA), claiming the company illegally collected facial data for analysis without written permission and accusing them of profiting from her facial biometric identifiers (Spanjaart, 2022).

Similarly, in September 2023, the Equal Employment Opportunity Commission (EEOC) brought forward a case against iTutorGroup for discrimination using an automated hiring system and settled out of court for $365,000. They had hired tutors based in the US but had programmed the software so that it automatically reject female applicants aged 55 or older and male applicant aged 60 or over, which resulted in the rejection of more than 200 qualified applicants based on age (EEOC, 2024c).

These cases exhibit just some of the legal action that has been taken against employers and AI-rec-tech vendors in relation to hiring discrimination and privacy matters. But how does law and regulation fit into the conversation around gender bias and AI-rec-tech design? As I found out, law and regulation were integral to vendors' framing of gender bias.

This is the third and final chapter summarising and analysing my empirical findings. In this chapter, I discuss the last of three concurrent frames to answer my main research question: how do AI-rec-tech vendors frame gender bias in hiring? I argue that the third way in which vendors frame gender bias in hiring is 'statistical' and, therefore, signalled by numerical outcome disparity between different demographic groups. As I show, this interpretation is rooted in anti-discrimination law.[27]

The law is not my area of expertise. However, during the early stages of my research, I repeatedly encountered references to legislation and regulation while discussing bias with vendors. Given this early finding, I sought to interview law and regulation experts. In contrast to clients, who I was unable to access, regulators[28] were much more forthcoming. Perhaps this was because, in comparison to clients, there was little reputational risk. Ultimately, due to their existing knowledge of how AI-rec-tech worked and intersected with the law, these regulators were important for me to speak with. Due to their existing knowledge, they were able to directly comment on the practices of vendors without needing more information from vendors themselves, something which might not have been possible if speaking with psychologists in relation to issues of 'underlying science' discussed in Chapter 5, for example.

While I still use vendor interviews as the foundation for my analysis regarding how gender bias is framed, in this chapter I also bring in interviews with regulators to help understand the vendors' views and practices in this complex area. In addition, as with the

[27] Similarly to the previous chapter, this chapter discusses bias in relation to all protected characteristics, but with a focus on gender.
[28] Henceforth, I use 'regulators' as a catch-all term to refer to the lawyers or regulators.

other chapters, I weave in data from conferences, websites, and reports, along with relevant legislation and regulatory documents, to expand upon my interview findings.

In this chapter, I show how understandings of gender bias were often discussed as outcome-based and statistical. While the foundational underpinning of measuring the statistical impact of systems is crucial, there was, however, little effort from vendors to embed the social understandings of bias discussed in Chapters 4 and 5 within this technical measurement. Rather, the vendors borrowed numerical guidance from anti-discrimination law intended for employers, and this forms the basis of their third frame.

I analyse the vulnerabilities of this frame using Goffman's idea on 'keys' (Goffman, 1974), which highlights the different understandings of existing law by vendors and regulators. This illustrates vulnerabilities in both the vendors' framing of gender bias in this way, and the original framers from whom they have borrowed this specific frame, the regulators.

Before beginning in earnest, two important points must be noted.

Firstly, this chapter is almost solely in relation to US law and regulation and to the five US vendors. While the vast majority of vendors did discuss statistical output regarding demographic groups in relation to bias, the vendors who particularly emphasised this framing of gender bias were mostly based in the US, and therefore referenced US law and regulation.[29] Having said this, some of the vendors based outside of the US also

[29] Laws are passed by a branch of Congress and signed by the president, and they establish requirements or prohibitions. Regulations are published by executive branch agencies to clarify their interpretation of a law and how it will be implemented (USDHHS, 2024).

referenced US law and regulation in relation to bias, likely because they are selling into these territories. I also bring in examples from the EU AI Act, GDPR, and the UK Equality Act 2010 throughout this chapter to illustrate similarities or differences with US law and to broaden the conversation to be relevant to other contexts.

To that end, as can be seen in Table 6 below, most regulators who I interviewed were based in the US.[30] Most of them worked at the EEOC, a bipartisan federal agency responsible for enforcing federal laws around discrimination against a job applicant or an employee because of the person's race, colour, religion, sex (including pregnancy and related conditions, gender identity, and sexual orientation), national origin, age (40 or older), disability, or genetic information (EEOC, 2024e). It typically doesn't investigate discrimination until a complaint is brought to the agency, and it periodically issues guidance and regulation (Bogen & Rieke, 2018). One of the interviewees was also from the Office of Federal Contract Compliance Programs (OFCCP), which oversees the additional legal and regulatory requirements (e.g. to take affirmative action and not to discriminate) which apply to federal contractors, companies, and organisations that provide services or products to a government agency (Bogen & Rieke, 2018).[31] And one of the interviewees worked at the European Commission, a part of the executive of the European Union which operates as a cabinet government, with 27 members of the Commission headed by a President, and includes an administrative body of about 32,000 European civil servants (Commission, 2024a).

[30] This table expands on Table 5 in Chapter 3, providing more information on the role and expertise of these regulators.

[31] In addition to these two federal bodies, AI-rec-tech would fall under the purview of relevant state specific laws and the FTC, a federal agency which deals with consumer protection and competition jurisdiction, pursues law enforcement, and develops policy (FTC, 2024). I tried to reach out to the FTC for an interview but had no response. As Ajunwa (2023b) points out, the FTC has yet to release specific guidelines on how purveyors of AI-rec-tech systems ought to advertise their wares. This might be part of the reason why they didn't respond to my request(s).

Pseudonym/Name[32]	Institution/Government Body
AI Regulation Expert at European Commission *Role and expertise:* This expert contributed to drawing up the European Commission's proposal for the EU AI Act.	European Commission [EU]
Keith E. Sonderling *(Commissioner)* *Role and expertise:* Keith is one of five presidentially appointed members of the EEOC. Commissioners develop and approve policies, issue charges of discrimination where appropriate, and authorise the filing of suits. Keith has published numerous articles on the benefits and harms of using AI-based technology in the workplace and specifically in hiring (EEOC, 2024f, 2024d).	Equal Employment Opportunity Commission (EEOC) [US]
Senior Staff at OFCCP	Office of Federal Contract Compliance Programs

[32] All law and regulation experts indicated in their ethics form whether they wanted to be named or anonymised.

Role and expertise: This expert is a litigator and policy strategist who has expertise on the intersection of equality and technology.	(OFCCP), US Department of Labor [US]
Bradford J. Kelley *(Former Chief Council to Commissioner Sonderling)* *Former role and expertise:* At the time of interview, Bradford provided legal and policy advice on federal employment anti-discrimination laws to Commissioner Sonderling. At the EEOC, he worked on cases, policies, and regulations under all the statutes enforced by the Commission. He was also a key advisor regarding the EEOC's Initiative on AI and Algorithmic Fairness.[33]	(Formerly) EEOC [US]
Brandi Meredith[34] *(Senior Attorney Advisor to Chair Charlotte A. Burrows)* *Role and expertise:* Brandi currently represents the US government in employment discrimination cases and	EEOC [US]

[33] This was launched in 2021 by Chair Charlotte A Burrows. It was designed so that the EEOC could examine more closely how existing and developing technologies change the way employment decisions are made in the hope that it would guide employers, employees, job applicants, and vendors to ensure that these technologies are used fairly and consistently with federal equal employment opportunity laws (EEOC, 2024a).

[34] Brandi Meredith and Eunice Ikene were interviewed together, in a 2-2-1 interview with me, as noted in Chapter 3.

provides legal and strategic advice to EEOC Chair Charlotte A. Burrows. She also serves as co-advisor on AI and Algorithmic Fairness Initiative.	
Eunice Ikene *(Deputy Chief of Staff to the Chair)* *Role and expertise:* Eunice is deputy chief of staff to Chair Charlotte A. Burrows, supporting the agencies work and particularly the AI and Algorithmic Fairness Initiative.	EEOC [US]

Table 6: Law and regulation experts: role and expertise

Secondly, I must note that this chapter differs slightly from the previous two. While the previous chapters considered bias in terms of design features experienced by the candidate or employer, this chapter focuses on the design of impact assessments. These are invisible to users but influence how the system functions under the hood.

6.1.1 Chapter outline

First, I answer sub-RQ#1 for this frame. I describe how vendor companies conceptualise gender bias through legal terms, which encompasses the statistical idea of the 4/5th rule. This rule is outlined in the Uniform Guidelines on Employee Selection Procedures 1978 (henceforth Uniform Guidelines) which, in turn, are rooted in Title VII of the Civil Rights Act 1964. Here, I argue that this frame exhibits a statistical conceptualisation of gender

bias in hiring, specifically equating gender bias with adverse impact, and lack of gender bias with demographic parity. This, I therefore show, is a borrowed frame.

Second, I answer sub-RQ#2 for this frame. I demonstrate that this statistical framing is operationalised through designing impact assessments which calculate ratios between different demographic groups. Impact assessments are currently self-imposed and self-regulated by vendors. There is variation in how impact assessments are conducted and uncertainty around the details, but I outline that, generally, impact assessments are conducted periodically on accumulated client data.

In this section, I also explore two ways in which impact assessments might not capture the full picture through measuring adverse impact statistically: firstly, there might be a lack of data on some (potentially more vulnerable) groups which could skew the results, and secondly, crucial details of how impact assessments are performed remain uncertain (or, as I argue, 'invisible out-of-frame'). I also point out that performing these assessments on accumulated client data is de-contextualised and, therefore, doesn't reflect actual hiring decisions. The effect these impact assessments have on these systems when used in context is therefore uncertain. However, I note that while this frame might be vulnerable in some ways, at least the vendors have a foundational underpinning of framing bias in relation to statistics; it is certainly important to reflect on the outcome of these systems.

Thirdly, I use Goffman's theory on 'keys' (Goffman, 1974) to highlight two further vulnerabilities of this frame. These vulnerabilities emerge from showing how and where views of the vendors and regulators diverge. I therefore discuss two different sets of

'keys'. The compliance 'keys' highlight the various ways that vendors and regulators interpret compliance with existing anti-discrimination law. This highlights the de-contextualisation vulnerability: the impacted function of the frame when the operationalisation is intended for a different context. The sufficiency 'keys' show the different ways that vendors and regulators interpret the sufficiency of anti-discrimination law in relation to bias. This highlights the loophole vulnerability, which applies to the original framers – the regulators – from whom the vendors have inherited this frame: there is a loophole in existing anti-discrimination law which impacts the function of the frame.

Fourthly, in the discussion, I summarise the chapter in relation to my research questions. In addition, I make a case for the value of using 'keys' to shed light on the strengths and vulnerabilities of this frame. Moreover, I outline how these vulnerabilities raise several unanswered questions which will be necessary to address in future research.

6.2 Bias as adverse impact: taking root in Title VII of the Civil Rights Act 1964

In this section, I answer sub-RQ#1 for this frame: how do AI-rec-tech vendors conceptualise gender bias in hiring?

For vendors based in the US, along with others, conceptualisations of gender bias were often related to statistical outcome. For example, Frida Polli, CEO of Pymetrics told me:

> "If I had to choose, I would choose outcome reporting every day, all day. Doesn't mean we don't need to know stuff about explainability, what are you putting into your algorithms, all of that stuff is also critical, but I think if we're not focused on outcomes, we're missing an important piece."

Specifically, this statistical outcome referenced the 4/5th rule. The 4/5th rule is a statistical measurement which discerns 'adverse impact' (or disparate impact; I use them interchangeably) between demographic groups in the outcome or scores of hiring decisions. It is one way to legally challenge a hiring decision on the grounds of discrimination.[35]

The 4/5th rule is a ratio of the proportion of successful pass rates (or selection rates) of the highest passing demographic category versus another demographic group, meaning the ratio of comparing the two should not be less than 0.8. Therefore, the success of any race, sex, or ethnic group as compared to the most successful group, for example male and female, should be no less than 0.8.

In the Uniform Guidelines §1607.4 'Adverse impact and the 'four-fifths' rule', it reads:

> "A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact." (EEOC, 1978)

As I understand it, this cannot be used intersectionally, e.g. to compare white females with Black males, or Hispanic males with Black females. It must be the majority sex, race, or ethnic group in comparison to another sex, race, or ethnic group. This draws attention to the fact that regulators, and vendor companies, often perceive protected characteristics separately rather than intersectionally. As shown in Chapter 5, there are, however, important intersections to consider regarding the design and effects of AI-rec-

[35] In the Uniform Guidelines, §1607.16 'Definitions', adverse impact is defined as "a substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group." It is one element of measuring the effect an employment practice has on a class protected by Title VII, and it is often used interchangeably with disparate impact (EEOC, 1978).

tech systems, such as how gender, race, and class intersect to impact skill development and subject selection in educational settings.

This rule is embedded within anti-discrimination law in the US. It is outlined in the Uniform Guidelines on Employee Selection Procedures (henceforth: Uniform Guidelines) which were released by the EEOC in 1978. These provide guidance about what types of employer conduct is permissible in assessing job applicants (Andrews & Bucher, 2022) and are a federal regulation rooted in Title VII of the Civil Rights Act 1964, which forms the basis of regulatory oversight regarding discrimination in employment and prohibits discrimination with respect to protected characteristics (Raghavan et al., 2020).

Adverse impact and the 4/5th rule are spelt out by Alex Vaughan, Former CSO of Pymetrics in relation to defining gender bias:

> "So, the way that the Equal Employment Opportunity Commission defines bias in this context […] is um, the ratio of pass rates […] so just to define what that is, like, if there are 100 men and 100 women who apply to a job, and let's say that 20 people end up getting hired. If 12 men get hired and 8 women get hired. Um then that's 12 out of, ur, 100 men got hired so 12% and 8% of women got hired. And then the ratio of those two is 8/12, which is 2/3, which is .66. […] And that ratio ideally should be perfectly equal – 1 – um, but, you know, there's statistical noise in the world, you can't control who applies, maybe, you know, one day more qualified women apply, one day more qualified men apply, so there's a range of outcomes um that follows a heuristic that basically says that any ratio above .8 or 4/5 is mostly acceptable and won't trigger, um, a tonne of scrutiny."

Here, the legal concept of the 4/5th rule is seen to constitute the definition of gender bias by the vendor; gender bias is understood in relation to adverse impact.

A statistical definition of bias in terms of the 4/5th rule was, indeed, how many of the vendors conceptualised bias and they see this to be mandated by these government bodies:

> "So, there's two way we adjust gender bias […] make sure it's technically validated and, and um, you know, passes what in the US they call the adverse impact study. Whether it's handicaps, demographics, white, black, Hispanic, female, you know, gender non-specific, that the results are always gunna be gender neutral." [Company Red CEO]

> "Title VII makes it really important that we are demographically neutral […] And what the EEOC and the OFCCP care about is the fraction of people who pass and comparing impact ratios across demographics." [Alex Vaughan, Former CSO of Pymetrics]

This is echoed in one of the Pymetrics whitepapers, which says: "In the US, the rule for fair hiring, legally defined as the 4/5ths rule, holds that relative to the highest passing group, other groups should be passing at the 80% rate." (Pymetrics, n.d.-b)[36]

Essentially, this is a measurement of statistical parity (or near parity) between two demographic groups. This has also been referred to as 'demographic parity' – equal average probability across two groups (IFOW, 2020), 'outcome error parity' - where protected groups receive equal proportions of positive outcomes or equal proportion of errors (Rovatsos et al., 2019), or 'classification parity' - where there are equal passing rates between defined groups (Sanchez-Monedero et al., 2020). It is statistical measures such as this which can lead vendors and academics alike to claim that these systems result in minimal to no adverse impact (Mondragon, 2023), which, to some, means no bias.

This frame exhibits a statistical conceptualisation of gender bias in hiring, specifically as adverse impact, in line with the 4/5th rule. This frame is therefore a borrowed frame[37]

[36] These systems are not only subject to Title VII and the Uniform Guidelines; this is just one tool that legislators can use. They are also subject, for example, to Title I of the Americans with Disabilities Act (ADA) of 1990, Section 503 of the Rehabilitation Act of 1973, and the Age Discrimination in Employment Act (ADEA) (Sonderling et al., 2022).

[37] A borrowed frame is one which has its details specified by another party, and which has been adopted and applied in a different context.

from the statistical test laid out in Uniform Guidelines regulation. It is a frame, originally, formed by regulators.

Crucially, adverse (or disparate) impact should be distinguished from 'disparate treatment'. Disparate treatment is another ground for discrimination as defined by the Uniform Guidelines which focuses on the failure to provide applicants from different demographic background, or categories produced under Title VII, with equal treatment and opportunity (Raghavan et al., 2020). As Bradford J Kelley (EEOC) outlined to me, disparate treatment is "direct or intentional", whereas disparate impact is "unintentional discrimination that usually arises, you know, based on a neutral application that disproportionately affects one group". Therefore, disparate treatment claims arise when an employer intentionally discriminates on the basis of protected characteristics, whereas disparate impact applies to an employer using a supposedly neutral practice but it has discriminatory effects based on protected characteristics (Raub, 2018). Ajunwa (2020) discusses how AI can present opportunities for both disparate treatment and disparate impact. This is also discussed by Sonderling et al. (2022), who outline that employers using AI systems could be liable under Title VII because of either. It might also be due to unequal opportunity to complete the test effectively (disparate treatment) that we see differences in statistical outcome (disparate impact).

Here, however, the vendors tended to focus on the statistical aspect in relation to bias, and therefore usually referred to disparate impact. Of course, thinking about disparate treatment was part of thinking about these statistics, and arguably the design measures in Chapters 4 and 5 were attempts to level the playing field for all applicants. Overall, then, the US vendors conceptualised gender bias in statistical terms regarding outcome,

specifically in relation to the 4/5th rule which seeks demographic parity in outcome ratios. It is crucial that vendors consider how these systems might impact different demographic groups, and how that might manifest in quantitative terms. Therefore, the foundation underpinning this conceptualisation is important. However, as I go on to show, the fact this is a borrowed frame means this might not be the most suitable statistical measurement for this context. This conceptualisation, while only briefly outlined here, continues to unfold throughout the following sections.

6.3 Measuring adverse impact

I now move on to answer sub-RQ#2 in relation to this frame: how do AI-rec-tech vendors attempt to operationalise addressing gender bias in their system design?

Overall, companies addressed statistical bias – in the form of adverse impact - through impact assessments (or adverse impact studies, as some of them referred to it). As the Company Red Partner said to me:

> "Well, the only way that you say see how this, whether this gender bias, or any bias is addressed, is an adverse impact study."

Impact assessments and audits are two terms commonly used, but not consistently distinguished. Audits are thought to inspect algorithmic bias or regulatory compliance, and impact assessments are seen to concern risk before usage or impact evaluation after use (Ada Lovelace Institute, 2020). Some might argue that the impact assessments which vendors conduct are a form of audit (Wilson et al., 2021), especially if they are conducted by a third party. While audits have been thought of as mostly relying on statistical definitions of bias and/or fairness (IFOW, 2020), there is a move towards encouraging

audits to assess socio-technical factors of systems too (Rhea et al., 2022; Sloane et al., 2022; Vecchione et al., 2021). Impact assessments as vendors discuss them are technical, but there is a need to perform socio-technical audits alongside. By socio-technical, I mean systems which consider both social and technical influences and impacts, aiming to consider both in the design of the system (Baxter & Sommerville, 2011; Mumford, 2006; Murray & Trist, 2016; Selbst et al., 2019).

Although not a design feature which can be experienced on the front- or back end of the system, these impact assessments are part of the vendors' system design and, indeed, they reflect overall system design too. They're not mandatory by legal standards, and they're not dictated by any external requirements or specific impact assessment templates (Kak & West, 2023). Currently, these adverse impact assessments are self-imposed and self-regulated, and they feed back into the design of algorithms or question design. However, I must caveat this with the fact that while most vendors internally design and conduct these tests, some vendors have them carried out by independent, external companies. Regardless, this still influences the internal function and design of the technology, and so chimes with my definition of AI design provided in Chapter 1.

It is unclear exactly how vendors (or external auditors) conduct these impact assessments. Certainly, there is variation in how it is done, and more research is needed to explore various elements of these assessments. However, there are some generalisable trends. One common factor, as we see across the quotes below, is that these tests aim to **measure parity** between demographic groups and thereby aim to abide by the 4/5th rule. For example, Frida Polli, CEO of Pymetrics, spoke about how all the algorithms they release

are pre-tested to make sure they fall above the disparate impact threshold (by which they mean the 4/5th threshold) and, if they don't, they won't release them.

Further to this, from my own data, I saw the following trends:

- **Timing:**

Vendors tend to carry out these impact assessments either at periodic intervals, for example, once a year or once every three years, or when there is a change in their test design.

The Company Red Partner, for example, told me that they get these impact assessments carried out periodically by an independent statistical firm. And the Company Red CEO told me that: "we've specified and done every three years for the last like thirty years, ok? So, check on that". The Company Red Partner also told me they might carry out these impact assessments if there are any changes in the test, because this effectively nullifies the previous impact assessment results. In these cases, they accrue the data (from across all their clients) and when they have enough, they go through the cycle of obtaining an impact assessment again.

In the case of Company Yellow, they tell me they do these impact assessments annually:

> "I do wanna point out to you that annually we review those statistics and based on gender, different genders and we also look for disparate means and standard deviations." [Company Yellow IOP1]

Alex Vaughan, Former CSO of Pymetrics also emphasised that they do this "regularly" and that if they don't meet this threshold (the 4/5th threshold), they will go back and

change elements of design before running the impact assessments again once they have accumulated more data.

- **Data:**

The impact assessments tended to be conducted on accumulated client data, although the nature of this data is unclear. Therefore, the data is not contextualised or in relation to each client and their hiring decisions, but it is data accrued from across multiple clients. As seen in the previous section, the Company Red Partner noted that they accumulate data across all their clients. The Company Red CTO also told me that they run these tests on 'respondents' data, e.g. people who applied for jobs, collected across multiple clients:

> "And um, even in the last validation that we did which was in 2021, you know, we sampled ur 10,000 of our respondents, ur, random sample, and then look at the scoring of our, um, sample by different things. Could be gender, could be race, could be location, to make sure that there's no bias inherent in the assessment." [Company Red CTO]

Frida Polli, CEO of Pymetrics also told me that they aggregate statistics from client data over a 12-month period. If this data produces statistics that are well above the disparate impact threshold, then they think that's success regarding bias reduction. They want to move as close as possible to a ratio of 1 across all groups:

> "If we can continue to produce statistics that are well above the disparate impact threshold, on an aggregate basis, we think that's success." [Frida Polli, CEO of Pymetrics]

Alex Vaughan, Former CSO of Pymetrics also told me about their impact assessments:

> "So, we actually did a meta-analysis […] um, across all of our clients, where we looked at impact ratios, um, in particular for male-female. And these are candidates who are self-identifying as male and female. […] um, and our ratio there, actually depending on the timeline you look at is really, really close to 1 […] And just to say it because I'm proud of it, um, you know, a lot of historical hiring bias in the US has to do with, ur, other demographic comparisons, in

> particular black-white and white-hispanic, um, and those are both in really excellent territory as well. You know above .93, .95, .97 for a lot of them."

As mentioned earlier, these statistics reflect ratios within a singular set of demographics – for example, males and females, but they do not function intersectionally. This limits our ability to see the impact on the most vulnerable intersectional groups in comparison to the most dominant, such as a ratio of white males and black females, and it also gives the illusion that these protected characteristics function separately, with these statistical goals acting as more of a 'tick box' exercise.

While there was no explanation to me of why the vendors run impact assessments on aggregate data across clients, rather than measuring clients individually, I suspect it is because this data would be for clients to analyse and publish, not the vendor. In addition, as we saw in Chapter 4, in many cases, vendors don't know who the clients go on to hire and therefore likely cannot perform these impact assessments individually.

- **Groups for testing:**

In cases where vendors don't have access to which candidates have been hired by their clients, it is unclear how the vendors dictate which candidates have been most 'successful' and therefore determine which groups to assess. Moreover, it is unclear how they choose which minority groups to assess within the ratios.

As the Company Red CSM told me, this is something which might change, but I was uncertain about how these 'minority' vs. 'majority' groups are dictated:

> "The first thing we look at when we look at bias has to do with simply cutting the data by demographic categories and running t-tests, you know, on the differences,

> you know, between males and females or any other categories, you know. So, we might look at African Americans compared to white people or look at Asian Americans compared to white people […] it's gunna always be the minority group compared to the majority group. Um and those analytics are based on EEOC guidelines. So, the US Equal Employment Opportunity Commission guidelines." [Company Red CSM]

However, it is unclear how the companies have arrived at the conclusion to test these groups specifically, or what was meant by 'minority' and 'majority' groups.

Considering these three trends, in addition to the fact that this approach regards protected characteristics singularly rather than intersectionally, there are two additional ways in which these tests might not capture the full picture.

Firstly, there might be a lack of data on some groups.

Although the vendors are not allowed to use demographic information within the algorithm, as discussed in Chapter 4, they do collect this data to run these impact assessments. However, it's also optional for the candidates to give this information:

> "We try our best, right, to try and get in our data, um, people to disclose basically who they are. But they don't have to, yeah." [Company Red Partner]

> "The demographic assessments are optional, and many people choose not to answer them, or they answer some not others which is… it's an interesting paradox because people don't wanna answer them because it seems unfair, but we can't know it's fair unless people answer them." [Company Blue CSO]

Vendors want people to disclose their demographics so that they can conduct these impact assessments. However, there is no way of knowing whether people from marginalised or vulnerable demographics are less likely to provide their details and how this might affect the results of vendors' impact assessments. Studies in the past have demonstrated that people of colour 'cover' their identity in case it might be used to discriminate against

them, for example through the 'whitening' of resumes (Kang et al., 2016). Additionally, as noted in Chapter 4, some people might not feel there is an option which describes them, particularly if there are not inclusive gender options available. This could result in them leaving the form blank. And in cases where vendors don't offer gender options beyond the binary, these tests certainly won't show the whole picture in terms of the effects on those outside or between the binary. Overall, it is possible that vendors lack the full picture in these instances, perhaps more so with people from marginalised or minority backgrounds. Of course, this raises questions around which candidates are more likely to fill in the demographic details and how this might affect the ratios.

Alex Vaughan, Former CSO of Pymetrics also pointed out that this type of analysis is easier with groups where there are more data:

> "To be blunt, that is almost... that is very rarely relevant for gender bias. And part of the reason for that is that there's less noise in terms of gender bias because, you know, approximately one half of candidates identify as male and approximately half identify as female, and so you have a lot of data about folks."

If there is not enough data on a group, and the adverse impact ratio is lower than 0.8, the Uniform Guidelines provide exceptions where this won't be seen as adverse impact. For example, in §1607.4, it says that: "Greater differences [than the 4/5th rule] in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group". In these cases the EEOC might aim to use evidence over a long period of time or might look at how that selection procedure had an impact when used elsewhere to determine adverse impact (EEOC, 1978).

There are, therefore, ways for employers to get around this statistical requirement, and this means that any demographic groups where there isn't enough data might be excepted from claims of adverse impact, even if their ratio is extremely low. As we saw in Chapter 4, some vendors don't collect data on non-binary participants, perhaps because they don't want to risk the ratios being too disparate - they might be worried about how it would appear to their prospective clients. This is an aspect of the law which should be reviewed, and which I therefore pick up on in Chapter 8.

Secondly, while crucial details of how impact assessments are performed are uncertain, we know that the tests themselves are de-contextualised from specific client decisions.

It is uncertain how vendors are dictating 'success' or 'pass rates' to calculate the ratio, particularly if/when they don't have access to client data on who has been successfully hired. The Company Blue CSO pointed this out, telling me that they cannot tell how their clients are using the system and therefore can't measure adverse impact properly:

> "So, we can't really measure adverse impact directly because we don't, you know, we don't know who they're actually hiring, and we don't even necessarily know all the applicants they get, it's whoever they choose to send the assessments to." [Company Blue CSO]

For other vendors, information about how the most successful group is being determined and, therefore, how the demographic groups are being decided and the ratios calculated for these assessments, is 'invisible out-of-frame' (a frame vulnerability outlined in Chapter 5). This is missing information around impact assessments by vendors which affects how these assessments function and what they reflect. Further investigation of how candidates are labelled 'successful' will make for important future research. For example, if vendors don't know who their clients have hired, what are they measuring as

'success' rates? Is this to do with high scores on their tests? And if so, what does that mean? If the test delivers a profile instead of a linear score, which profile characteristics are they selecting as successful? Or are they selecting minority and majority groups randomly as dictated by societal patterns and trends?

Additionally, calculating these ratios with accumulated client data means that these ratios are de-contextualised and so the ratios likely do not reflect specific instances of hiring. This, therefore, renders these impact assessments somewhat meaningless; they won't show some of the disparities in ratio which occur for their clients. Instead, these impact assessments are likely conducted as an extra precaution by vendors to re-assure their clients.

In the context of the employer's decision-making, things might be quite different. For example, these impact assessments might be conducted by vendors on someone who scores 'high' on the test (whatever the vendor determines that to mean). However, as we saw in Chapter 4, due to the hybrid nature of these systems, it might be that clients choose to ignore these scores in practice.

However, it is crucial to explore outcomes of these systems to infer their impact. Therefore, although this frame might be vulnerable in certain ways, the foundational underpinning of thinking about bias as statistical is a crucial one.

6.4 Different 'keys': the dissonant tunes of vendors and regulators

I have now discussed how the statistical frame is understood and, broadly speaking, how it is operationalised.

In this section, I shed light on two further vulnerabilities of the statistical frame. I do this by discussing the misalignment between the vendors and regulators when it comes to two areas: compliance and sufficiency.

I use Goffman's idea on 'keys' (Goffman, 1974) to illustrate the differences in opinion between vendors and regulators. Goffman's idea of 'keys' ingeniously plays on the analogy of music, particularly the idea of modulation and the transposition of tunes to sound similar but also, altogether, different (Jameson, 1976). 'Keys' essentially represent an understanding and set of conventions which can be systematically transformed into something similar - but altogether not the same – by the activity of 'keying'. Goffman defines 'keys' as "the set of conventions by which a given activity, one already meaningful in terms of some primary framework, is transformed into something patterned on this activity but seen by participants to be something quite else" (Goffman, 1974, pp. 43–44).

In this case the 'keys' surround the law and regulations – a set of conventions which are understood by the regulators in one way and another way by the vendors. For Goffman, 'keying' is one way in which frames are vulnerable. It is a "systematic transformation" whereby something that is "already meaningful in accordance with a schema of interpretation" is transformed (Goffman, 1974, p. 45). This transformation might alter only slightly the activity, but it utterly changes what the participant would say is going on. Therefore, while people can play in the same 'key' and be in harmony, they can also

be in different ‘keys’, causing a dissonance. Ultimately, then, ‘keys’ signify differences in interpretation and understanding.

I show here how the different keys of these two parties highlight vulnerabilities of the statistical frame.

6.4.1 The compliance keys: the de-contextualisation vulnerability

The first set of ‘keys’ which show different understandings of vendors and regulators (and sometimes, although rarely, amidst vendors themselves) is in the realm of compliance.

When it came to discussions of bias and measuring it statistically, many of the vendors claimed that they were EEOC compliant. This seemed to be a way of signalling that their systems reduced or removed statistical bias through their impact assessments and was therefore in line with the 4/5th rule in the Uniform Guidelines.

One website said that their system was “EEOC – adverse impact-compliant science”, another said that their company was “GDPR and EEOC compliant”. During a conference, one vendor asserted that clients should ensure that vendors are EEOC compliant and pass all adverse impact studies before they consider buying it, to ensure that it doesn’t discriminate (‘Using Psychometric to Make Better Hires, More Quickly’, 2022).

These claims might be designed to re-assure clients that vendors are thinking about legal compliance, and to put clients at ease if they’re nervous about bias emerging through use

of AI-rec-tech. The Company Red HOS told me that when their clients are worried about using these systems due to bias, they assure them that they're 'EEOC compliant':

> "A lot of the bigger companies, for probably the reasons why we're talking about here, they are, they're worried about using anything for hiring, just because of bias. Um, you know, and it would take them a while […] we say look, there's no bias,[38] and you know it's, it's EEOC compliant and it doesn't… we have an adverse impact study."

It has been found previously that vendor companies advertise their systems as EEOC compliant. Kelly-Lyth (2021) notes that US-based algorithmic vendors often advertise their compliance with the 4/5th rule. In a study of 18 vendors, looking at publicly available data including their websites, reports, and webinars, Raghavan et al., (2020) also found that three of these vendors explicitly mentioned the 4/5th rule and an additional four advertised compliance or claimed to control adverse impact more generally.

However, in a different 'key', many of the regulators told me that vendors could not be 'EEOC compliant'. As Keith E. Sonderling (EEOC) told me:

> "Right, but nobody's EEOC compliant because we don't give out compliance audits."

Brandi Meredith (EEOC) told me, similarly:

> "[W]e do not certify any products, right [laughs] […] if people are advertising 'we're EEOC compliant' we ask them to take that down because we have no… our agency does not review products… [or] certify them."

Keith also told me that it wasn't just that vendors couldn't be compliant because the EEOC don't issue compliance audits, but also that the law doesn't apply to the vendors. The law only applies to the employer:

[38] As noted in Chapter 4, there is variation in whether participants asserted that their systems reduce or remove bias.

> Keith E. Sonderling: "Every law, every law we enforce here at the EEOC applies to the companies who *use* the artificial intelligence." [emphasis added]
> Collett: "Not the vendor companies?"
> Keith E. Sonderling: "Not the vendor companies."

Therefore, it is not the vendors who can or need to be compliant, because "[u]nder the current legal framework in the United States, employers—not vendors—are generally liable for employment discrimination." (Sonderling et al., 2022, p. 16). Bradford J. Kelley (EEOC) reiterated this to me:

> "[Y]eah, overall, you know the laws really only apply to employers."

This was, indeed, confirmed by my reading of the Uniform Guidelines, which specify in section §1607.1B: "These guidelines incorporate a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of Federal law prohibiting employment practices which discriminate" (EEOC, 1978).

Therefore, these guidelines and anti-discrimination laws do not apply to vendor companies. This means that it doesn't matter how vendors have tested the system or what they have said to the employer about their EEOC compliance, it is still only the employers who are ultimately liable:

> "Under US law […] the employer is liable for any employment decision […] when we come in, we're only looking at you [the employer], and we don't care what the vendor told you, we don't care what reports the vendor shows you that they're compliant, we only care about how it was applied to your workforce. And even if they gave you 100% assurances in their brochures, and studies that this product would not discriminate and do what it's told, that doesn't matter, I'm only coming after you [the employer], and the fines, and the, you know, the federal discrimination and the investigations and lawsuits are going to be against you, it doesn't matter what anyone told you, the only thing that matters is the result and that's the state of the US law." [Keith E. Sonderling, EEOC]

Vendors, therefore, are not required by law to conduct these assessments and are not subject to these guidelines and legislation. Perhaps this also explains why it was so difficult for me to gain access to the employers – because ultimately, they are the ones liable for employment discrimination, and there is more for them to lose if any problematic practices were to be exposed.

Indeed, these statistical measures are designed to measure employment decisions, not aggregated client data. As seen above, when the vendors conduct impact assessments – internally or using external companies - they usually conduct them on aggregated datasets from their clients, meaning these results are de-contextualised. Therefore, they are somewhat irrelevant to fulfilling the requirements of these guidelines for their clients:

> "Of course, they say they're EEOC compliant, but these vendors are using […] tests on the 4/5th rule, based upon you know these aggregate datasets […] but from our perspective, nobody can say that, because to be compliant you [the employer] have to test your own dataset. And when the EEOC comes in and when the federal investigators come in, we only look at the results of it used on your [the employer's] workforce for that exact position. So, anyone who is saying that they've tested the AI programme, unless they've literally gone into that employer, and they've looked at that employer's dataset, and they've looked at that employer's applicant pool […] for that exact position based on that individual job description for that one job being used for AI, and then done a test and shown whether or not there's disparate impact on that test, and if it passed the 4/5ths rule, which is still not even a set in stone rule and the EEOC doesn't even use it all the time, then you could say you're compliant, but until then, and that's not happening." [Keith E. Sonderling, EEOC]

Keith E. Sonderling (EEOC) noted that this is one of the main reasons why the EEOC are trying to raise awareness - so that the employers understand the legal situation, and can ask the right questions to vendors, for example, are you going to test this on my dataset?

However, it is important to acknowledge that not all vendors did claim to be compliant. Therefore, the different 'keys' here aren't just between the vendors and regulators, but

amidst vendors (to be clear, not within companies, but this variation emerged between different companies). For example, one website says they "can help support clients' legal compliance with the EEOC", thereby implying that they themselves were not compliant. This was also mentioned in one of the interviews:

> "So, I know some companies say that they're EEOC compliant, there is no 'compliant'; you can't be EEOC compliant. […] assessments cannot be EEOC compliant, plainly […] It's guidance, it's not standards of compliance." [Company Blue CSO]

Overall, though, regulators exhibit a different 'key' to the vendors. While some vendors say they are EEOC compliant, regulators say this is not possible and that, moreover, this statistical measurement is not intended for use on aggregated datasets by vendors. The 4/5th rule requires the dataset being tested to be that of a specific employer for a specific job position. In addition, the law is not aimed at vendors, but rather, at employers, meaning they cannot be compliant when they're not liable.

This dissonance between 'keys' raises a vulnerability for this frame – the vendors' concept of gender bias is rooted in the 4/5th rule, but this was never intended to be used as a statistical measure in the context of vendors' impact assessments on accumulated client data, with the vendors making decisions around what 'success' means. This frame therefore highlights the **de-contextualisation vulnerability.** This de-contextualisation impacts the function of the frame as this measurement is misplaced, and therefore there are uncertainties about the usefulness of this type of impact assessment by the vendors. It is uncertain whether this makes a difference in practice or what it will show on accumulated client data.

In some ways, though, the vendors are going beyond what is required of them by law. At least they're attempting to measure the output, even if it's, perhaps, not the right measurement to use. Statistical impact on different demographic groups is certainly an important thing to consider in discussions of bias. However, the fact this frame is borrowed means that it hasn't considered thoroughly how to embed the social framings of bias – intrapersonal and structural – into the statistical measurements in the context of periodic impact assessments within vendor companies. Additionally, as discussed above, this is also reflected in the way this statistical goal captures protected characteristics separately rather than intersectionally and, therefore, this conceptualisation of bias does not reflect the holistic picture of how these systems might be impacting the most vulnerable in society.

There are a couple of further questions raised by this de-contextualisation vulnerability shown by these different 'keys' surrounding compliance with anti-discrimination law. While I cannot answer these questions with my own data, they indicate areas for future research.

Firstly, as mentioned, it raises questions about whether these impact assessments of vendors on aggregated client data make a difference in practice to their client's statistical demographic outcomes? Depending on this, it will be important to explore, in future, the most appropriate statistical guidelines for vendors' impact assessments during development.

Secondly, for the vendors who do claim to be 'EEOC compliant', why do they claim this? Perhaps vendors advertise this to signal to their clients that they are considering bias and

discrimination, and to re-assure them. As Raghavan et al. (2020) discuss, adopting the 4/5th rule might cater to employers' aversion to legal risk (Raghavan et al., 2020).

Thirdly, who should be liable for discrimination in hiring using AI-rec-tech? Bradford J. Kelley pointed out that "the reasons the employers are responsible is because they're the ones you know, using that". We saw in Chapter 4 that there are elements of the design which are optional to use, and that clients are given autonomy to make decisions surrounding, for example, the skills and behaviours included in the desirable candidate profile.

There have been moves, however, to ensure that the vendors also bear some responsibility for bias in these systems. For example, the Algorithmic Accountability Act in the US was re-introduced in 2023 by Representative Yvette Clarke. Having previously been introduced in 2019 and 2022, it failed both times to make it out of the 116th and 117th Congress respectively, and we are yet to see whether this new introduction will progress further. This would look to shift some responsibility onto certain AI-rec-tech vendors[39] by requiring them to perform impact assessments on automated decision systems before deployment, and it would require the FTC to create guidelines for assessing the impact of decision processes (*Algorithmic Accountability Act of 2023*, 2023; Mökander et al., 2022). As we have seen, vendors are already doing these impact assessments on a periodic basis. However, this move would formalise the requirement and outline specific processes for certain vendors performing pre-implementation impact assessments.

[39] This would only apply to large vendors who had either a greater than $50mn annual gross turnover, or a $250mn equity value for a 3-year period (*Algorithmic Accountability Act of 2023*, 2023) which, in my opinion, is a limitation of this Act; it should apply to all vendor companies, however large or small.

Moreover, the EU AI Act would bring more responsibility to vendors as well as employers (Sonderling et al., 2022). This risk-based legal framework dealing with AI aims to provide AI developers, deployers and users with clear requirements and obligations regarding specific uses of AI (Commission, 2024b).[40] In the proposal, AI-powered recruitment technology is labelled as high risk, meaning it poses significant risk to fundamental rights, health, or safety (European Commission, 2021).

The EU AI Act is trying to put more liability onto the vendors instead of just on the employers:

> "AI providers, those who are developing and placing those systems on the market […] to make them actually responsible for how those systems are developed, designed, ur, ask them to comply with certain mandatory requirements, properly document everything, make risk assessments, ensure that data quality is very good and the system is checked for biases, properly tested before it's actually used, ur, and especially for, yes, for discrimination, bias examination, and also the risk mitigation measures are really important, they're put in article 9 and 10 of our proposal. Um and then also to properly document the system, give some important transparency, instructions to use to the users, to whom they are selling the systems, so they know the limitations, the capabilities, what kind of oversized measures they have to exercise, and how to monitor the system. So […] [we] ask them to make sure that the system are checked for their conformity before they're placed on the market." [AI policy expert at the European Commission]

As mentioned above, vendors are already testing for statistical bias pre-implementation, but without any specific or tailored guidelines. These guidelines are, in my opinion, desperately needed and desired by vendors themselves and are not yet specific or useful enough for vendors in the current EU AI Act proposal.

[40] In June 2023, the European Parliament adopted its negotiating position on the Act, which included a ban on emotion recognition systems in the workplace (European Parliament, 2023).

The expert from the European Commission also talked about the AI Liability Directive,[41] an initiative complimentary to the EU AI Act, which deals with the question of who would be liable if an employee (or job candidate) thinks they're being discriminated against. This would make it easier to bring a claim against the vendor company in court rather than the employer, who might not have had full transparency about how the system works. Of course, though they also note that the provider is "limited only to the extent of their responsibilities and competencies" and therefore, "if the user has, for example, not used it appropriately, um, it could also be made responsible" [AI policy expert at the European Commission]. Therefore, it does not shift liability solely onto the vendors, but allows for a more balanced picture and makes the regulation more flexible in terms of whether it could be aimed at either the vendor or the employer.

While I agree that it is important to balance liability and responsibility between vendors and employers, there needs to be more specific and comprehensive guidance for vendors, who are, as we will see, open to more specific and stringent regulation.

Overall, the two 'keys' surrounding compliance with anti-discrimination law highlight a vulnerability of this frame: the de-contextualisation vulnerability. While some vendors claim that their systems are EEOC compliant, the regulators, along with a few exceptional vendors, demonstrate that vendors cannot be EEOC compliant because the law doesn't apply to them. Rather, it applies only to employers. The impact assessments which the vendors conduct (or which are conducted by a third party) do not use data from individual clients, and therefore do not count towards compliance measures for the client.

[41] The directive would create a rebuttable 'presumption of causality' to ease the burden of proof for victims to establish damage caused by an AI system and would furthermore give national courts the power to order disclosure of evidence about high-risk AI systems suspected of having caused damage (Madiega, 2023).

Ultimately, therefore, while consideration of statistical outcome is important, the specific conceptualisation and operationalisation of this frame are de-contextualised; the vendors have borrowed a frame from regulation which is not intended for them. This forms a vulnerability, as these measurements are not necessarily apposite for periodic vendor impact assessments and are therefore misplaced. This raises questions about the effectiveness of these impact assessments, along with a clear need for considerations about the best measurements to use in these contexts.

6.4.2 The sufficiency keys: the loophole vulnerability

The second way in which the vendors and regulators differed concerned legal sufficiency: whether the laws and regulations which formed the blueprints for this frame were sufficient. This was not a vulnerability on the side of the vendors, but, rather, on the side of the regulators, from whom this frame was borrowed.

Contrary to common understanding, vendors seemed to be relatively open to more specific and stringent regulation. Perhaps they use the 4/5th rule for want of more appropriate guidance, and to ensure or reassure their clients that they are aware of existing legislation. This shows from the fact that, even though they borrow this frame to form their own understanding and practices, they do, at times, criticise it.

Vendors discussed how the Uniform Guidelines around the 4/5th rule exhibited some deficiencies, meaning it might not always capture adverse impact. On the other side, the regulators spoke about how these laws in themselves are sufficient for tackling

discrimination and bias. Instead, they pointed out that there were other areas which needed more attention, such as enforcement.

Firstly, the vendors' 'key'. As seen above, vendors generally used the Uniform Guidelines and the 4/5th rule to ground their conceptualisation and operationalisation of bias as statistical. Nevertheless, they did recognise that this guidance surrounding federal law can exhibit insufficiencies.

One participant pointed out that while Title VII means that a ratio of 0.8 or above wouldn't trigger investigation, if the ratio did fall below this, during investigation it would be incumbent upon the employer to demonstrate that the bias reflects a valid selection criterion. If this selection criteria are deemed valid, and therefore if no other less discriminatory alternative were possible to the employer, then this might be sufficient to avoid charges. In other words, if the outcome falls below the desired ratio, and if it so happens – as argued by the employer – that the men who applied were simply more qualified for the job on average than the women, and this was linked to job performance, then it is perfectly legal to use an assessment that identifies those top performing men and hire them at a higher rate than women.

Alex Vaughan, Former CSO of Pymetrics explained this further:

> "[T]he way Title VII works […] that is a trigger for investigation, after which, ur, it is incumbent on the employer to demonstrate that even if there was bias, that that bias reflects, um, a valid selection criteria that actually relates to job performance. So, US law believes, for better or worse, um, that if it so happens that men are more qualified for a job on average, then it is perfectly legal to use an assessment that identifies those high performing men, um, and hires them at a higher rate than women. But the important thing is that that determination has to be valid, meaning that it has to be linked to job performance."

This ability to sidestep the 4/5th rule therefore surrounds 'business necessity'. As §1607.11 of the Uniform Guidelines reads: "Thus, the persons who were in the class of persons discriminated against during the period the user followed the discriminatory practices should be allowed the opportunity to qualify under less stringent selection procedures previously followed, unless the user demonstrates that the increased standards are required by business necessity" (EEOC, 1978).

Title VII case law and EEOC guidelines hold that business necessity can be claimed when the outcome of a hiring decision is predictive of future employment, even if the correlations will reflect historical injustices, discrimination against protected groups, or flaws in data (Barocas & Selbst, 2016). In this business necessity defence, employers would have to establish validity of the model and also justify why no alternative practice exists that would serve the employer's business objective equally well while reducing disparate impact (Jones, 2023; Raghavan et al., 2020). If the employer cannot justify how or why the measurement of the trait is job-related, then the business necessity rebuttal will fail (Raub, 2018).

It was clear to me from the way that the vendors raised this in discussion that there was some speculation about the ability to claim business necessity, and how this might play out in practice. It was sometimes seen by the vendors as a loophole in the legislation. For example, one of the Pymetrics whitepapers noted that:

> "A federal regulation loophole states that a hiring tool or methodology can be biased so long as it is 'job-related'. This means that if people who are successful in a role show specific characteristics, and all 'successful employees' are white (for example) due to the history of biased human hiring and promotion practices, then it is almost guaranteed that your job-related talent assessment tools will favor white applicants and disadvantage minorities." (Pymetrics, n.d.-b, p. 5)

This is not the first time that anti-discrimination law has been perceived as insufficient in its capacity to address bias (Ajunwa, 2021, 2023a). Bogen & Rieke (2018) argue that the EEOC should consider new regulations that interpret Title VII in light of predictive hiring tools, and Jones (2023) also argues that the EEOC and Congress should focus enforcement on vendors through new legislation and regulation around auditing and transparency requirements.

In the US, we see that alongside complying to federal laws, there have been more stringent state specific laws – perhaps as a reaction to a perceived insufficiency of federal laws for addressing discrimination arising from AI-rec-tech. Briefly, a few of them are:

- **Illinois: Artificial Intelligence Video Act (2020)**

  Candidates using recorded video interviews must be notified and their consent must be gained beforehand. Employers relying solely on AI analysis of a video interview to hire must report demographic data of candidates who applied, were interviewed, and hired (Artificial Intelligence Video Interview Act, 2020). However, Sonderling et al. (2022) point out that the law fails to define AI, and it is also silent about penalties, enforcement, and whether employers may refuse to consider applicants who refuse to consent.

- **Maryland: Use of Facial Recognition Services – Prohibition Act (2020)**

  This act prohibits employers from using facial recognition technology during pre-employment job interviews unless the applicant consents by signing a specified waiver (Hogan, 2019). Again, there are big gaps for interpretation regarding compliance and enforcement (Sonderling et al., 2022).

- **NYC: Automated Decision Tool Local Law 144 (passed 2021, effective 2023)**

> Employers who are using AI in hiring must notify candidates and inform them of the tool's characteristics. They must submit an annual independent audit to test the tool's disparate impact. They incur a cost of $500 for the first violation and up to $1500 for subsequent violations (DCWP, 2023). However, it does not mention how findings should be utilised, and the passing criteria for / nature of the audit (Fuchs, 2023; Sloane et al., 2022).

Some vendors have welcomed these state specific laws. For example, the founders of HiredScore, which tries to highlight promising candidates based on resumes and other data sources, and Pymetrics both supported the NYC bill (Simonite, 2021).

The vendors' 'key' points to the **loophole vulnerability** of the original frame. The business necessity clause – or loophole - within the Uniform Guidelines does compromise the integrity of the frame, just as the vendors point out. This was seen earlier, when Keith told me that "the 4/5ths rule, which is still not even a set in stone rule and the EEOC doesn't even use it all the time". The ratio is possible to sidestep, and this somewhat undermines the point of having an arbitrary measurement in place to operationalise the statistical understanding of bias altogether.

This 'key' of the vendors does also show that they place a high importance on statistical biases – they highlight loopholes and, in some cases, welcome state-specific regulation. Therefore, it is the vendors who highlight the vulnerability of the frame. The current guidance offers no concrete direction for what vendors should be measuring within impact assessments, and it also bears a loophole which means this regulation is insufficient for operationalising its chosen statistical standard. Vendors, in this sense,

seem open to more regulation and future research should explore more appropriate statistical measurements which vendors can use to test their systems during development and pre-deployment.

Secondly, then, the regulators 'key'. Those who I spoke with thought that the current federal laws were sufficient to tackle bias and discrimination in hiring.[42]

Regarding the 'business necessity' clause, Brandi Meredith (EEOC) said to me:

> "I wouldn't call it a loophole. [Laughs] I think that it's… […] you know, it was drafted in the 70s so there have been things that have changed, or mechanisms that have changed, and so best practices still apply under that document."

I found that it was common for the experts within these government bodies to tell me that these laws, while old, were sufficient:

> "[M]ost of the employment laws are from the 1960s or so […] But even with that, the laws, like, the way that we apply the law will be the same regardless of, you know, the age of these laws. So even if it's a law from 1964 like the Title VII of the Civil Rights Act, it applies with equal force to new and emerging technologies." [Bradford J. Kelley, EEOC]
>
> "I say our laws here are old, but they're not outdated, you know, they apply with equal strength to decisions that humans were making since the 1960s as they to do technology." [Keith E. Sonderling, EEOC]
>
> "I think that, you know, that we feel like our statutes do apply in this context […] I definitely think that we have the tools and it's a matter of deploying them, I think. We have some pretty good tools." [Eunice Ikene, EEOC]

This picks up on, and rebuts, another reason why sometimes these regulations are seen as insufficient – they were designed and adopted before these systems existed. As Kroll et al. (2016) note, the tools currently available to courts and policymakers were developed to oversee human decision-makers, and therefore might fail when applied to technology.

[42] These viewpoints might have been different had I spoken to people at a state level.

In the case of the UK, similarly, it is only historic laws – originally designed to be applied to human decision-making - which can be applied to discrimination through AI hiring. Bias in algorithmic decision-making in recruitment is currently governed primarily by the UK Equality Act 2010 and the Data Protection Act 2018. While these laws do not lay down a statistical rule of thumb akin to the 4/5th rule (Binns et al., 2020), these tangential laws also exhibit similar sentiments to that of Title VII's 'business necessity' clause. Sanchez-Monedero et al. (2020) points out that the UK Equality Act 2010, article 19(2) says that bias can be justified if it is shown to be a 'proportionate means of achieving a legitimate aim', without legitimacy even being defined.

It is noticeable how similar the regulators answers were surrounding sufficiency of the laws. It shows, perhaps, that this is a united take from government on the sufficiency of existing anti-discrimination law. These regulators thought existing law and regulation was sufficient to regulate AI-rec-tech in the same manner as it regulates humans and did not acknowledge that the 'business necessity' clause presented a loophole.

When I looked more closely at the interviews with regulators, I realised that, perhaps, this was a way of balancing regulation and innovation. This was something mentioned by a few of the regulators:

> "[W]e don't wanna stifle innovation […] but at the same time, focusing on the risk, you know, what are the risks involved for the employment discrimination purpose." [Bradford J. Kelley, EEOC]

Sonderling et al. (2022) also argue for a deregulatory approach to using AI in the context of hiring, which both ensures anti-discrimination employment practices, but will also maintain and spur further innovation. They argue that, instead, self-regulatory approaches

(e.g. self-audits and self-regulation) for vendors are highly effective mechanisms to manage risk because of the rapidly changing nature of the underlying technology and the existing legal mechanisms that can already help reduce these risks without stifling innovation.[43] Perhaps this is unsurprising given that Sonderling is a Republican commissioner, and the party has traditionally promoted de-regulation of corporations. However, as shown above, the vendors are already self-auditing and self-regulating, but are both open to further specific regulation and are also using a borrowed frame intended for employers which has no guarantee of making any difference to actual employment decisions.

Part of the reason behind the different 'key' of the regulators might therefore be an attempt to strike a balance between innovation and regulation, but it is deeply interesting to see that the regulators are the ones who are reticent to spend time bringing in more stringent regulation, and the vendors are the ones who seem to call for it. The vendor companies seem to be the ones using the federal laws in which to root their conceptualisation and operationalisation of gender bias in the statistical frame for lack of more appropriate guidance, scrutinising these existing regulations, and supporting more state-specific regulation.

These different 'keys' might be occurring because regulators are not allowed to scrutinise these laws, and because vendors long to have solid measurements which show the impacts of their systems are not detrimental, especially as this might re-assure their clients. It

[43] In the UK, the current Conservative government also illustrates a pro-innovation approach to regulation, arguing that regulation is not always the most effective way to support responsible innovation (DSIT, 2023).

might also be because vendors are genuinely concerned about bias and discrimination and want to ensure these systems do not cause harm to vulnerable groups.

However, this is not all to say that these regulators think existing law and regulation is fool proof. Indeed, I found that they were focused on different issues to those of the vendors.

The main insufficiency highlighted by the experts was that of enforcement.[44]

Regulators often noted that enforcing these laws was difficult. Many job candidates either don't know how to bring claims forward or, for some reason, cannot. To bring a case to court, the claim needs to be brought to the EEOC, but in many instances, people don't even know that these systems are being used on them. It might also be that they see the claim as being more effort than it's worth in terms of their time, or they worry about the future consequences on their reputation. In addition, even if the EEOC made the case a priority, its limited resources might constrain it's efforts – it currently receives nearly 10,000 new charges annually and it faces a persistent backlog (Kim, 2017).

When I asked if they get the claims often, Keith E. Sonderling (EEOC) told me:

> "I can't talk about that by law, whether or not we get claims, but I can tell you we've never filed a public lawsuit relating to true artificial intelligence[45] […] I think it's just a lack of awareness. I've been out there saying like, look, if you know this is being used, I'm willing to do this, I'm willing to bring one of these, um, cases."

[44] I did also consider the financial barrier, but it is free to bring a charge to the EEOC (EEOC, 2024b).
[45] This was at the time of interview (November 2022).

He continued by telling me that often, prospective employees don't know they're being subject to AI, perhaps they just think it is someone saying they weren't qualified. They just get an email saying, 'you are not qualified for this position', and so they don't know why that's happened.

This is also pointed out by Brandi Meredith (EEOC), who talks about the difficulty of people bringing complaints to them:

> "I think that a particular challenge for our agency - because we're primarily… our enforcement mechanisms, like, are charge-driven - so people bring complaints to us and if they don't know they have a complaint then obviously it's a challenge for them to bring it to us. Um, and so we do have other enforcement mechanisms we can use like commissioners' charges and things like that."

I was told about commissioner's charges multiple times throughout various interviews. Essentially, this gives the senate, the commissioners specifically, the ability to file cases of their own. But again, this takes knowledge, for example through a whistle-blower. In addition, Eunice Ikene (EEOC) also told me about third party charges:

> "It could be a workers centre or, um, some other community-based organisation could file charges on behalf of a group of individuals or on behalf of an individual and so I think there's a lot of flexibility there."

Therefore, while Sonderling et al. do argue for a de-regulatory approach for vendors, they, along with the other regulators interviewed, also argue that federal agencies should fully utilise the tools at their disposal to address discrimination committed by employers – EEOC commissioner charges, direct investigations, and encouragement and incentivisation for companies to inform applicants about what data is used in the hiring process (Sonderling et al., 2022).

There are similar concerns about enforcement and claims in other jurisdictions. For example, there is a similar issue in terms of candidates knowing about bias and discrimination with the GDPR. Article 22 of the GDPR allows candidates to object to decisions which have legal or similar effects being made about them using their personal data and by solely automated means (e.g. recruitment using an automated tool). They can request for a human to make the decision or reconsider it (Sanchez-Monedero et al., 2020). However, the candidate might not realise that this system is being used. For this reason, MEPs shaping the EU AI Act want to boost citizens' right to file complaints against AI systems and receive explanations of decisions based on high-risk AI systems (European Parliament, 2023).

Overall, the second way in which vendors and regulators are out of 'key' surrounds the sufficiency of the current legal landscape to address bias. While vendors place high importance on statistical bias and use the 4/5th rule in their framing of gender bias for lack of more specific or tailored guidance, they are also critical of the insufficiencies of existing regulation. In another 'key', regulators see these laws as sufficient to tackle bias in hiring. This highlights the vulnerability of the regulators original frame to loopholes, which means the function of this frame is compromised considering the 'business necessity' clause, which could be used to excuse stark disparity in outcome ratios between demographic groups.

6.5 Discussion

Increasingly, AI-rec-tech is gaining the attention of those in the legal domain. We see this in the recent legal action around discrimination and privacy brought against employers

using these systems (EEOC, 2024c; Henry, 2022; Spanjaart, 2022), through the EU AI Act proposal labelling these systems as 'high risk' (European Commission, 2021), and through the influx of state-specific laws in the US to regulate these technologies (DCWP, 2023; Artificial Intelligence Video Interview Act, 2020; Sloane et al., 2022; Sonderling et al., 2022).

In this chapter, I have demonstrated how anti-discrimination law has influenced the vendors' framing of gender bias in hiring through discussing the statistical frame. Here, I have focused on the five vendors in the US, who generally showed a consistent frame, although other vendors did also mention statistical outcome in relation to bias.

To answer sub-RQ#1 in relation to this frame, I showed that gender bias was conceptualised statistically, specifically as adverse impact. Bias was indicated through the ratio of the highest selected demographic group, versus another demographic group, being lower than 0.8. However, as I understand it, this does not apply intersectionally, for example to look at the ratio between white males and black females, which is certainly a limitation of this regulation. In contrast, an un-biased situation was perceived to be statistical parity – or near parity, for example a ratio of 0.8 or above between these two groups. This frame was outcome-driven and was rooted in the 4/5th rule, itself based on Title VII of the Civil Rights Act 1964.

I then answered sub-RQ#2 for this frame, exhibiting how vendors operationalised addressing this form of gender bias through adverse impact assessments. These assessments are self-imposed and self-regulated. Here, I showed some general trends about these impact assessments which I found in my data: they are often conducted

periodically on accumulated client data, and there is uncertainty around how demographic groups are selected to be assessed.

I also noted in this section that there are two ways, in addition to the lacking consideration of intersectionality, in which these impact assessments might not capture the full picture statistically. Firstly, the lack of data on some demographic groups (such as non-binary individuals) might skew the statistics. This might leave room for the loophole in the Uniform Guidelines where a greater difference than 0.8 is not classed as disparate impact if based on small numbers. Secondly, crucial details about impact assessments are unclear or de-contextualised. For example, it is unclear how are vendors dictating 'success' or 'pass rates' to calculate the ratio, particularly if/when they don't have access to client data on who has been successfully hired. This operationalisation therefore demonstrates, I argue, an 'invisible out-of-frame' vulnerability, and more research will be required to discern details about these impact assessments. Additionally, calculating these ratios with accumulated client data without knowing who was hired means that these ratios are de-contextualised and do not reflect specific instances of hiring. Therefore, they are unlikely to be a consistently meaningful reflection of their systems in use.

However, it is positive that vendors are thinking about bias in terms of outcome and impact. Therefore, I argue that while there are vulnerabilities in this frame, the foundational underpinning of conceptualising bias as outcome-based and statistical is important.

Next, I explored two further vulnerabilities of this frame. I conducted this analysis using Goffman's idea on 'keys' (Goffman, 1974). Understandings can be in harmony (in the

same ‘key’), or dissonant (in different ‘keys’). In this case, the law and regulations in which this frame was rooted was understood by the regulators in one way and by the vendors in another.

Regarding the compliance ‘key’, I discussed the de-contextualisation vulnerability. I found that some vendors stated they were ‘EEOC compliant’. In contrast, regulators asserted that vendor compliance is not possible. Not only do the EEOC not issue compliance audits, but this guidance does not apply to them (it is aimed at employers) and is not intended to be used in this context. Moreover, the vendors often run their adverse impact assessments on aggregated datasets, which is somewhat meaningless in relation to this guidance of the 4/5th rule which is intended for specific employment decisions. For the liable party – employers - to comply, this statistical test must be run on the employer’s data. Therefore, this frame is de-contextualised, which leaves it vulnerable by compromising its function; it is uncertain how the vendors’ assessment on accumulated client data might work to improve statistical bias in practice.

This vulnerability – brought to the surface through the different ‘keys’ surrounding compliance - raised several questions and points:

1. **Claims** – why do vendors claim to be EEOC compliant? Perhaps it is to signal that they are considering bias and to reassure their clients. Regardless, this claim should stop being advertised on vendors’ websites and to their clients; it is misleading in terms of where the legal liability lies.
2. **Liability** – who should be liable for adverse impact? In Chapter 4, it became clear that employers have options and control over aspects of system use, but does this mean that vendors should have no liability for bias and discrimination taking place

through these systems? There are moves elsewhere to put more legal responsibility onto the vendors (e.g. EU AI Act, Algorithmic Accountability Act 2023 in the US). I support a shift towards more joint liability for disparate impact emerging through system usage, as vendors are also responsible for transparency with their clients about system functionality. In addition, this might help to form specific and more appropriate statistical tests for vendors to use in the context of development and pre-deployment.

3. **Impact assessment design** – questions were raised around how vendors are conducting their impact assessments. Unfortunately, my interviews did not go in the right direction to answer this, but in the future research it will be especially important to understand how impact assessments are being conducted when vendors don't have access to client data about who has been hired. What are they using to measure 'success' or 'pass rates'? Why do they not perform these on specific client data? And how are they choosing which groups to perform these assessments on? In addition, do these pre-implementation impact assessments make a difference to the employers' ratios in practice?

Regarding the sufficiency 'keys', I showed that some vendors say there are deficiencies in existing regulation. Ratios below 0.8 can be justified by business necessity. In another 'key', regulators argued that the laws, while old, are sufficient and enable a balance of regulation and innovation. However, they saw insufficiencies around enforcement. Therefore, I argue that vendors highlight vulnerabilities in the regulators frame, which they have borrowed in want of more specific and tailored regulation. Specifically, it sheds light on the loophole which means that this chosen statistic for employers doesn't apply in some circumstances, which surround the employer's business needs. This vulnerability

also highlights, once again, that there is no guidance specifically for vendors on development or pre-deployment impact assessments, something which is desperately needed, and which vendors seem open to receiving.

There are questions and points brought to the surface through the different 'keys' surrounding sufficiency of the law:

- **Nature of measurement** – there are multiple questions surrounding the suitability and adequacy of the 4/5th rule. For example: is the 4/5th rule, or statistical parity, the best way to quantify bias for employers, let alone for vendors? How should measurements take place pre-implementation (for vendors) and post-implementation (for employers)? What happens if the employer is only hiring one person? How should vendors measure aggregated datasets to best highlight biases? We need regulation which covers what vendors should measure during development and pre-deployment, so that these companies aren't having to borrow statistical measurements intended for different context. Moreover, there should be a re-consideration of whether the 4/5th rule – and the 'business necessity' clause – are the most effective means for measuring employer bias.
- **State-specific laws** – there are some aspects of these laws which need tightening or clarifying. For example, the Illinois Act fails to define AI and is silent about penalties and enforcement, and similarly the one in Maryland is silent about enforcement (Sonderling et al., 2022). Additionally, the NYC bill does not mention the nature of the audit that is required (Fuchs, 2023; Sloane et al., 2022).
- **Enforcement** – how might enforcement be encouraged, especially on the assumption that claims need to be brought forward (as is the case with the EEOC)? More awareness should be raised for candidates so that they know how and why

> to make claims or seek advice on what could constitute the basis for a claim. Ajunwa (2023b), for example, has urged the EEOC and FTC to launch an education campaign to inform applicants about these rights so they can make use of them, and to create a website to help job applicants query automated hiring platforms.

Regarding the nature of measurement, this does raise an important consideration. It is crucial to measure the impact that these systems have on demographic groups. Numerical knowledge is important. But as Merry (2016) argues, if it's not connected closely to qualitative forms of knowledge, it leads to oversimplification, homogenisation, and the neglect of surrounding social structure. Raghavan et al. (2020) also argue that the 4/5th rule should not substitute for critical analysis into the mechanisms by which bias and harm manifest in an assessment: "the legal operationalization of the 4/5th rule as a definition of bias runs the risk of downplaying the importance of examining a system as a whole" (Raghavan et al., 2020, p. 478).

The statistical frame is a vital one, but both within and alongside these statistics, there must be more consideration of the nuances of the underlying social elements of the data, design, and use of systems.

Given that the social conceptualisations of gender bias were thorough in Chapters 4 and 5, it was surprising that vendors didn't consider these more fully in the statistical frame, using them to shape their measurements and impact assessments, particularly seeing as the 4/5th rule is not intended for their use. Perhaps borrowing the frame of existing regulation was due to re-assuring their clients and because, as discussed, there is no more

suitable numerical guidelines. However, the social framings of bias should be more embedded in impact assessments for vendors during development and pre-implementation.

Moreover, as the only pre-dominantly technical frame of the vendors, there was a lack of discussion on data input and algorithmic bias, with the core focus being on output. However, there should be more discussion of this going forward. As Kim (2017a) argues, statistics cannot address many of the reasons that automated decision processes might be biased in the first place, such as biased datasets or biased inputs, missing variables, skewed training data, selection of biased target variables.

Recently, there has been a turn to considering socio-technical audits, especially in the context of hiring technology. Some recognise that reducing concepts such as fairness and bias to quantifiable metrics might disguise the assumptions built into the systems (Kak & West, 2023). Sloane et al. (2022) have argued that technical audits – such as impact assessments - often fail to examine the assumptions that underpin the claims made by such tools, and so they introduce a socio-technical approach to inspect the assumptions underpinning the systems. Rhea et al. (2022) also discuss how audits are often concerned with adverse impact, but they present a socio-technical audit methodology developed to measure the underlying assumptions operationalized by systems, including the stability of personality prediction systems used in the hiring domain.

The vision from this thesis, and indeed this chapter, is that a pairing of socio-technical audits and statistical impact assessments might occur in parallel to capture all the frames assessed. Alongside the intrapersonal and structural frames, the statistical frame could be

revised to reflect upon the most effective ways to measure the impact on the most vulnerable groups.

This chapter has highlighted the importance that vendors place on statistical measurements for defining bias. This was further exaggerated through discussion of the different 'keys' of vendors and regulators, where vendors were seen to go further than required by law (when discussing compliance) and to point out deficiencies in existing regulation (when discussing sufficiency). However, these different 'keys' also illustrated vulnerabilities of the frame, particularly around the impact of using de-contextualised statistical measures and the 'business necessity' loophole in this existing regulation.

This frame, along with the others, also allows me to look at one side of the bi-directional process of the social shaping of technology (MacKenzie & Wajcman, 1999). Foremost, it reflects how law and regulation – itself the outcome of social processes - can shape the way technology is designed. However, both vulnerabilities show that, while the foundation of this frame is important, the conceptualisation and operationalisation need to be re-considered in terms of suitability and sufficiency. More needs to be done to integrate the social frames into the technical, including statistical impact assessments, data, and algorithmic design.

6.6 Conclusion

In this chapter, I have outlined the last of the three frames. I have highlighted how this is rooted in legal ideas surrounding discrimination, specifically the 4/5th rule and adverse impact, which is outlined in the Uniform Guidelines based on the Civil Rights Act 1964.

I showed how this was operationalised by vendors through impact assessments to test the ratios on the performance of different demographic groups. Moreover, I demonstrated that vendors and regulators demonstrate different 'keys' surrounding the law in relation to the vendors' operationalisation of bias. This dissonance was specifically in two areas surrounding the law: compliance and sufficiency. In the discussion section, I argued that crucial areas for future research stemming from these vulnerabilities include the nature of impact assessments for vendors and whether the current regulation is the best means of auditing these systems post-implementation.

In the following chapter, the discussion chapter, I bring the three frames together to explore the tensions, commonalities, and implications.

CHAPTER 7: DISCUSSION

7.1 Introduction

In this chapter, I aim to summarise the three frames outlined in this thesis. Additionally, I go a step further by drawing these findings together and evaluating the connections between the frames. Overall, I show that while these frames are and can be discussed separately, they are intrinsically connected and, together, provide a socio-technical framing of gender bias. This helps to further answer my research question.

Moreover, I recommend some short-term actions which can be taken by AI-rec-tech vendors, clients, and regulators to address the points which emerged in studying the vulnerabilities of these frames. This thesis does not aim to argue for or against the existence of these systems, but rather, it sees their production and use as inevitable. This research therefore aims to be constructive and productive in providing recommendations for action and future research which, I hope, will reduce the negative impacts stemming from the use of AI-rec-tech.

7.2 Thesis summary: the three frames

In this thesis, I have focused on the way(s) in which eight international AI-rec-tech assessment vendor companies conceptualise gender bias and how this is operationalised in the design of their technology. This study therefore offers the first in-depth, empirical, qualitative exploration of what lies behind the design of AI-rec-tech.

In **Chapter 1**, I outlined how AI-rec-tech systems can be used for the sourcing, screening, assessment, and selection of candidates (Ajunwa & Schlund, 2020; Black & van Esch, 2020; IFOW, 2020; Rab-Kettler & Lehnervp, 2019; Raghavan et al., 2020) and that there is a growing global market for their use (Köchling & Wehner, 2020; Sloane, 2023). I argued that studying predictive algorithms in this context is vital; these systems impact people's lives and have the potential to do so along gendered and racialised lines (Eubanks, 2017).

In **Chapter 2**, I demonstrated how my research focus and research questions emerged from reviewing surrounding literature. Assessing the literature around AI-rec-tech, I found that research on bias demonstrated limited empirical exploration of vendor companies, and none in relation to how they attempt to address bias through system design. This is the gap in existing work which this thesis aims to fill. I also explained my theoretical framework, an adapted version of Goffman's 'frames' (1974). I argued that this enables me to capture vendors' understanding of gender bias (their conceptualisation) and how this is mobilised in system design (operationalisation). It also provides the analytical tools to evaluate the vulnerabilities and strengths of these frames. This chapter therefore narrowed down the scope of my project and discerned my research questions:

RQ: How do AI-rec-tech vendors frame gender bias in hiring?

> **Sub-RQ#1:** How do AI-rec-tech vendors *conceptualise* gender bias in hiring?
>
> **Sub-RQ#2:** How do AI-rec-tech vendors attempt to *operationalise* addressing gender bias in their system design?

In **Chapter 3**, I discussed my methods. This included qualitative online interviews with both AI-rec-tech vendors and regulators; walkthroughs and autoethnography of the

assessments; and content research on conferences, websites, reports, and regulatory/legal documents. Interviews, my primary method of data collection, allowed me to capture the "knowledge, attitudes, and behaviours of individual producers" which are "the product of professional and organizational processes" (Vliegenthart & van Zoonen, 2011, p. 111). This was vital for me to explore, given the purpose of my research concerned the thought(s) and understanding(s) underlying design.

In **Chapters 4-6**, I answered my research question and sub-research questions. I did this by explaining three frames which demonstrate how vendors conceptualise gender bias in hiring and how they operationalise these understandings in system design. I also analysed the vulnerabilities and strengths of these frames, focusing more on the vulnerabilities. This enabled me to outline areas for action and future research. These frames were discussed discretely, although, as I will outline below, they are certainly not mutually exclusive.

- **The intrapersonal frame:** in **Chapter 4**, I discussed the intrapersonal frame (the first of three answers to my main RQ). Vendors conceptualised gender bias to be a symptom of human perception relating to stereotypes, homophily, and subjectivity (answering sub-RQ#1). This understanding was a strength of the frame; it was the most thorough conceptualisation of bias demonstrated by the vendors, as could be seen from the extent to which it resounded with existing literature and research.

  I showed that vendors attempted to 'remove' or 'improve' human perceptual biases through anonymisation, re-defining fit, and bias reminders (answering sub-RQ#2). This meant that, to the vendors, AI-rec-tech was superior to human recruiters when it came to addressing gender bias in hiring.

Additionally, I discussed the vulnerability of 'breaks in the frame', which I based on Goffman's idea of 'breaking frame' (Goffman, 1974). I defined this as inconsistencies between the conceptualisation of gender bias and the ability to fully address this bias in technological design. These breaks emerge where human perceptual biases can infiltrate the hiring process while vendors are designing these systems and while clients are using these systems.

Specifically, I outlined five different 'breaks in the frame': (1) the optionality break – optionality to use design features such as candidate anonymisation; (2) the hybrid break – human involvement in final selection stages and interviewing; (3) the ignorable break – design features such as bias reminders can be ignored in favour of human judgement; (4) the profile-shaping break – humans play a role in shaping the skills that are desirable for their employees; and (5) the contradictory break – despite their conceptualisation of bias, there continues to be stereotypical presentation of gender.

Given that some breaks can be avoided, and some can't, this frame is inevitably broken, meaning vendors cannot claim to remove human or intrapersonal biases through AI-rec-tech, or even improve the effects of human biases with any certainty.

- **The structural frame:** in **Chapter 5**, I discussed the structural frame (the second of three answers to my main RQ). This frame, along with the statistical frame, were not as commonly mentioned by the vendors as the intrapersonal frame, but they were still a consistent trend and were spoken about within most vendor companies.

  I showed how vendors framed gender bias in hiring in relation to inequalities in education and opportunity in society, or 'structural injustice' (answering sub-

RQ#1). While vendors were realistic about the scope and ability of their technology to impact the root cause of structural injustice, they sometimes perceived their systems as allowing people to be liberated from these inequalities due to hard work, ambition, and innate skill. I argued that this resonates with narratives of the 'American dream'. Vendors operationalised addressing this bias through neutralising questions and assessing skills/behaviours rather than qualifications (answering sub-RQ#2).

I noted the strength of this conceptualisation through showing how it maps onto evidence of labour market inequalities globally. I also demonstrated how the frame is vulnerable to 'out-of-frame' activity (Goffman, 1974). This included 'out-of-frame permeability': realities out-of-frame, or out of focus, which might permeate and therefore impact the frame in design. For example, structural injustices in education could still impact candidates' skills/behaviours. I also discussed 'invisible out-of-frame' factors: crucial factors of the design which are invisible to the researcher, but which impact the function of the system. For example, not fully understanding the underlying 'science' and training data used to measure these skills/behaviours, as well as the construction of the skills themselves.

- **The statistical frame:** in **Chapter 6**, I evaluated the statistical frame (the third of three answers to my main RQ). While all vendors mentioned statistical outcome of the assessments regarding demographics, the specific elements of this frame in this chapter mainly pertained to US-based vendors. The foundation underpinning this frame was a strength; it is grounded in attempting to measure the impact of the systems in numerical terms. However, upon closer investigation, there were vulnerabilities to the frame in terms of both understanding and design.

As vendors explained, their conceptualisation of statistical bias stemmed from the 4/5th rule outlined in the Uniform Guidelines which is itself based on Title VII of the 1964 Civil Rights Act. I explained that the 4/5th rule is a ratio between the most successful demographic group and any other demographic group, which, according to this rule, should be no less than 0.8 (answering sub-RQ#1). Moreover, I showed how this statistical measurement is operationalised largely through internally conducted impact assessments which calculate the ratio between demographic groups (answering sub-RQ#2). There was variation in how these were conducted, but generally they happened periodically on accumulated client data. I argued that these might not capture the full picture due to lack of data on some groups and crucial details of how impact assessments were being performed remaining uncertain ('invisible out-of-frame').

Following this, using Goffman's theory on 'keys' (Goffman, 1974), I highlighted two further vulnerabilities which emerged from showing how and where views of vendors and regulators diverged. This occurred around 'compliance': where vendors sometimes understood themselves to be 'EEOC compliant', regulators asserted this wasn't possible. This highlighted the de-contextualisation vulnerability – the fact this frame was de-contextualised from its originally intended context (employer's final decisions, rather than accumulated client data) meant it did not reflect actual hiring decisions. It also occurred around 'sufficiency': where vendors perceived current law and regulation to be somewhat insufficient, the regulators tended to view the laws as sufficient. This showed the loophole vulnerability – this applied to the regulators (the original framers, from whom the vendors have borrowed this frame), highlighting a loophole of 'business necessity' in existing anti-discrimination law. Overall, I argued that the

focus of vendors on statistical outcome is important; it allows an estimation of how the intrapersonal biases and structural biases are intersecting with results. However, future work needs to evaluate whether this chosen statistical measurement is the most suitable one; I argue this frame could more effectively incorporate the social biases discussed by vendors into its numerical measurements.

7.3 Notes on the nature and characteristics of the three frames

There are a few brief points to note about the nature and characteristics of the frames before delving into the relationship between them.

Firstly, throughout discussion of the frames, it became clear that gender bias was treated similarly to biases in relation to other protected characteristics. While vendors gave examples relating to gender - largely because I was asking them about gender bias - other protected characteristics were often referenced during our discussions. As seen in Chapter 5, for example, alongside gender, there was discussion of class and race in relation to educational opportunity. Therefore, this work and these frames are potentially more widely applicable to other protected characteristics as well as gender.

Secondly, it is interesting to reflect upon whether bias was spoken of as binary (either there or not there) or a gradation (reducible, but always there to some extent). This varied between frames and participants, with no clear trend between companies (e.g. everyone in one company wasn't claiming they could 'remove' bias, and therefore perceiving it in binary terms). In Chapter 4, I showed that some participants did claim that these systems

can “remove that subconscious bias” [Company Red HOS] (seeing it as a binary phenomenon), while others said that it is “impossible to take all bias out” [Head of Impact at Pymetrics], or that it is “really hard to eliminate bias altogether” [Company Yellow IOP2] (perceiving it as gradation). In Chapter 5, bias was often recognised as a gradation which could not be fully removed; structural injustice was a social reality being reflected through these systems, but something these companies could have a “small positive impact” [Company Red CSM] in addressing. Although, again, some participants did speak of ‘eliminating’ bias in this context, perceiving it in binary terms. This variation highlights that collective understandings are constituted of personal and individual views which aren’t always the same.

Thirdly, I want to consider whether there are any major gaps in how the vendors conceptualised gender bias in hiring. To reflect upon this, I revisit the literature review. Overall, the vendors considered many of the key areas discussed in the literature relating to gender bias in hiring. As argued throughout, generally, I found their conceptualisations of bias to be thorough, complex, and well-grounded. In Chapter 4, I showed how the vendors’ framing of gender bias as intrapersonal resonated with a wealth of literature surrounding stereotypes and homophily in hiring and at work. Additionally, vendors exhibited both social (Chapters 4 and 5) and technical (Chapter 6) interpretations of bias, which was mentioned in the literature review as a common way of dividing understandings of bias in AI.

There were some areas of literature which do not emerge in the vendors’ understanding of gender bias. For example, there was no discussion of gender bias as ‘exclusion’, a theme which I showed in the literature review to be prevalent in work around gender bias

in AI. This was highlighted by the fact that - within some vendor companies - there was no mention about moves to include non-binary or gender fluid representations in avatars, chatbots, or in the demographic forms before tests, as noted in Chapter 4. This, I hope, can be remedied through action from vendor companies. Additionally, there was a lack of discussion surrounding candidate experience, another prominent body of literature concerning AI-rec-tech discussed in Chapter 2. This is understandable, seeing as it's not common to define bias in relation to experience. However, it was a conspicuous gap in our conversations, and it might be useful for vendors to consider how candidates experiences change the way they complete assessments and how this might intersect with issues of bias.

Fourthly, it's important to note that vendors aimed to balance diversity and validity through their frame operationalisations. Some of the most 'valid' predictors of job performance are sometimes thought to be associated with racial-ethnic and gender subgroup predictor score differences, generally disadvantaging non-White people and women (De Soete et al., 2013; Madera & Abbott, 2012; Ployhart & Holtz, 2008; Pyburn et al., 2008; Trindel et al., 2021). This is commonly referred to as the 'diversity-validity' dilemma. Employers generally want to identify diverse but high-quality candidates (Madera & Abbott, 2012; Trindel et al., 2021). Emphasis has long been placed on strategies to minimise these differences, for example, encouraging affirmative action to increase the representation of disadvantaged groups (De Soete et al., 2013; Kravitz, 2008; Rottman et al., 2023).

The vendors evidently tried to balance diversity and validity in their system design. For example, in Chapter 4, I showed how vendors anonymised candidates so that the primary

focus was on job-relevant skills, and in Chapter 5, I demonstrated how vendors aim to measure skills and behaviours rather than qualifications, which could be proxies for protected characteristics. Both show a focus on 'getting the right person for the job' (validity) while also reducing the influence of factors such as human perceptual biases or educational background/qualifications to increase diversity. However, as outlined in Chapter 5, assessing skills and behaviours - depending on what they are - might still act as proxies for protected characteristics. Therefore, more research is needed into how employers/vendors are deciding desirable skills for jobs (e.g. what is meant by validity?), and whether/how assessment of skills and behaviours align with protected characteristics (e.g. does/how does the validity-diversity dilemma play out in practice, and what does this signal?). I discuss this more in Chapter 8.

Fifthly, it is interesting to consider the idea of 'frame-shifting' (Lorino et al., 2017). In the context of negotiation between retailers and suppliers, Lorino et al. (2017) illustrate the idea that when one frame fails, there can be a subsequent shift to another frame which narrates the situation in a different way. Transferring this to the context of the vendors' frames, the fact there are multiple frames means there could be frame-shifting in the face of vulnerabilities arising in one frame. For example, it might be that if the statistics don't meet exactly the threshold desired, this could be explained through the structural frame. Or, if someone asks whether there is proof that these systems can reduce human biases, the vendors might offer statistics in reply. In other words, this multiplicity of frames might help the vendors to justify the function and use of their systems.

Lastly, regarding the social shaping of technology (MacKenzie & Wajcman, 1999), I aimed to touch upon not only how designers are shaping technological systems around

their own social views, but also what roles technology is playing in changing, reflecting, or reinforcing aspect of our society. I showed in my empirical chapters how the views of the vendors were shaped by resounding narratives and realities in our society, such as human perceptual biases and inequalities in education and opportunity. While it is clearer to see how the technology reflects – and attempts to subvert – certain social realities, it is much harder to discern how these systems might be shaping society until we begin to gather more data on their impacts.

Now, I go on to discuss several considerations about how the frames relate to one another.

7.4 Common 'vulnerabilities'

When it comes to the relationship between frames, it is crucial to note the common vulnerabilities. In Chapter 2, I defined vulnerabilities as complexities within or surrounding the frame which impact its integrity or function. Crucially, this means either the vendors' understanding, or their design, is impacted by a factor which could be either within or outside of their control.

In the findings chapter, I translated Goffman's ideas on vulnerabilities to analyse the three frames of the vendors:

- 'Breaks in the frame' (Chapter 4): internal inconsistency between understanding of gender bias and ability to address it through system design.
- 'Out-of-frame' activity (Chapter 5): external or unknown factors impacting the function of the design.

- ‘Keying’ (Chapter 6): contradictions or differences in understanding which impact the function of the design.

These vulnerabilities can concern the understanding and design itself (as seen in Chapter 4) or can be in relation to other stakeholders (as seen in Chapters 5 and 6). Vulnerabilities also highlight directions for future research which will help discover more about these frames and their effects.

Interestingly, though, some of these frames are vulnerable in more than one way. This was demonstrated in Chapter 6, where I showed that uncertainty around how vendors conduct impact assessments could also be considered ‘invisible out-of-frame’.

Given that I focused on the vulnerabilities rather than the strengths of the frames to clarify areas for future research, I do not concentrate on common strengths here. However, it is worth noting that the first two frames (intrapersonal and structural) showed strengths in their understanding of gender bias, which either resounded with existing literature (Chapter 4) or evidence in society (Chapters 4 and 5). As discussed in Chapter 2, a strength can be in relation to an understanding alone (if it is backed up by data or theory) or can refer to how an operationalisation effectively addresses this concept through counteraction/mitigation. While the conceptualisation of gender bias in Chapter 6 had an important foundation underpinning it, the conceptualisation itself was based on a mis-adopted regulatory understanding and, as I argued, demonstrated vulnerabilities due to this.

While I generally focused on specific vulnerabilities for each of frame, there are ways in which that vulnerability might also be applicable to other frames. Here are some examples:

- **Client's usage is 'invisible out-of-frame' and potentially a different 'key':** in Chapter 4, I delineated that there is a lack of understanding in AI-rec-tech literature about how clients use AI-rec-tech systems. This could be recognised to be 'invisible out-of-frame' activity to which this frame is also vulnerable. Client usage encompasses crucial information which might change how the design functions. Additionally, it's possible that clients might not concur with how vendors represented their motivation(s) for system usage. As discussed in Chapter 4, vendors didn't interpret diversity of candidates to be a priority for their clients, but clients might disagree with this understanding. This would demonstrate different 'keys', which might impact how clients are using these systems optional design features.
- **Permeability as a 'break in the frame':** in Chapter 5, I showed that while some of the vendors perceived the assessment of skills and/or behaviours to be neutral, structural injustices in education and society might still impact these factors. While I analysed this as 'out-of-frame permeability', it could also be seen as a 'break in the frame', because the operationalisation of assessing skills and behaviours might be unable to address the conceptualisation of bias as structural injustice.
- **Impact assessments as 'invisible out-of-frame':** in Chapter 6, I discussed how internal impact assessments conducted by vendors to assess the outcome of these systems are largely opaque, particularly in relation to the 4/5th rule. This

knowledge could also be considered 'invisible out-of-frame', and therefore vulnerable to impacting the function of the design.

- **'Break in the (statistical) frame':** in Chapter 6, I discussed how the vendors conceptualise gender bias in relation to the 4/5th rule. However, given that their operationalisation is likely to occur on accumulated client data, this does not actually measure the ratio of demographics in the context of specific companies. This could be interpreted as an inconsistency between the operationalisation of gender bias and their underlying understanding, and therefore as a 'break in the frame'.

These common vulnerabilities add an extra layer to answering my main research question; they show how frames can be vulnerable in multiple ways and highlight further areas for future research which I incorporate into the agenda for future work in the following chapter.

Additionally, while this thesis presents three ways in which frames can be vulnerable, there are bound to be other ways in which frames might have their integrity or function compromised. For example, in *Frame Analysis,* Goffman also outlines the vulnerability of 'fabrication': the "intentional effort of one or more individuals to manage activity so that a part of one or more others will be induced to have a false belief about what is going on" (Goffman, 1974, p. 83). This could manifest in design which intentionally appears to address a particular conceptualisation of bias, but in fact does nothing of a sort. This could only be discovered with more in-depth research on how these systems work under the hood along with exploration of vendors' intentions. Frames could also be vulnerable to 'ordinary troubles', as Goffman discusses, occurring when people 'mis-frame' due to

'ambiguity' around the primary framework or common understanding (Goffman, 1974). This might occur when a conceptualisation is completely detached from common understandings of a term, which we do not see occurring with any of the vendors in this study.

7.5 'Laminating' frames: separate, sequential, and socio-technical

'Lamination' is outlined by Goffman as multiple layers of experience. Goffman speaks of this specifically in relation to re-keying within one frame, where each different 'key' is "adding a *layer* or a *lamination* to the activity*"* (Goffman, 1974, p. 82). Here, I transform this analogy to refer to layers or laminations of different frames. Lamination, therefore, elicits an exploration of how these three frames relate to one another: can they operate all at once? Are they sequential, related, or self-contained?

Again, this consideration adds another layer to answering my main research question; it examines how the vendors frame gender bias in hiring, considering whether these frames are inseparable, ordered, or completely discrete.

In Chapters 4, 5, and 6, I outlined the frames separately. Indeed, they were often spoken of distinctly by vendors. In these chapters, it was evident that some companies didn't mention various conceptualisations or operationalisations of gender bias, showing that the frames can be construed independently of one another. An example of this occurs in Chapter 6, where US-based vendors were the ones primarily framing bias in line with the Uniform Guidelines based on Title VII.

However, here, I argue that while these frames can be seen separately, they are intimately related. More than this, when held together – laminated – it is evident that these frames are somewhat sequential.

As discussed in the literature review, biases in AI systems are often divided into social biases and technical biases (Friedman & Nissenbaum, 1996; Hall & Ellis, 2023; Lopez, 2021). This resonates with the frames, which are predominantly social or technical. The first two frames relate more to social realities and how to account for these inequalities in the design - and the third is more technical – in contrast to the first two, it is an outcome-based measurement, related to how we calculate these realities in terms of the impact of these systems.

In theory, the two social frames (intrapersonal and structural) direct and underlie the technical frame (statistical), as they should outline the disparities to be measured. The intrapersonal and structural frames are deeply related to one another. Structural injustices are built on a legacy of human perceptual biases such as stereotypes, and in turn these human perceptual biases are reinforced by structural inequalities in society.

These frames are therefore sequential (the social precedes the technical) and are all important in their own ways. While simply a technical frame wouldn't make sense (as it wouldn't be clear what to measure), solely a social frame would be lacking without consideration of the outcome or how this might impact the data or algorithmic design. Generally, I found that all vendors discussed both social and technical elements of bias in interviews.

Consequently, when laminating the frames, we see they can either be held separately, but are also intrinsically related and sequential. Therefore, to summarise how these three frames answer the main research question would be to say that - when all the frames are held together - the vendors frame gender bias in hiring *socio-technically*.

The term 'socio-technical' is often used to refer to systems involving a complex interaction between humans and machines, or, stemming from this, it can denote a method of system design which considers both social and technical factors and aims to optimise for both (Baxter & Sommerville, 2011; Mumford, 2006; Murray & Trist, 2016; Selbst et al., 2019).

Overall, I use the term 'socio-technical' to capture how vendors, whether knowingly or not, consider a combination of both social and technical aspects of bias in the framing of their systems, in terms of both understanding and design. This is evident from the consideration of social forms of bias - for example human perceptual biases (Chapter 4) and structural biases in human institutions such as education (Chapter 5), and technical forms of bias - such as statistical disparity (Chapter 6). Similarly, Selbst et al. (2019) describes ML systems that consist of a combination of technical and social components as 'socio-technical systems'; fairness cannot exist as a purely technical standard, but system design must think about people and their complex environments, relationships, organisations, and social norms in tandem.

As has become evident throughout this thesis, there are limitations to both the isolated social and technical framings of bias. But additionally, while these frames are sequential

and - together - socio-technical, there are limitations to the translation of this social framing into an appropriate technical one.

In Chapter 6, I showed that technically, not only was there a lack of discussion about data input and algorithmic bias (Zhou et al., 2022) with the core focus being on output, but the technical framing also showed lack of consideration of the social framings. Instead, it was focused on borrowing a regulatory frame which itself seemingly contained little consideration of underlying social factors. Additionally, there is a lack of consideration for intersectionality, and how different protected characteristics intersect within the nexus of power to cause vulnerability to some intersectional groups. In Chapter 5, disparity in education and opportunity was shown not only to disproportionately impact young girls, but to deeply intersect with race and class. However, in Chapter 6, with the employment of the 4/5th rule, this resulted in seeing each protected characteristic – gender, race, class – to be singular, and, therefore, biases within these might be seen as almost a 'tick box' exercise rather than the reality: a complex web of interacting inequalities.

Therefore, while the vendors did exhibit a socio-technical perception of gender bias, there is more to be done to embed these social understandings into the technical inputs, design, and outputs of the system. I hope to show ways for this endeavour to progress throughout the action recommendations and future research agenda.

7.6 Frame consistency

As noted throughout the three findings chapters, vendors were not always united in their precise views. For example, in Chapter 6, I showed that many of the vendors claimed to

be ‘EEOC compliant’, but the Company Blue CSO notably went against the grain and told me that it wasn’t possible for vendors to be EEOC compliant. This signifies how interviews within companies highlight both individual and collective views which aren’t always united.

In terms of operationalisations, too, vendors weren’t always aligned. For example, re-defining fit, discussed in Chapter 4, was primarily used by Company Red, whereas bias reminders, discussed in the same chapter, was adopted by more companies, along with using skills/behaviours to measure ability, outlined in Chapter 5, which was used by all companies.

However, broadly speaking, the frames were consistent across vendor companies. The intrapersonal frame (Chapter 4) was consistent across all companies, the structural frame (Chapter 5) within most companies – around five out of eight, and the statistical frame (Chapter 6) was mentioned by all companies, although I focused on the specific framing of the US vendors.

This consistency was evident within interviews, but also in their public- and client-facing documents, in conferences, and on their websites. Consistency is not common in research on frames, which can find a wide range of frames are interpreted by respondents, involving dominant frames evoked by government and economic elites, and critical frames developed by respondents (Gamson, 1992). Particularly, it is interesting considering that the vendor companies spanned across Asia, Europe, and the US. However, perhaps it is less surprising within this research given the topic of bias is widely

discussed, and the fact I studied a narrow range of companies rather than concentrating on the opinions of un-connected individuals.

The source of this commonality can only be speculated upon at this point, but there are various reasons why this uniformity might have occurred. Only future research will be able to determine the reason(s) with more certainty. But these potential reasons are important to consider; again, it adds another layer to answering my main research question by deliberating on why these ideas might arise.

Firstly, this consistency might have emerged from internal discussion and alignment. This likely happens during formal and informal meetings, water-cooler chats, and formal media or interview training where company values and 'party-lines' are discussed, used, and transfer naturally via osmosis.

Secondly, these ideas might arise from industry conferences, webinars, and/or training programmes (e.g. unconscious bias training or similar). In these settings, AI-rec-tech might be discussed and positioned in relation to concepts such as bias, fairness, and discrimination. It's possible that some of these narratives about how AI-rec-tech reduces bias are disseminated in these settings, or that vendor companies simultaneously conduct market research into other companies when locating their own ideas on gender bias.

Thirdly, the vendors' understanding(s) of bias tend to resonate with common narratives in society surrounding bias. In relation to the human perceptual biases noted in Chapter 4, for example, there is much discussion in the media about stereotypes in the workplace, subconscious biases, and hiring in our image (Carnegie, 2023; Grant, 2018; S. Iqbal &

Kar, 2021; Kramer, 2022). It is likely that the vendors have picked up on these narratives and used them to frame their own perceptions whilst also appealing to potential clients. This could also be the case with academic work, which seemed to resonate with the vendors' views, particularly around stereotypes and similarity bias discussed in Chapter 4. Given the degree of resonance with this work, it wouldn't be surprising if vendors were reading books and articles on these topics. In Chapter 4, for example, Frida Polli, CEO of Pymetrics cited *Thinking Fast and Slow* during our interview, showing that she is connected to mainstream academic work in these areas, aware of literature and common rhetoric, and is perhaps using it to position their systems in relation to bias. Indeed, this makes sense, as many of the individuals within vendor companies, including Frida, do come from a background of prolonged academic study.

Fourthly, specifically surrounding the statistical frame, the consistency of this frame might be connected to the multiple legal controversies surrounding these technologies, as discussed in Chapter 6 (EEOC, 2024c; Harwell, 2019a; Spanjaart, 2022). It's possible that vendor companies wish to reassure their clients and the public that they are 'complying' with the law and that they are open to more stringent regulation to appease any legal claims involving or referencing them.

It is crucial to note that the regulators were also consistent in their views. This was unsurprising, largely because most regulators I interviewed were based at the EEOC. It is likely that they will have a unified view on the matter because of their legal obligation to represent federal government, common understanding of the law, and have likely undertaken training for media communication or interviews.

7.7 Action recommendations

Lastly, I go beyond my research question(s) to consider how analysing these frames has highlighted actions which might address these vulnerabilities in the short and immediate term. This is - crucially - different to my recommendations for a future research agenda, from which I hope work will produce separate recommendations around AI-rec-tech design, regulation/law, and impact assessments.

Design features – recommendations for vendors:

- **Alteration of stereotyped chatbots and avatars:** in Chapter 4, I showed that despite the vendors' discussion of gender stereotyping as one aspect of human perceptual biases, some companies continue to use gender stereotyped women as representations in chatbots and avatars. Previous work has highlighted that feminising AI systems can have consequences for how women are viewed and treated in society (Bergen, 2016; Dillon, 2020; UNESCO, 2019; Woods, 2018). Vendors should alter chatbots and avatars to be more inclusive and less stereotypical. This de-feminisation and de-sexualisation of chatbots and avatars is crucial for how women are viewed in real social contexts, particularly in a world where OpenAI's most recent version of ChatGPT includes a feminised voice assistant inspired by 'Her', a film where a man falls in love with his voice operating system. Their new voice assistant can flirt, listen, be interrupted, sing, and read facial expressions, and there are attempts to build this into the iPhone in replacement of Siri (Robison, 2024).

- **Inclusive gender options:** in Chapter 4, I also showed that some vendors still only use male/female options for gender on demographic forms for candidates before assessments. This maps static categories onto diverse bodies (West et al., 2019). It is essentialist (Scheuerman et al., 2019), exclusionary (Keyes, 2018), and it bakes in the binary (Bivens & Haimson, 2016). This means that non-binary or gender non-conformist candidates have to either choose an incorrect option or refrain from using the service, generally making these participants feel excluded or uncomfortable (Scheuerman et al., 2021). It is important for vendors to change this to reflect people's identities, so that people are included and feel seen (Keyes et al., 2021), as well as to gather knowledge on these groups to see the full picture of who these companies are hiring. I suggest that vendors follow the proposition of M. K. Scheuerman et al. (2021) on their pre-assessment forms, using options: woman/female (cis or trans); man/male (cis or trans); non-binary; my gender is not listed above (open text box); unsure/questioning; prefer not to answer.
- **Remove affect recognition:** as noted in Chapter 5, while only three of the vendors I studied sold interview systems, only one of them used automated video technology (the others provided synchronous video platforms with no algorithms attached). They didn't use facial recognition technology, but only speech recognition to provide a transcript and score tone, speed, and so on. Affect recognition, video and/or speech, should be removed from AI-rec-tech systems. It is highly unlikely that these systems can infer skills and capabilities from this data (Raghavan et al., 2020), and, more than this, it is unlikely to be inclusive of disabled candidates, for example those with cleft lip or palate, achondroplasia, Down's syndrome, or those with speech impairments or deaf affect (Tilmes, 2022) or those who aren't speaking in their native language.

- **Anonymisation:** in Chapter 4, I outlined how - in some systems - recruiters are blinded to identifying information on candidates before a shortlist is formed. I also discussed in Chapter 4 how anonymisation of this sort has been shown to improve diversity (Goldin & Rouse, 1997). Currently, some companies do not offer this feature, and those who do give their clients the option to de-anonymise when they desire. I recommend that all companies should make it mandatory for pre-shortlisted candidates to be anonymised to recruiters; this might reduce the possibility of subjectivity, stereotypes, or homophily infiltrating decisions. Of course, this will need to happen alongside structural and social changes. As vendor companies see, there are still differences in how various demographic groups answer or complete the assessments.
- **Improving accessibility:** assessments should be as inclusive as possible, catering to people with visual and speech impairments, for example colour blindness, ADHD, dyspraxia, and those with mobility issues which impact their ability to use a computer.
- **Encourage candidates to complete demographic information:** vendors should consider an encouraging and transparent message to candidates when it comes to filling in their demographic information. They should explain transparently and concisely how this data is used and why this will help them to assess bias during auditing, which in turn helps future candidates.

Claims and marketing – recommendations for vendors:

- **'Removal' of bias:** in Chapter 4, I discussed how vendors cannot claim to remove human bias completely; human perception remains in the construction of skills, labelling and selection of data, and final shortlisting/decision-making. Companies

can only claim to provide design features which might mitigate elements of human perceptual biases and should remove all claims of the sort which say that their systems can 'remove' or 'eliminate' human biases.

- **'EEOC compliance':** as mentioned in Chapter 6, vendors cannot claim to be 'EEOC compliant' when the Uniform Guidelines only apply to employers.
- **Information to clients:** some of the design features such as re-defining fit and bias reminders (Chapter 4) are optional for clients to use to diversify their employees (in theory, although, in Chapter 8, more research is called for to assess this in practice). To compliment this, vendors should provide their clients with a concise information pack on why diversity will help their company productivity, and this should be tailored to each client to encourage them to use these features.

Law and regulation – recommendations for regulators, vendors, and employers:

- **Clarification of state-specific laws:** in Chapter 6, I discussed how there were some elements of state-specific laws on AI-rec-tech which are currently opaque and unspecified. This includes the definition of AI and specification of penalties and enforcement procedures in the Illinois Artificial Intelligence Video Act (2020) and the pass criteria or audit type in the NYC Local Law 144 (2023) (Fuchs, 2023; Sonderling et al., 2022). Along with the other unclarified parts of these state laws, these need clarification and subsequent revision.
- **Enforcement:** until more specific recommendations on law and regulation can be formed, there needs to be an increase in enforcement mechanisms for current law and regulation. This includes increasing awareness for candidates so that they know how and why to make claims or seek advice on what could constitute the basis for a claim. I agree with Ajunwa (2023) that the FTC and EEOC should

launch an educational campaign to inform applicants of their rights and create a website to help them query these platforms. In addition, vendors and employers should directly provide information to their applicants and employees on their rights and the laws which intersect with the use of these technologies.

- **AAA:** US Congress should move forward the AAA 2023 and, within it, should consider opening it to all vendors, not just those with more than $50mn annual gross turnover or $250mn equity value for a three-year period. In addition, they should consider adding specific statistical tests for vendors to perform pre-implementation. In doing so, this provides vendors with guidance on testing during development and pre-implementation.

Auditing – recommendations for vendors and clients:

- **Collaboration between vendors, clients, and academics/independent bodies:** rather than testing solely on their accumulated data, as discussed in Chapter 6, clients and vendors should collaborate to test the outcome of these technologies in practice. As a gold standard, this should be performed in collaboration with independent partners, for example with academic institutions through an open science framework, which would allow companies to share their data with these institutions in a protected way. This should be an iterative and ongoing relationship that represents real and contextualised decisions within client companies for all stages of decision-making. This would also be interesting for vendors, who would be able to determine whether they are producing significantly different outcomes for different clients. This might enable them to see where they should focus their attention in aiding clients with issues surrounding bias.

- **Socio-technical audits:** in Chapters 5 and 6, I discussed how, alongside statistical impact assessments, socio-technical audits should be conducted on vendor companies. Just as both the social (Chapters 4 and 5) and technical (Chapter 6) framings of bias are important, so testing systems should assess the social underpinnings and impacts along with the outcome of AI-rec-tech. These two processes (socio-technical audits and statistical impact assessments) should be done in parallel (Balayn & Gürses, 2021). In the following chapter, I discuss how there is work to be done on creating a standardised socio-technical audit which can be used for each stage of AI-rec-tech (sourcing, screening, assessment/interview, selection) as well as discerning the most effective statistical measurement for both pre- and post-implementation. However, in the meantime, alongside impact assessments, companies should make use of socio-technical audits within the social sciences (Rhea et al., 2022). For example, Sloane et al. (2022) introduce a socio-technical approach to auditing tools to inspect the assumptions underpinning the system which can be used by developers, employers, or the general public. It looks at the funnel stage, goal, data, function, assumptions, and epistemological roots. While this would be more workload and an additional cost if performed by an external independent body, to avoid this barrier to implementation, this should again be done through collaboration between vendors and clients, potentially in collaboration with academics or independent bodies.

These recommendations offer some short-term suggestions. These frames highlight that bias needs to be treated as a complex socio-technical ecosystem rather than something which is solvable solely through technical alterations (Barocas & Selbst, 2016), because

there are “some problems that cannot be fixed by a technical solution at all” (West et al., 2019, p. 18). The frames in design certainly illustrate the inseparable intersection between law, design, usage, and societal structures. Therefore, these links must be considered, and future work must focus on the surrounding structures and other stakeholders who impact the use and effect of AI-rec-tech systems, along with system design.

7.8 Conclusion

In this chapter, I have summarised the three frames outlined in this thesis, reviewing how they answer my research questions and considering their nature and characteristics. In addition, I continued to answer my main research question by considering several other factors.

I discussed ‘common vulnerabilities’, arguing that while I focus on specific vulnerabilities in each of the chapters for each frame, frames can be, and are, vulnerable in multiple ways. Moreover, I outlined what frames look like when ‘laminated’ and argued that while they can be seen as separate, these frames are also sequential and present an overall socio-technical framing of gender bias in hiring. However, I argue, there is more to be done to effectively embed the social understanding into the technical. I also discussed the consistency of frames between vendors and outlined how this might be due to internal discussions, industry webinars/conferences/training and common narratives in society.

The chapter culminated in several action recommendations for vendors, regulators, and clients for the short and immediate term, which will address some points arising from

frame vulnerabilities. The remaining points and questions raised through discussing frame vulnerabilities are incorporated into the future research agenda, proposed in the following chapter.

CHAPTER 8: CONCLUSION

8.1 Thesis contributions

Most people born today are likely to be subject to an AI-rec-tech system at some point in their working life. By 2030, the AI recruitment market is set to be costed at around $942.3mn (Dhapte, 2024), and it was estimated in 2020 that around 98% of Fortune 500 companies already use AI or data-driven systems in at least one of these stages of hiring (IFOW, 2020).

The impact of these systems on people's lives and livelihoods cannot be under-estimated; we are all, and will continue to be, affected by these technologies. This might occur directly, during applications for jobs, or indirectly, due to their influence upon who is hired into government, education, news media, and social media – sectors which affect us all in one way or another.

Overall, this project supplies an important contribution to better understanding what lies behind the design of AI-rec-tech, specifically regarding gender bias. Using rare empirical data from eight international AI-rec-tech vendors, I argued that there are three dominant frames which encompass how these companies perceive gender bias in hiring and how they attempt to address it in system design: (1) the intrapersonal frame, (2) the structural frame, and (3) the statistical frame. In exploring these frames, I answered my main research question: how do AI-rec-tech vendors frame gender bias in hiring? Along with my two sub-research questions, which aimed to investigate the vendors' understanding(s)

of gender bias (sub-RQ#1) and how this understanding is mobilised in system design (sub-RQ#2).

Furthermore, I analysed these frames, shedding light on their strengths and vulnerabilities. I argued that the vendors' conceptualisation(s) of gender bias were largely consistent, well-grounded, and complex. However, vulnerabilities emerged when reflecting upon the relationship between understanding and design, and the connection between system design and external societal factors such as structural inequalities and opposing stakeholder views. I highlighted these vulnerabilities through adapting Goffman's terminology around 'breaks in the frame' (Chapter 4), 'out-of-frame' activity (Chapter 5), and 'keying' (Chapter 6).

Lastly, I concluded that these three frames – while they can be held separately – are deeply related, sequential, and when seen together, exhibit a socio-technical understanding of gender bias. Attention to both the social and technical elements of bias is crucial; relying on a solely social or technical view of bias would be insufficient. However, as I argued, there should be more effort from the vendors in future to integrate the social understanding into the technical.

Exploration of the three frames also prompted me to recommend immediate action which might be taken to further address issues which surfaced during discussion of frame vulnerabilities (Chapter 7), and future research which would further address the remaining significant questions surrounding gender bias in relation to AI-rec-tech (Chapter 8).

Through all of this, this thesis has provided manifold contributions theoretically, methodologically, practically, and to existing literature and research.

Theoretically, this thesis contributes **the first piece of work to apply Goffman's theory on 'frames' to explore AI design**. This fresh use of theory contributes an innovative and creative mechanism for analysing the relationship between our social worlds and how aspects of it are transposed to AI design. Moreover, it provides new vocabulary through which to discuss and critique technological design such as 'break in the frame', 'out-of-frame', and 'keying'. In doing so, it contributes to literature within sociology which focuses on AI systems - particularly AI design - and answers the call of many sociologists to harness sociological tools to study the social roots of emerging technologies (Joyce et al., 2021; Liu, 2021; Sloane & Moss, 2019; Zajko, 2022). It is important for researchers to analyse how social realities influence and are affected by emerging technologies; this offers us ways to think about what our technology is doing and why, who it might affect and how, and what role we want it to play in our society. In addition, this application of frames is flexible, and could be applied to other AI systems, other technologies, or to how users (or those subject to systems) interpret these systems.

Moreover, this thesis also contributes new theory in the form of a **new taxonomy of gender bias in AI**. It does this through providing three frames of gender bias as perceived by vendors: the intrapersonal frame, the structural frame, and the statistical frame. These frames, while specific to AI-rec-tech, are transferable, and could be applied to other AI systems. They could also be relevant to biases surrounding race, ethnicity, class, and so on. Moreover, they are open to addition and revision in these different contexts. Once again, this offers a lens to reflect upon what our technology captures and what it might

reflect into society, which speaks to the wider aim of assessing the social impacts of technology.

In Chapter 3, I also presented a **new methodological concept for qualitative research on elites which entails access issues: 'the hustle'**. This illustrates how access issues impact research and researcher – it requires the researcher to act as networker, influences how much data can be collected, dictates research design, and alters interview dynamics (this is also published in altered form in Collett, 2023). These are important considerations, particularly in a world where researchers will increasingly need to explore settings where elites are making decisions which affect the masses. The concept allows researchers to reflect upon the barriers to accessing elite participants, as well as how to overcome them, and it captures how and why research which encounters access issues can still be rigorous and meaningful. While applied to interviews and elites in this thesis, it is also a valuable and transferable concept for ethnography or any other qualitative research on participants who are difficult to access.

Practically, in Chapter 7, I provided a set of **short-term action recommendations** which emerged from analysing the vulnerabilities of the frames. These offered practical recommendations to vendors, employers, and regulators about the design, auditing, marketing, and regulation of these systems. If implemented, I believe these will contribute to lessening the negative impacts of these systems in the immediate future. I have already engaged with some of the vendors on the action points, as they have had access to this thesis to decide their anonymity status. In some cases, I have been invited for talks, webinars, and discussions concerning these suggestions. I believe this engagement is a meaningful contribution of this work, and it was always part of my hope; it is through

this sort of collaboration that we might be able to shift the needle on the real-life impacts of AI systems.

Moreover, concerning the long-term, I provide a **detailed future research agenda concerning gender bias and AI-rec-tech** in this chapter. This also emerged from analysing the vulnerabilities of the frame, along with broader discussions stemming from frame analysis. This contributes to outlining the next steps for research on AI-rec-tech design, implementation, and impacts, by providing tailored suggestions based on the gaps and questions raised by my empirical findings. I hope, therefore, that this thesis forms the basis of a productive legacy of future work on this starkly under-researched topic.

To existing literature and research, this thesis contributes both directly to existing work and to broader research topics.

Directly, this work contributes multiple **new insights to existing literature on AI-rec-tech**. Firstly, regarding the topic of bias and assessment AI-rec-tech, this work fills a gap regarding vendors' understanding of gender bias and their attempts to address it. It offers rare empirical data which has been called for to investigate this (Dencik & Stevens, 2021; Stone & Dulebohn, 2013; Tambe et al., 2019). Much existing work on AI-rec-tech and bias has been based on theory or publicly available data (Drage & Mackereth, 2022; Raghavan et al., 2020; Tilmes, 2022), but empirical data is needed to delve deeper into what lies behind the design and what constitutes the design itself. AI-rec-tech companies are extremely difficult to access, and so obtaining this data was significant. This is important as it enables us to go forward with a better understanding of vendors' approaches and how the systems work, and this, in turn, allows a clearer vision of where

biases could arise. For example, this might occur due to de-anonymisation, employers choosing to 'clone' existing employees, external structural injustices, or insufficient or unsuitable statistical measures around impact.

Secondly, it also contributes to existing work on AI-rec-tech in legal studies. It does this, for example, by showing elements of system functionality (which has so far been lacking in legal studies literature), exhibiting that vendor companies are generally open to regulation, showing the current disconnect between regulators and vendors, and illustrating how/where more regulation is needed (as shown in Chapter 6, particularly around impact assessments specific to pre-implementation within vendor companies). In addition, it shows how influential law and regulation are on the ethical stances which these companies adopt. All of this contributes to knowledge within legal studies literature on these systems, necessarily shedding light on the views of the vendors in this conversation and the crucial next steps for regulating these technologies.

Moreover, this thesis makes a direct contribution to **literature on the use of AI in HR** by demonstrating critical avenues for future work in this area. AI is not only used within recruitment, but also within management – for example, systems which monitor employee performance through phone logs and emails, along with facial recognition and wearable devices to track activity and conditions (TUC, 2018) and systems which aid decisions around promotion and firing (Bales & Stone, 2020). While these offer flexibility to work from home (Millington, 2017), they might also exhibit differences in time away from work and therefore productivity due to differences in childcare and domestic work, which women tend to do more of, especially since Covid-19 (Craig & Churchill, 2021; Giurge et al., 2021; Sarker, 2021). The findings of this thesis, particularly in Chapter 4,

show how human interpretations of high performance can be integrated into these systems as ideal outcomes. AI management systems, too, should consider how human perceptual biases within the workplace are changing performance results, and how continued structural injustices, such as differences in care and domestic norms for men and women, are interacting with workplace performance which is being tracked by these devices.

More broadly, this thesis contributes to several other important conversations around the place of AI in our society and workplaces.

While incorporating the voice or perceptions of job candidates is not something this research does directly, this thesis could contribute to the practical endeavour of **the democratisation of AI development.** This aims to enable more people to participate in AI development and design processes in order to incorporate a range of diverse voices including those impacted (Ovadya, 2023; Seger et al., 2023; Simons, 2023). This study highlights the processes of these technologies: it shows that humans are still involved in final decisions given the 'hybrid model', that more emphasis is put on 'skills' and 'behaviours' than on qualifications, and that in some cases companies might work to 'clone' their high performers, in others they might aim to 'diversify'. All this illuminates elements of how these systems work. These facts should be of interest to many, considering that most people will likely be subject to these systems at some point in their working life. For future work to be able to centre the voice of workers, and to democratise the design and implementation of systems, these workers must understand and be informed about their underlying rationale and functionality so that they have sufficient context and can deliberate effectively (Ovadya, 2023). This thesis might form a part of the puzzle in that endeavour.

Separately, as outlined in Chapter 2, the underlying epistemological foundation of this work surrounds the **social shaping of technology** (MacKenzie & Wajcman, 1999). This work also contributes to this literature by demonstrating how and where societal ideas have contributed to shaping the design of the technology. For example, the vendors adopt views about human perceptual biases surrounding stereotypes, homophily, and subjectivity, all of which, I showed, are largely prevalent in society. These social views shaped the technology as the vendors attempted to counteract or mitigate these realities. This matters because it shows the power that AI designers have to adopt or ignore these views or realities, and this can play a role in how their systems function. While this thesis doesn't show how the technology shapes society, and doesn't claim to, it lays the groundwork for this by directing our focus to how and where the impact might happen. For instance, looking at impacts of design features such as anonymisation and re-defining fit, and whether skills/behaviour measurement is equalising the playing field considering structural injustices.

This thesis also contributes to conversations around **wider ethical issues concerning the use of AI**.

Most significantly, this thesis contributes to discussion on *gender bias.* It contributes to this literature by showing what is being considered and what is not being considered in the design of AI-rec-tech. Long-standing ideas of bias in hiring - from stereotypes to homophily - are being considered during the design of systems, with an attempt to counteract them. Regarding these considerations, though, this thesis also shows the limitations of what AI-rec-tech can do to reduce or remove these biases and thereby

highlights the limitations of technology to account for biases in hiring. The 'breaks in the frame' (Chapter 4) showed this, demonstrating the impossibility of removing human biases from AI, as did 'out-of-frame permeability' (Chapter 5), which displayed that the outcomes of systems are likely to still be influenced by a broader ecosystem of structural inequalities. Therefore, the thesis contributes to literature on gender bias in AI by demonstrating that there is a need for change to structures to address these issues, not just changes to system design. This was also highlighted by the 'out-of-frame permeability' vulnerability (Chapter 5), concerning the need to equalise educational structures, and the 'loophole vulnerability' (Chapter 6), which showed that changes to the law and regulation around these systems is necessary to ensure more integration of social inequalities into the statistical. The thesis also highlights what is not being considered in this framing, particularly around exclusion of certain genders from data collection and contradictions around stereotypical presentations of gender in their avatars and chatbots, which, I argued in Chapter 7, must change.

This research also contributes to conversations around other ethical issues concerning the use of AI. In terms of *transparency*, this thesis shows that vendors have been transparent to an extent and (at least some) are willing to participate in conversations around bias. Academics have long called for transparency of tech companies, particularly around the design of their systems (Felzmann et al., 2020). However, there is seen to be a lack of transparency in some areas as shown by the 'invisible out-of-frame' vulnerability in Chapter 6. This highlights how there is a long way to go in terms of learning about the underlying science and decisions which feed the data and algorithmic design, and further contributes to the idea that AI systems are often 'black boxes', even for academic researchers. Moreover, it shows that clients were not forthcoming to participate this

research, despite the fact there is a need for more transparency from clients to carry out future research. This might help to consider ways around this opacity, such as incentivisation for employers' transparency, anonymisation, and NDAs.

Regarding *privacy and surveillance*, this work contributes to this conversation by showing that we need more information on how this data on candidates could be used for other purposes. Collated, this data on candidates could be used maliciously if it got into the wrong hands, for example within social media to manipulate and misinform, something which is particularly pertinent in a year where over half the world will be voting in political elections. While there is assurance from vendors that this data is not used for purposes other than auditing and impact assessments, there is more to be done to ensure that it cannot be used for other purposes.

Lastly, around *autonomy/control* of the systems, as we saw in Chapter 4, the use of AI-rec-tech encourages human-in-the-loop or 'hybrid' use. This contributes to knowledge about the fact that human-technological interactions take place when using these systems and highlights that the human perceptual biases will still be present, at least to some extent, in the design and/or use of these systems.

This thesis is also **relevant to work on other AI technologies**, particularly those which will interact with people's skills and behaviours and, therefore, will impact their pursuit of and applications for work. For example, intelligent tutoring systems within schools which help to build skills in a personalised way might change children's skill and behavioural development (Gillani et al., 2023). Skills and behaviours have been shown to be the common way in which candidates are assessed, and so it will be important for

future work to pay attention to how AI in education is changing skill development and whether this correlates to certain schools or demographic groups. In addition, generative AI systems are also changing the development of skills, including coding and computer programming (Yilmaz & Karaoglan Yilmaz, 2023). Not only will these present opportunities for more people to garner these skills, but it might also change the need for employees who exhibit these skills, as these systems can carry out these tasks independently. It will be vital to assess how generative AI is changing skillsets and behaviours, and how this might interact with the use and results of AI-rec-tech assessments.

There is also a contribution to be made to the discussion on **corporate social responsibility and accountability concerning the impact of AI systems** (Adensamer et al., 2021; Broeders & Taylor, 2017). As I showed in Chapter 4, the vendors tended to see themselves as service providers, while ultimately the responsibility for the use of the system, and therefore the outcome, lay with their clients. However, in Chapter 6, I outlined that vendors were using regulatory guidelines to form their views and practices concerning gender bias. Some were even open to more stringent regulation which would direct their pre-implementation and developmental testing.

This highlights a tension between placing the responsibility for system outcome on their clients, but also seeking ways to mitigate harms through their own responsibility. In my opinion, this tension highlights the current lack of accountability placed on AI-rec-tech vendors in existing regulation. As shown in Chapter 6, it is only employers who can be liable for discriminatory outcomes in hiring decisions when AI-rec-tech is used. Vendors are not liable or responsible for these outcomes. Even though they might adopt the 4/5th

guidelines in their framing, perhaps to re-assure their clients, vendors don't always see themselves as responsible for system use and outcomes. However, this does not mean they are not responsible. As I have argued throughout, there should be a shift towards joint responsibility for the impact of these systems, as is encouraged in the EU AI Act and the AAA. Vendors are responsible for the design of the systems, its capabilities, and what is shared with clients surrounding system functionality, even if they might not be responsible for precise or problematic decisions of the client when they use the technology.

Lastly, this thesis contributes to literature surrounding **digital and economic inequalities**. As seen with 'out-of-frame permeability' in Chapter 5, these systems are likely to be to the advantage of those who have been informed about their usage, who have grown up playing games, had access to digital devices of their own, internet access, and time to research these systems and the job, or who are comfortable self-promoting, all of which, together, lend themselves to more privileged circumstances (Rivera, 2016; Ticona, 2021, 2022; Trammell, 2023). These findings therefore support the case that AI-rec-tech is likely favouring the most privileged in these respects, rather than equalising the playing field.

8.2 Limitations

There are several limitations to this study which must be noted. While these limitations are important to consider, I demonstrate why they were manageable, or why the research was worth these limitations.

- **Gender binary:**

This thesis is limited in its engagement with non-binary genders. In many of the discussions, I refer to male and female, as this is the terminology used by vendors or in the research which I cite. While I have considered individuals who identify between or outside of binary gender in my definition of gender (Chapter 1) and in the design of the systems regarding the presentation of gender (Chapter 4), it is a limitation that I don't centre their experiences even more. I hope that future work can further consider the implication of these systems for people who identify between or outside of the gender binary, and that academic research in general might be more open to framing discussions of gender in more inclusive terms.

- **Wider applicability:**

I collected data from eight international AI-rec-tech vendors for this thesis, along with six regulators across the US and EU. These findings are relevant to the companies and regulators within my sample, excepting that in some cases the frames don't apply to all vendors or individuals within these companies. These frames might be applicable to companies outside of this sample and, indeed, I am sure that they are in many ways. However, there is uncertainty of the wider applicability of these findings. This data was so rare to obtain though that regardless of knowing its wider applicability, this study is still extremely valuable.

- **Global North-centric:**

This thesis is limited in its ability to speak to global challenges. Given that the companies in this sample are all based in the Global North, I recognise this work does not do justice to the complexity of these issues in the Global South and in developing economies such as India or China, both of which are global leaders in AI deployment (IBM, 2022). This omission is largely due to there being little information on the use of AI-rec-tech in these countries. I hope that future work can do more to address the use of these systems in the Global South.

- **Interview dynamics - honesty:**

As outlined in Chapter 3, I was limited in knowing how honest participants were being during interviews. It is true that participants might have chosen to reveal certain viewpoints during interviews which may or may not have been truthful. This might have stemmed from a desire to appear in a certain light, or due to organisational restrictions, privacy, and/or possible defensiveness. This could also - in theory - be worsened by the fact they knew I was recording the interview, which might have reduced their willingness to volunteer sensitive information. However, regardless of this limitation, it was still valuable for me to hear the narratives that vendors presented. In addition, I suspect the chances that vendors were dishonest was low. For the most part, these discussions did not touch upon sensitive topics, and all these companies were willing to participate in this study, likely showing they feel they have nothing to hide in the context of this research.

- **Interview dynamics – rapport:**

Initially, I suspected that one limitation of this research would be my own scepticism about the effects of AI-rec-tech. However, the limitation turned out to be quite the opposite: I came to develop high levels of rapport with some of my participants, which I was aware might make me biased in their favour. As Gorman *et al.* (2005) point out, there is the question of selective reporting in these cases. However, I attempted to stay alert to my own biases throughout data collection and analysis, and I have done my best to stay true to the data.

- **Sample – size:**

The sample size of this research is not large. However, as I considered in Chapter 3, obtaining this empirical data was a significant achievement due to the difficulty with accessing these companies. In addition, the data I collected was rich, and therefore while it is important to recognise my sample size, I argue that this does not affect the rigour and implications of my research.

- **Sample – company type:**

While it would have been desirable to interview client companies alongside vendor companies, this was unfortunately not possible due to access difficulty. In Chapter 3, I outlined how I contacted 92 client companies, got responses from ten, and managed to interview four. The four client interviews I did conduct presented me with insufficient data to answer my intended research questions concerning client usage and motivations. This is a limitation as it narrows the data sources; in this research, I always use the same source (e.g. publicly available data from vendor companies, interviews with vendor

companies, conferences with vendor companies), excepting the regulators whose interviews I use in Chapter 6.

However, this allowed me to focus more specifically on vendors instead of attempting to study both clients and vendors. Given the richness of my data, this has been extremely valuable to conversations around design. I hope that future work can engage more with clients, and that clients who use these systems will become more transparent with researchers.

- **Sample – demographic details:**

In terms of the sample, another limitation is that I do not have the demographic breakdown of my interviewees. While this information would have been interesting in many respects, I did not gather this data for anonymity reasons.

- **Sample – vendor companies:**

Vendor companies which agreed to participate are likely those which feel they have nothing to hide and therefore are potentially less controversial. This has been noted as being an issue with qualitative research before; those who say yes to participation are the ones who might take a particular angle (O'Connor et al., 2008). However, given the rarity of accessing vendors, it was still extremely valuable for me to engage with these companies, and this is more of a consideration rather than a certainty.

- **Variations in systems and viewpoints:**

Variations in system design complicates analysis slightly; not all assertions in this thesis apply to every AI-rec-tech system. This also applies to the claims made by some of the participants - in Chapter 4, for example, I saw a range of claims about whether systems can remove or reduce bias. However, while this is a limitation which must be noted (e.g. that not all conclusions apply to all companies or all individuals), the data showed obvious trends and the frames were generally consistent.

- **Confidentiality and anonymisation:**

This research has a mixture of companies which are named and anonymous.

As outlined in Chapter 3, this meant that I had to draft the whole thesis under the assumption that, perhaps, all companies would want to remain anonymous. This meant that I couldn't expand on intricate details about companies and system design as I had to balance the ability to add this extra information with a consistent tone and approach to companies throughout the thesis. However, functioning under this assumption did provide a regularity to how I discussed each company, which was an advantage to drafting the thesis this way. This combination of anonymisation statuses also meant I couldn't include as many screenshots or visual displays as I would otherwise have hoped, but I feel this has been handled well with replicated images and text.

Overall, I think the advantages are worth the limitations on this front. As outlined in Chapter 3, this mixture occurred because I offered vendors the decision about whether to remain anonymous or be named. They made this decision as informed participants,

having seen the full draft of my thesis. This was important due to copyright, as it allowed me to credit the companies who were happy to be named and comply to 'fair use' for those who wished to remain anonymous (Copyright, Designs and Patents Act, Section 30, 1988).

The fact that I have had access to this number of companies is significant, and offering vendors this option was a big re-assurance to participants for various reasons. It catered to the different approaches and wishes of the companies, and therefore this flexibility enabled me to access some companies which otherwise would not have been willing to participate.

8.3 Agenda for future research

Having discussed the contributions of this thesis and outlined its limitations, I round off by looking ahead. Here, I offer a long-term future research agenda which stems largely from the vulnerabilities discussed when explicating the three frames, along with other points which have been raised throughout the thesis. This agenda aims to provide a roadmap for the continuation of research on AI-rec-tech and gender bias surrounding design, regulation, implementation, and effects - or 'framing effects' (Björnehed & Erikson, 2018).

This agenda would require more transparency from various stakeholders, including AI-rec-tech clients. It is likely that clients didn't want to participate in this research due to anxiety about the potential legal implications. However, future work must find solutions to this reticence, for example, with anonymisation of companies and appropriate NDAs.

This research is crucial so that we can learn more about the use and impacts of these systems.

Moreover, this research agenda should aim to broaden its scope by conducting more work in the Global South and in emerging/developing economies. This will help to explore whether, in these contexts, the implications of these technologies are different. While the Global North is often treated as universal, it is crucial for us to increase research on the Global South; AI researchers should play a role in bringing forward perspectives from social positions that have historically been marginalised (Gehl Sampath, 2021; Zajko, 2022). In addition, future research would benefit from being more specific to industry/sectors (e.g. consulting, banking, food, retail, etc.), contract type (e.g. part-time, full time, hourly, zero hours), company size (e.g. small, medium, large), and job type, so that recommendations can be more specific and actionable and so that we might be able to see differences between workers, particularly focusing on those who are most vulnerable.

8.3.1 Client implementation: research within sociology & science and technology studies

The process of clients implementing these systems is still opaque. However, research on this topic will answer important questions around why and how employers are using these systems, and this will be a crucial part of the puzzle to discern the potential impact on candidates.

In cases where clients design the skills or behavioural profile of their desired candidate (as outlined in Chapter 4), future research should explore how this process occurs, how 'top performers' are determined, and what trends (if any) these exhibit demographically. Moreover, it should be explored how clients use the optional features such as anonymisation and re-defining fit. This would help to highlight more about clients' motivations for use (Sloane et al., 2022). Additionally, research should explore to what extent clients use these systems in a 'hybrid' way, e.g. if they use them primarily as a support or whether they rely on them. This will help to inform knowledge on how and why systems are being used and would therefore be helpful in shaping pre- and post-implementation audits for clients.

Separately, as noted in Chapter 1, there is currently little information on the exact scale at which these systems are used. More work needs to be done to investigate the scale at which these systems are being implemented, and this should be calculated by industry, sector, and company size.

8.3.2 Impacts of background, education, and opportunity: research within education

In Chapter 5, I discussed the vulnerability of 'out-of-frame permeability' regarding how external factors such as education and opportunity might impact on candidates so-called soft skills and behaviours. More research is needed to determine how education and socialisation impact the soft skills and behaviours outlined in these tests, particularly along gendered lines and particularly in light of new uses of AI in educational settings. This will help to shed light on how these measurements might correlate to background

and could lead to recommendations on more inclusive ways to measure skills in system design or on policy/curricula in education.

I also pointed out in Chapter 5 how there should be more research to determine the extent to which hard skills still influence people's success on these tests and the industries/sectors/jobs in which this is more prevalent. It would be interesting to explore whether MOOCs intersect with the likelihood of people getting these jobs and, if so, to what extent. This should shed light on professions where there is disparity in relation to hard skills and reflect upon/recommend ways this might be remedied in educational settings. This research is about re-focusing work on the reformation of structures, not just individuals (as argued for in Chapter 5) and might help to shape educational policy that equalises opportunity for all children and adults to have good access to teaching, resources, and a wider curriculum.

8.3.3 Underlying science: research within psychology

In Chapter 5, I discussed how there are 'invisible out-of-frame' vulnerabilities which impact the system design and functionality. Largely, these centre around the underlying 'science' of these tests, with some claiming these tests are based on pseudoscientific principles (Ajunwa, 2021; Sloane et al., 2022). This should explore how vendors are shaping the measurement of skills and behaviours, the underlying psychometric/neuroscientific tests/behavioural tests that these assessments are based upon, and what implications this might have for minority or vulnerable groups. I also discussed this in Chapter 4, where I raised that it is currently unclear what certain skills mean and how they are measured: what does it mean to be creative, risk-averse, or

empathetic? In other words, the epistemological roots of these tests (Sloane et al., 2022) deserve much more attention from psychologists.

8.3.4 Audits and impact assessments: research collaboration between social data science, statistics, and sociology

I pointed out in Chapter 5 that there is a need for socio-technical audits to assess systems along with statistical impact assessments. Future research should aim for collaboration between social data science and sociology to design a standardised socio-technical audit which can be used for each stage of AI-rec-tech (sourcing, screening, assessing, selecting). This will be crucial to capture not just the technical or social aspects of bias in isolation, but to realise the socio-technical nature of bias and the many ways it might be initiated or manifested.

In parallel, future research should address the following areas around statistical impact assessments:

(a) *How vendors are currently performing impact assessments*. This should explore when the tests occur, how they are measuring success rates without hiring information from clients, and how they are choosing which groups to perform these assessments on. Moreover, does this pre-implementation testing make a difference to successful candidates in practice?

(b) *The most effective statistical measurement to assess the impact of these tests on minority/vulnerable groups in context.* Currently, the statistical measurement used

revolves around the 4/5th rule, which, as shown in Chapter 6, could be subject to exceptions. Future research on AI-rec-tech should consider the most effective statistical tests for vendors (pre-implementation if vendors cannot access client datasets) and clients (post-implementation, perhaps in collaboration with vendors). This should consider questions around the adequacy of the 4/5th rule, particularly in these contexts, and consider the 'business necessity' clause. In addition, this work should reflect upon how this measurement might be flexible in different scenarios, for example if only one person is to be hired, in which case perhaps these measurements should capture the workplace or team, rather than just that cohort of employees. Overall, future research should aim to outline the most effective way to measure bias statistically both pre- and post-implementation in terms of how the social framings of bias can be considered within the statistical.

8.3.5 Law and regulation: research within law

In Chapter 6, I outlined how the vendors tend to conceptualise gender bias in relation to the 4/5th rule from the Uniform Guidelines. However, I also discussed how it is possible for the companies to justify a ratio below 0.8 if there is a 'business necessity' or due to small numbers of respondents.

Legal scholars should revisit the guidelines and consider how they could be used and potentially abused in practice. There should be consideration about whether there are ways the Uniform Guidelines might be updated in order to reduce the risk of employers using the 'business necessity' clause to their advantage (Bogen & Rieke, 2018), how the use of justifying low ratios due to 'small numbers' from one group might be weaponised,

and whether these statistical guidelines in the Uniform Guidelines should be updated (see above).

8.3.6 Candidates: research within psychology

As discussed in Chapter 6, it is currently uncertain which candidates are not completing the demographic details prior to tests and how this might change impact assessment ratios. More research is needed to investigate which groups are less likely to fill out these details, why, and to determine ways that candidates might be encouraged to complete this information if they are comfortable to do so. This is crucial so that the impacts of these tests can be detected with more accuracy.

In Chapter 2, I discussed how existing work on candidate perception has found that participants perceive human decisions to be fairer and more just than algorithmic decisions in hiring tasks (Acikgoz et al., 2020; Bauer et al., 2004; Lee, 2018). Future research on candidate preference should focus on shifting this to more practical terms for system designers and administrators of talent acquisition systems (Yarger et al., 2019). For example, it would be interesting to explore why there is a general preference for human interviewers, and whether/how this might reflect problematic aspects of design or user experience which could be remedied. This work should aim to bring the voices of those impacted by these systems into the conversation around design and development. In addition, it would be useful to use this research to consider what materials would be most informative to candidates surrounding bias, processes, functionality, and in what form this information should be provided.

8.3.7 Impact on demographics: research within social data science

Future work should investigate the quantitative impact of the design features on different genders. For example, from Chapter 4, how do re-defining fit or bias reminders impact the diversity of those candidates shortlisted and selected? Similarly, as discussed in Chapter 5, how does question neutralisation and skill/behaviour measurement impact the diversity of candidates shortlisted or selected? This research could be conducted through comparative studies with human recruiters.

Moreover, future research must quantitatively measure the impact of systems regarding diversity in the outcome of hiring decisions. This will help to inform debate on the impact of these systems on workplace diversity, and perhaps in which sectors/industries this might be more prevalent. While there has been one study done of this nature, by Bursell & Roumbanis (2024), this research remains narrow and needs expansion. This type of work should be conducted on a much wider scale, across countries, industries, and company types, to build a comprehensive picture of the impacts of these systems on various demographic groups and to see whether, how, and why they might be impacting more precarious or vulnerable workers in different ways.

8.4 Conclusion

AI-rec-tech systems are only becoming more widespread, and they affect us all in one way or another. They will continue to have an impact the lives and livelihoods of millions. However, research is only beginning to explore AI-rec-tech through empirical data and, as I have shown above, there is a long way to go.

At the beginning of this thesis, I outlined Kyle's story. Kyle was 'red-lighted' for multiple jobs by an AI-driven personality assessment. As Kyle's story demonstrated, these systems deserve our attention, particularly around how they might be affecting minority or vulnerable groups.

This thesis has provided an important contribution to this endeavour. To understand how these systems might affect people from different walks of life, as an initial step, we need to go straight to the source to investigate what underlies and constitutes their design and rationale. This lays the groundwork for future work, which can, and should, investigate the impact of systems in much greater detail.

The future research agenda which I have put forward aims to act as a springboard for more specific and targeted work on the design, use, and impact of AI-rec-tech. Until that research takes place, however, this thesis provides a critical contribution to assessing the design of these AI tools in relation to their messy, complicated social contexts.

REFERENCES

Acemoglu, D., Laibson, D., & List, J. A. (2014). Equalizing superstars: The internet and the democratization of education. *American Economic Review*, *104*(5), 523–527. https://doi.org/10.1257/aer.104.5.523

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, *28*(4), 399–416. https://doi.org/10.1111/ijsa.12306

Acilar, A., & Sæbø, Ø. (2023). Towards understanding the gender digital divide: a systematic literature review. *Global Knowledge, Memory and Communication*, *72*(3), 233–249. https://doi.org/10.1108/GKMC-09-2021-0147

Ada Lovelace Institute. (2020). *Examining the Black Box*. https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf

Adams, J. T. (1931). *The Epic of America*. Routledge.

Adams, R., Hess, U., & Kleck, R. (2015). The Intersection of Gender-Related Facial Appearance and Facial Displays of Emotion. *Emotion Review*, *7*(1), 5–13. https://doi.org/10.1177/1754073914544407

Adams, T., Holman Jones, S. L., & Ellis, C. (2015). *Autoethnography*. Oxford University Press.

Adensamer, A., Gsenger, R., & Klausner, L. D. (2021). "Computer says no": Algorithmic decision support and organisational responsibility. *Journal of Responsible Technology*, *7–8*(June), 1–10. https://doi.org/10.1016/j.jrt.2021.100014

Adler, L. (2021). *Hire with Your Head*. John Wiley & Sons.

Agarwal, P. (2020). *Sway unravelling unconscious bias*. Bloomsbury Sigma.

Ahmed, S. (2012). *On being included: racism and diversity in institutional life*. Duke University Press.

AIHR. (2021). *The Top 22 Virtual HR Conferences to Attend in 2021*. AIHR Digital. https://www.digitalhrtech.com/virtual-hr-conferences/

Ajunwa, I. (2020). The Paradox of Automation as Anti-Bias Intervention. *Cardozo Law Review*, *41*(5), 1671–1742. https://doi.org/http://dx.doi.org/10.2139/ssrn.2746078

Ajunwa, I. (2021). Automated video interviewing as the new phrenology. *Berkeley Technology Law Journal*, *36*(3), 1173–1226. https://doi.org/https://doi.org/10.15779/Z38RX93F1Q

Ajunwa, I. (2023a). *The quantified worker : law and technology in the modern workplace*. Cambridge University Press.

Ajunwa, I. (2023b). *Unfair Automated Hiring Systems Are Everywhere*. Wired. https://www.wired.com/story/unfair-automated-hiring-systems-are-everywhere-ftc/

Ajunwa, I., & Greene, D. (2019). Platforms at work: Automated hiring platforms and other new intermediaries in the organization of work. *Work and Labor in the Digital Age*, *33*, 61–91. https://doi.org/10.1108/S0277-283320190000033005

Ajunwa, I., & Schlund, R. (2020). Algorithms and the Social Organization of Work. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (1st ed., p. 1000). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.52

Alford, R. R. (1998). *The craft of inquiry : theories, methods, evidence*. Oxford University Press.

Algorithmic Accountability Act of 2023, (2023). https://www.congress.gov/bill/118th-congress/house-bill/5628/text?s=1&r=68

Allal-Chérif, O., Yela Aránega, A., & Castaño Sánchez, R. (2021). Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence. *Technological Forecasting and Social Change*, *169*. https://doi.org/10.1016/j.techfore.2021.120822

Altemeyer, B. (2019). Making the business case for AI in HR: two case studies. *Strategic HR Review*, *18*(2), 66–70. https://doi.org/10.1108/shr-12-2018-0101

Anderson, E. (2017). *Private government : how employers rule our lives (and why we don't talk about it)*. Princeton University Press.

Andrews, L., & Bucher, H. (2022). Automating Discrimination: AI Hiring Practices and Gender Inequality. *Cardozo Law Review*, *44*(1), 145–202.

Angwin, J., & Larson, J. (2016). *Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say*. ProPublica. https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say

Balayn, A., & Gürses, S. (2021). *Beyond Debiasing: Regulating AI and its inequalities*. https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf

Bales, R. A., & Stone, K. V. W. (2020). The invisible web at work: Artificial intelligence and electronic surveillance in the workplace. *Berkeley Journal of Employment & Labor Law*, *41*(1), 1–62.

Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context. *Information Systems Frontiers*, *24*(3), 857–875. https://doi.org/10.1007/s10796-021-10223-8

Barger, R. P. (2020). Democratization of Education through Massive Open Online Courses in Asia. *IAFOR Journal of Education: Technology in Education*, *8*(2), 29–46.

Bargh, B. (2022). *Estée Lauder settles out of court with staff who were 'sacked by algorithm.'* Cosmetics Business. https://cosmeticsbusiness.com/est-e-lauder-settles-out-of-court-with-staff-who-were-sacked-by-algorithm--199624

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, *104*(3), 671–732. https://doi.org/10.15779/Z38BG31

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Bastian, R. (2019). Why We Need To Stop Talking About Diversity of Thought. *Forbes*. https://www.forbes.com/sites/rebekahbastian/2019/05/13/why-we-need-to-stop-talking-about-diversity-of-thought/

Bateson, G. (1955). A theory of play and fantasty. *Psychiatric Research Reports*, *2*, 39–51.

Bauer, T., Donald, T., Kyle, M., & Ana, C. (2011). Applicant reactions to technology-based selection. In T. Nancy & A. Seymour (Eds.), *Technology-Enhanced Assessment of Talent* (pp. 120–223). Jossey-Bass.

Bauer, T., Truxillo, D. M., Paronto, M. E., Weekley, J. A., & Campion, M. A. (2004). Applicant Reactions to Different Selection Technology: Face-to-Face, Interactive Voice Response, and Computer-Assisted Telephone Screening Interviews. *International Journal of Selection and Assessment*, *12*(1–2), 135–148. https://doi.org/10.1111/j.0965-075X.2004.00269.x

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, *23*(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo* (2nd ed.). Sage Publications.

Beitin, B. K. (2012). Interview and Sampling: How Many and Whom. In J. F. Gubrium, J. A. Holstein, A. Marvasti, & K. D. McKinney (Eds.), *The SAGE Handbook of Interview Research: The Complexity of the Craft* (2nd ed., pp. 243–254). SAGE Publications, Inc. https://doi.org/10.4135/9781452218403

Benjamin, R. (2019). *Race after technology: abolitionist tools for the new Jim code*. Polity Press.

Bergen, H. (2016). 'I'd Blush if I Could': Digital Assistants, Disembodied Cyborgs and the Problem of Gender. *Word and Text*, *VI*(1), 95–113.

Berger, R. (2015). Now I see it, now I don't: researcher's position and reflexivity in qualitative research. *Qualitative Research*, *15*(2), 219–234. https://doi.org/10.1177/1468794112468475

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*,

21(1), 33–46. https://doi.org/10.1007/s11092-008-9064-9
Biesta, G. (2015). What is education for? On Good education, teacher judgement, and educational professionalism. *European Journal of Education*, *50*(1), 75–87. https://doi.org/10.1111/ejed.12109
Binns, R., Gilbert, A., Imafidon, A.-M., Johnston, T., Simons, J., Mountfield, H., & Thomas, A. (2020). *Mind the gap: how to fill the equality and AI accountability gap in an automated world*. https://uploads-ssl.webflow.com/5f57d40eb1c2ef22d8a8ca7e/5fad57ea4869b00f399bd2c3_IFOW-ETF-Mind the gap (v9-12.11.20).pdf
Bivens, R. (2017). The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. *New Media & Society*, *19*(6), 880–898. https://doi.org/10.1177/1461444815621527
Bivens, R., & Haimson, O. L. (2016). Baking Gender Into Social Media Design: How Platforms Shape Categories for Users and Advertisers. *Social Media + Society*, *2*(4), 1–12. https://doi.org/10.1177/2056305116672486
Björnehed, E., & Erikson, J. (2018). Making the most of the frame: developing the analytical potential of frame analysis. *Policy Studies*, *39*(2), 109–126. https://doi.org/10.1080/01442872.2018.1434874
Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it? *Business Horizons*, *63*(2), 215–226. https://doi.org/10.1016/j.bushor.2019.12.001
Black, J. S., & van Esch, P. (2021). AI-enabled recruiting in the war for talent. *Business Horizons*, *64*(4), 513–524. https://doi.org/10.1016/j.bushor.2021.02.015
Blacksmith, N., Willford, J., & Behrend, T. (2016). Technology in the Employment Interview: A Meta-Analysis and Future Research Agenda. *Personnel Assessment and Decisions*, *2*(1), 12–20. https://doi.org/10.25035/pad.2016.002
BLS. (2024). *USUAL WEEKLY EARNINGS OF WAGE AND SALARY WORKERS FOURTH QUARTER 2023*. https://www.bls.gov/news.release/pdf/wkyeng.pdf
Bogen, M., & Rieke, A. (2018). *Help wanted: An examination of hiring algorithms, equity, and bias*. https://doi.org/10.1007/978-1-349-20745-9_2
Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *30th Conference on Neural Information Processing Systems (NIPS 2016)*, 1–9. https://doi.org/https://doi.org/10.48550/arXiv.1607.06520
Boucher, A. (2017). Power in elite interviewing: Lessons from feminist studies for political science. *Women's Studies International Forum*, *62*, 99–106. https://doi.org/10.1016/j.wsif.2017.05.003
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, *13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846
Brenner, M. (1985). Intensive Interviewing. In M. Brenner, J. Brown, & D. V Canter (Eds.), *The research interview : uses and approaches*. Academic Press.
Bringer, J., Johnston, L., & Brackenridge, C. (2004). Maximizing Transparency in a Doctoral Thesis 1: The Complexities of Writing About the Use of QSR*NVIVO Within a Grounded Theory Study. Qualitative Research. *Qualitative Research*, *4*(2), 247–265. https://doi.org/https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1177/1468794104044434
Broeders, D., & Taylor, L. (2017). Does Great Power Come with Great Responsibility? The Need to Talk About Corporate Political Responsibility. In M. Taddeo & L. Floridi (Eds.), *The Responsibilities of Online Service Providers* (pp. 315–323). Springer. https://doi.org/10.1007/978-3-319-47852-4_17
Broussard, M. (2023). *More Than a Glitch*. MIT Press.
Bukodi, E., & Goldthorpe, J. H. (2019). *Social mobility and education in Britain : research, politics and policy*. Cambridge University Press.

Bunz, M., & Braghieri, M. (2022). The AI doctor will see you now: assessing the framing of AI in news coverage. *AI and Society*, *37*(1), 9–22. https://doi.org/10.1007/s00146-021-01145-9

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR*, *81*, 77–91.

Burke, P. J. (2012). *The right to higher education: beyond widening participation*. Routledge.

Bursell, M., & Roumbanis, L. (2024). After the algorithms: A study of meta-algorithmic judgments and diversity in the hiring process at a large multisite company. *Big Data and Society*, *11*(1), 1–18. https://doi.org/10.1177/20539517231221758

Bustelo, M., Suaya, A., & Viollaz, M. (2019). *The future of work in Latin America and the Caribbean*. https://doi.org/10.1144/M33.9

Butler, J. (1990). *Gender trouble : feminism and the subversion of identity*. Routledge.

Butler, J. (1993). Critically Queer. *GLQ: A Journal of Lesbian and Gay Studies*, *1*(1), 17–32. https://doi.org/10.1215/10642684-1-1-17

Butler, J. (2011). *Bodies that matter: on the discursive limits of "sex."* Routledge.

Byrne, D., Griffitt, W., Hudgins, W., & Reeves, K. (1969). Attitude Similarity-Dissimilarity and Attraction: Generality Beyond the College Sophomore. *The Journal of Social Psychology*, *79*(2), 155–161. https://doi.org/10.1080/00224545.1969.9922403

Calarco, J. M. (2018). *Negotiating opportunities : how the middle class secures advantages in school*. Oxford University Press.

Carli, L. L., Alawa, L., Lee, Y. A., Zhao, B., & Kim, E. (2016). Stereotypes About Gender and Science: Women ≠ Scientists. *Psychology of Women Quarterly*, *40*(2), 244–260. https://doi.org/10.1177/0361684315622645

Carlsson, M., & Eriksson, S. (2019). In-group gender bias in hiring: Real-world evidence. *Economics Letters*, *185*, 1–3. https://doi.org/10.1016/j.econlet.2019.108686

Carlsson, R., & Sinclair, S. (2018). Prototypes and same-gender bias in perceptions of hiring discrimination. *Journal of Social Psychology*, *158*(3), 285–297. https://doi.org/10.1080/00224545.2017.1341374

Carnegie, M. (2023). Breaking the office bias: smashing the stereotypes holding women back. *The Independent*. https://www.independent.co.uk/life-style/galaxy-the-ripple-effect/breaking-bias-workplace-stereotypes-b2366688.html

Carter, D. (2016). Hustle and Brand: The Sociotechnical Shaping of Influence. *Social Media and Society*, *2*(3), 1–12. https://doi.org/10.1177/2056305116666305

Catalyst. (2020). *Women Business Leaders: Global Statistics*. https://www.catalyst.org/research/women-in-management/

Cerón-Anaya, H. (2019). *Privilege at Play: Class, Race, Gender, and Golf in Mexico*. Oxford University Press. https://doi.org/10.1093/oso/9780190931605.001.0001

Chamorro-Premuzic, T. (2018). The Case For Boosting Cognitive Diversity In Organizations. *Forbes*. https://www.forbes.com/sites/tomaspremuzic/2018/09/03/the-case-for-boosting-cognitive-diversity-in-organizations/

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? *Industrial and Organizational Psychology*, *9*(3), 621–640. https://doi.org/10.1017/iop.2016.6

Charlwood, A., & Guenole, N. (2022). Can HR adapt to the paradoxes of artificial intelligence? *Human Resource Management Journal*, *December*, 1–14. https://doi.org/10.1111/1748-8583.12433

Clark, D. (2023a). *A level pass rates in the United Kingdom in 2023, by grade and gender*. Statista. https://www.statista.com/statistics/282976/a-level-results-in-the-uk-by-gender/#:~:text=In the United Kingdom in,24.3 percent of male students.

Clark, D. (2023b). *Proportion of GCSE entries awarded a C/4 or higher in the United Kingdom between 2000 and 2023, by gender*. Statista. https://www.statista.com/statistics/282484/gcse-pass-rate-in-uk-by-gender/#:~:text=Proportion of GCSE entries that,UK 2000-2023%2C by gender&text=In 2023%2C 71.7 percent of,level in the United Kingdom.

Code Signal. (2020). *How to Apply Data-Driven Recruiting to Hiring Engineers*. https://codesignal.com/resource/how-to-apply-data-driven-recruiting-to-hiring-engineers/
Code Signal. (2022). *About Us*. Code Signal Website. codesignal.com/about/
Code Signal. (2024). *Pre-screen*. Code Signal Website. https://codesignal.com/products/pre-screen/
Collett, C. (2023). The Hustle: How Struggling to Access Elites for Qualitative Interviews Alters Research and the Researcher. *Qualitative Inquiry*, 1–13. https://doi.org/10.1177/10778004231188054
Collett, C., & Dillon, S. (2019). AI and Gender Four Proposals for Future Research. *Cambridge: The Leverhulme Centre for the Future of Intelligence*. http://lcfi.ac.uk/resources/ai-and-gender-four-proposals-future-research/
Commission, E. (2024a). *About the European Commission*. European Commission. https://commission.europa.eu/about-european-commission_en
Commission, E. (2024b). *AI Act*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
Cools, H., Van Gorp, B., & Opgenhaffen, M. (2022). Where exactly between utopia and dystopia? A framing analysis of AI and automation in US newspapers. *Journalism*, *25*(1), 1–19. https://doi.org/10.1177/14648849221122647
Copyright, Designs and Patents Act, Section 30, (1988). https://www.legislation.gov.uk/ukpga/1988/48/section/30
Craig, L., & Churchill, B. (2021). Dual-earner parent couples' work and care during COVID-19. *Gender, Work and Organization*, *28*(S1), 66–79. https://doi.org/10.1111/gwao.12497
Crawford, K. (2021). *The Atlas of AI*. Yale University Press. https://doi.org/10.2307/j.ctv1ghv45t
Crawford, K., Dobbe, R., Dryer, T., Fried, G., & Al., E. (2019). *AI Now Report 2019*. https://ainowinstitute.org/AI_Now_2019_Report.pdf
Crawley, S. L. (2012). Autoethnography as Feminist Self-interview. In *The SAGE Handbook of Interview Research: The Complexity of the Craft* (2nd ed.). SAGE Publications, Inc. https://doi.org/10.4135/9781452218403
Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *The University of Chicago Legal Forum*, *1989*(1), 139–168.
Cukier, K., Mayer-Schönberger, V., & Véricourt, F. de. (2021). *Framers: human advantage in an age of technology and turmoil*. Dutton, Penguin Random House LLC.
D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.
Darling-Hammond, L. (2001). Inequality in Teaching and Schooling: How Opportunity Is Rationed to Students of Color in America. In B. Smedley, A. Stith, & L. Colburn (Eds.), *The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in the Health Professions*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK223640/
Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
David, C. C., Atun, J. M., Fille, E., & Monterola, C. (2011). Finding Frames: Comparing Two Methods of Frame Analysis. *Communication Methods and Measures*, *5*(4), 329–351. https://doi.org/10.1080/19312458.2011.624873
DCWP. (2023). *Automated Employment Decision Tools: Frequently Asked Questions*. https://www.nyc.gov/assets/dca/downloads/pdf/about/DCWP-AEDT-FAQ.pdf
De Soete, B., Lievens, F., & Druart, C. (2013). Strategies for dealing with the diversity-validity dilemma in personnel selection: Where are we and where should we go? *Journal of Work and Organizational Psychology*, *29*(1), 3–12. https://doi.org/http://dx.doi.org/10.5093/tr2013a2
Deakin, H., & Wakefield, K. (2014). Skype interviewing: reflections of two PhD researchers. *Qualitative Research*, *14*(5), 603–616. https://doi.org/10.1177/1468794113488126
Debusmann, B. (2023). *Affirmative action: US Supreme Court overturns race-based college*

admissions. BBC. https://www.bbc.co.uk/news/world-us-canada-65886212#:~:text=The US Supreme Court has,contentious issues in US education.

Dencik, L., & Stevens, S. (2021). Regimes of justification in the datafied workplace: The case of hiring. *New Media & Society*, *25*(12), 3657–3675. https://doi.org/10.1177/14614448211052893

Dennis, M. R., & Kunkel, A. D. (2004). Perceptions of men, women, and CEOs: The Effects of Gender Identity. *Social Behavior and Personality: An International Journal*, *32*(2), 155–171. https://doi.org/10.2224/sbp.2004.32.2.155

Deutsch, N. L. (2004). Positionality and the pen: Reflections on the process of becoming a feminist researcher and writer. *Qualitative Inquiry*, *10*(6), 885–902. https://doi.org/10.1177/1077800404265723

Dhapte, A. (2024). *AI recruitment market research report information by component*. Market Research Future. https://www.marketresearchfuture.com/reports/ai-recruitment-market-8289

Dickson, D. R., & Nusair, K. (2010). An HR perspective: The global hunt for talent in the digital age. *Worldwide Hospitality and Tourism Themes*, *2*(1), 86–93. https://doi.org/10.1108/17554211011012612

Dillon, S. (2020). The Eliza effect and its dangers: from demystification to gender critique. *Journal for Cultural Research*, *24*(1), 1–15. https://doi.org/10.1080/14797585.2020.1754642

Dilmaghani, M. (2021). The gender gap in competitive chess across countries: Commanding queens in command economies. *Journal of Comparative Economics*, *49*(2), 425–441. https://doi.org/10.1016/j.jce.2020.09.006

Dobbin, F., & Kalev, A. (2016). *Why Diversity Programs Fail*. HBR. https://hbr.org/2016/07/why-diversity-programs-fail

Domina, T., Penner, A., & Penner, E. (2017). Categorical Inequality: Schools As Sorting Machines. *Annual Review of Sociology*, *43*(1), 311–330. https://doi.org/10.1146/annurev-soc-060116-053354

Donahoo, S. (2023). Working with style: Black women, black hair, and professionalism. *Gender, Work and Organization*, *30*(2), 596–611. https://doi.org/10.1111/gwao.12838

Drage, E., & Mackereth, K. (2022). Does AI Debias Recruitment? Race, Gender, and AI's "Eradication of Difference." *Philosophy and Technology*, *35*(4), 1–25. https://doi.org/10.1007/s13347-022-00543-1

DSIT. (2023). *A pro-innovation approach to AI regulation*. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Dunham, C. R. (2017). Third generation discrimination: the ripple effects of gender bias in the workplace. *Akron L Review*, *51*(1), 55–98. https://doi.org/https://dx.doi.org/10.2139/ssrn.3019163

Durkheim, É. (1956). *Education and sociology*. Free Press.

Duttlinger, C. (2022). *Attention and distraction in modern German literature, thought, and culture* (1st ed.). Oxford University Press.

Dutz, R., Hubner, S., & Peus, C. (2022). When agency "fits" regardless of gender: Perceptions of applicant fit when job and organization signal male stereotypes. *Personnel Psychology*, *75*(2), 441–483. https://doi.org/10.1111/peps.12470

Eagly, A. H., & Karau, S. J. (2002). Role Congruity Theory of Prejudice Toward Female Leaders. *Psychological Review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (pp. 458–476). SAGE Publications Ltd. https://doi.org/10.4135/9781446249222.n49

Edel, A. (1979). *Analyzing concepts in social science*. Transaction Books.

Edo, A., Jacquemet, N., & Yannelis, C. (2019). Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Review of Economics of the Household*, *17*(1), 349–376. https://doi.org/10.1007/s11150-017-9391-z

Edwards, R., & Holland, J. (2013). *What is qualitative interviewing?* Bloomsbury Academic.
EEOC. (1978). Part 1607 - Uniform Guidelines on Employee Selection Procedures. In *Title 29 - Labor. Subtitle B - Regulations Relating to Labor (Continued). CHAPTER XIV - EQUAL EMPLOYMENT OPPORTUNITY COMMISSION.* https://www.govinfo.gov/content/pkg/CFR-2018-title29-vol4/xml/CFR-2018-title29-vol4-part1607.xml
EEOC. (2024a). *Artificial Intelligence and Algorithmic Fairness Initiative*. US Equal Employment Opportunity Commission. https://www.eeoc.gov/ai
EEOC. (2024b). *Frequently Asked Questions*. https://www.eeoc.gov/youth/frequently-asked-questions#Q1
EEOC. (2024c). *iTutorGroup to Pay $365,000 to Settle EEOC Discriminatory Hiring Suit*. US Equal Employment Opportunity Commission. https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit
EEOC. (2024d). *Keith E. Sonderling, Commissioner*. US Equal Employment Opportunity Commission. https://www.eeoc.gov/keith-e-sonderling-commissioner
EEOC. (2024e). *Overview*. US Equal Employment Opportunity Commission. https://www.eeoc.gov/overview
EEOC. (2024f). *The Commission and the General Counsel*. US Equal Employment Opportunity Commission. https://www.eeoc.gov/commission
Empson, L. (2017). Elite interviewing in professional organizations. *Journal of Professions and Organization*, *5*(1), 58–69. https://doi.org/10.1093/jpo/jox010
Espinosa, M. P., & Ferreira, E. (2022). Gender implicit bias and glass ceiling effects. *Journal of Applied Economics*, *25*(1), 37–57. https://doi.org/10.1080/15140326.2021.2007723
Eubanks, V. (2017). *Automating inequality : how high-tech tools profile, police, and punish the poor*. St. Martin's Press.
European Commission. (2021). *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS COM/2021/206 final*. European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
European Parliament. (2020). *Precarious work from a gender and intersectionality perspective, and ways to combat it* (Issue November). https://www.europarl.europa.eu/RegData/etudes/STUD/2020/662491/IPOL_STU(2020)662491_EN.pdf
European Parliament. (2023). *MEPs ready to negiotiate first ever rules for safe and transparent AI*. News. https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai
Evans, L., & Davies, K. (2000). No sissy boys here: A content analysis of the representation of masculinity in elementary school reading textbooks. *Sex Roles*, *42*(3–4), 255–270. https://doi.org/10.1023/A:1007043323906
Exley, C. L., & Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, *137*(3), 1345–1381. https://doi.org/10.1093/qje/qjac003
Eynon, R., Fry, J., & Schroeder, R. (2008). The ethics of internet research. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE Handbook of Online Research Methods* (pp. 22–41). SAGE Publications.
Eynon, R. (2022). DATAFICATION AND THE ROLE OF SCHOOLING: Challenging the status quo. In L. Pangrazio & J. Sefton-Green (Eds.), *Learning to Live with Datafication: Educational Case Studies and Initiatives from Across the World* (1st ed., pp. 17–34). Routledge. https://doi.org/10.4324/9781003136842-2
Eynon, R. (2023). Utilising a Critical Realist Lens to Conceptualise Digital Inequality: The Experiences of Less Well-Off Internet Users. *Social Science Computer Review*, *41*(3), 1081–1096. https://doi.org/10.1177/08944393211069662
Fabris, A., Baranowska, N., Dennis, M. J., Hacker, P., Saldivar, J., Saldivar, F. Z., & Biega, A. J. (2023). Fairness and Bias in Algorithmic Hiring. *ACM 2023*, *September*, 1–48.

https://doi.org/10.48550/arxiv.2309.13933
Faragó, F., Savoy, S., Sanchez, E. A., Khaleghi, N., Evans, L., Adams, E. L., & Palomino, E. (2022). Gender Typicality, Pressure to Conform to Gender Norms, and Anti-Fat and Appearance Stereotypes in Girls. *Archives of Sexual Behavior*, *51*(4), 2183–2198. https://doi.org/10.1007/s10508-021-02252-w
Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics*, *26*(6), 3333–3361. https://doi.org/10.1007/s11948-020-00276-4
Fisher, K. (1997). Locating Frames in the Discursive Universe. *Sociological Research*, *2*(3), 1–31. https://doi.org/10.5153/sro.78
Fiske, S. T. (2002). What we know now about bias and intergroup conflict, the problem of the century. *Current Directions in Psychological Science*, *11*(4), 123–128. https://doi.org/10.1111/1467-8721.00183
Florentine, S. (2016). *How artificial intelligence can eliminate bias in hiring*. CIO. https://www.cio.com/article/3152798/how-artificial-intelligence-can-eliminate-bias-in-hiring.html
Flores, K. (2021). *The Importance of Soft Skills & Hard Skills in Workforce Transformation*. Pymetrics Pygest Blog. https://www.pymetrics.ai/pygest/soft-hard-skills-workplace-transformation
Foley, L. J. (2012). Constructing the Respondent. In J. F. Gubrium, J. A. Holstein, A. B. McKinney, & M. & K. D. (Eds.), *The SAGE Handbook of Interview Research: The Complexity of the Craft* (2nd ed., pp. 305–316). SAGE Publications, Inc. https://doi.org/https://doi.org/10.4135/9781452218403
Forman, A., Glasser, N., & Lech, C. (2020). INSIGHT: Covid-19 May Push More Companies to Use AI as Hiring Tool. *Bloomberg Law*. https://news.bloomberglaw.com/daily-labor-report/insight-covid-19-may-push-more-companies-to-use-ai-as-hiring-tool
Forsythe, D. (1993). Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. *Social Studies of Science*, *23*(3), 445–477. https://doi.org/10.1177/0306312793023003002
Foucault, M. (1991). *Discipline and punish: the birth of the prison*. Harmondsworth: Penguin Books.
Friedman, B., & Nissenbaum, H. (1996). Bias in Computer Systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561
Friedmann, E., & Efrat-Treister, D. (2023). Gender Bias in Stem Hiring: Implicit In-Group Gender Favoritism Among Men Managers. *Gender and Society*, *37*(1), 32–64. https://doi.org/10.1177/08912432221137910
FTC. (2024). *About The FTC*. Federal Trade Commission. https://www.ftc.gov
Fuchs, L. (2023). Hired by a Machine: Can a New York City Law Enforce Algorithmic Fairness in Hiring Practices? *Fordham Journal of Corporate & Financial Law*, *28*(1), 185–222.
Gala, D. S., & Philips-Brown, M. (2024). *FairTargetSim: Interactive Variable Definition Simulator*. Hugging Face. https://huggingface.co/spaces/DaliaGala/FairTargetSim
Gallego, A., Krentz, M., Tsusaka, M., Yousif, N., & Brooks Taplett, F. (2019). *How AI Could Help—or Hinder—Women in the Workforce*. BCG. https://www.bcg.com/en-gb/publications/2019/artificial-intelligence-ai-help-hinder-women-workforce
Galos, D. R., & Coppock, A. (2023). Gender composition predicts gender bias: A meta-reanalysis of hiring discrimination audit experiment. *Science Advances*, *9*(18), 1–11. https://doi.org/10.1126/sciadv.ade7979
Gamson, W. A. (1992). *Talking politics*. Cambridge University Press.
Gartner. (2020). *Gartner HR Survey Shows 86% of Organizations Are Conducting Virtual Interviews to Hire Candidates During Coronavirus Pandemic*. Gartner. https://www.gartner.com/en/newsroom/press-releases/2020-04-30-gartner-hr-survey-shows-86--of-organizations-are-cond
Gehl Sampath, P. (2021). Governing Artificial Intelligence in an Age of Inequality. *Global Policy*, *12*(S6), 21–31. https://doi.org/10.1111/1758-5899.12940
Gerson, K., & Damaske, S. (2020). *The Science and Art of Interviewing*. Oxford University

Press, Incorporated. https://doi.org/10.1093/oso/9780199324286.001.0001
Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21
Gillani, N., & Eynon, R. (2023). Understanding and Improving Social Factors in Education: A Computational Social Science Approach. In T. Yasseri (Ed.), *Handbook of Computational Social Science*. Edward Elgar Publishing. https://doi.org/10.2139/ssrn.4323861
Gillani, N., Eynon, R., Chiabaut, C., & Finkel, K. (2023). Unpacking the "Black Box" of AI in Education. *Educational Technology and Society*, *26*(1), 99–111. https://doi.org/10.30191/ETS.202301_26(1).0008
Gillespie, T. (2014). The Relevance of Algorithms: Essays on Communication, Materiality, and Society. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies*. https://doi.org/https://doi.org/10.7551/mitpress/9780262525374.003.0009
Gillis, T. B., & Spiess, J. L. (2019). Big Data and Discrimination. *The University of Chicago Law Review*, *86*(2), 459–487.
Giurge, L. M., Whillans, A. V., & Yemiscigil, A. (2021). A multicountry perspective on gender differences in time use during COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(12), 1–7. https://doi.org/10.1073/pnas.2018494118
Goffman, E. (1974). *Frame analysis: an essay on the organization of experience*. Northeastern University Press.
Goldin, C., & Rouse, C. (1997). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. In *National Bureau of Economic Research Working Paper Series*. https://doi.org/10.4324/9780429499821-66
González, M. J. C., Cortina, C., & Rodríguez, J. (2019). The role of gender stereotypes in hiring: A field experiment. *European Sociological Review*, *35*(2), 187–204. https://doi.org/10.1093/esr/jcy055
Gorman, G. E., Clayton, P., Shep, S. J., & Clayton, A. (2005). *Qualitative research for the information professional : a practical handbook* (2nd ed.). Facet.
Gorska, A. M., & Jemielniak, D. (2023). The invisible women: uncovering gender bias in AI-generated images of professionals. *Feminist Media Studies*, *23*(8), 4370–4375. https://doi.org/10.1080/14680777.2023.2263659
Goyette, K. A. (2017). *Education in America*. University of California Press.
Grant, G. (2018). Similar-To-Me Bias: How Gender Affects Workplace Recognition. *Forbes*. https://www.forbes.com/sites/georginagrant/2018/08/07/similar-to-me-bias-how-gender-affects-workplace-recognition/
Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*(4), 945–967. https://doi.org/10.2307/20439056
Grigoropoulou, N., & Small, M. L. (2022). The data revolution in social science needs qualitative research. *Nature Human Behaviour*, *6*(7), 904–906. https://doi.org/10.1038/s41562-022-01333-7
Grugulis, I., & Vincent, S. (2009). Whose skill is it anyway?: "Soft" skills and polarization. *Work, Employment and Society*, *23*(4), 597–615. https://doi.org/10.1177/0950017009344862
Guenole, N., & Feinzig, S. (2018). *The Business Case for AI in HR With Insights and Tips on Getting Started*. https://research.gold.ac.uk/id/eprint/33662/
Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903
Guest, G., MacQueen, K. M., & Namey, E. (2012). *Applied Thematic Analysis*. SAGE Publications, Inc. https://doi.org/10.4135/9781483384436
Guyan, K. (2022). *Queer data: using gender, sex and sexuality data for action*. Bloomsbury Academic.
Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, *47*(7), 1264–1279. https://doi.org/10.1108/OIR-08-2021-0452

Halpin, B. W., & Smith, V. (2019). Recruitment: an undertheorized mechanism for workplace control. *Theory and Society*, *48*(5), 709–732. https://doi.org/10.1007/s11186-019-09362-4
Hammersley, M. (2008). *Questioning qualitative inquiry: critical essays*. SAGE.
Harris, C. G. (2018). Making better job hiring decisions using "Human in the loop" techniques. *CEUR Workshop Proceedings*, 16–26. https://ceur-ws.org/Vol-2169/paper-03.pdf
Harvey, W. S. (2011). Strategies for conducting elite interviews. *Qualitative Research*, *11*(4), 431–441. https://doi.org/10.1177/1468794111404329
Harwell, D. (2019a). HireVue faces FTC complaint over 'unfair and deceptive' practices. *The Washington Post*. https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices/
Harwell, D. (2019b). Rights group files federal complaint against AI-hiring firm HireVue, citing 'unfair and deceptive' practices. *The Washington Post*. https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices/
Haselton, M. G., Nettle, D., & Andrews, P. W. (2015). The Evolution of Cognitive Bias. In D. M. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 724–746). John Wiley & Sons. https://doi.org/10.1002/9780470939376.ch25
Hawkins, T. T. (2021). *Persona: The Dark Truth Behind Personality Tests*.
Hay-Gibson, N. V. (2009). Interviews via VoIP: Benefits and Disadvantages within a PhD study of SMEs. *Library and Information Research*, *33*(105), 39–50. https://doi.org/10.29173/lirg111
Heather. (2018). *30 Best Recruiting Conferences To Attend In 2019*. Harver. https://harver.com/blog/recruiting-conferences-2019/
Heidari, H., Barocas, S., Kleinberg, J., & Levy, K. (2023). Informational Diversity and Affinity Bias in Team Growth Dynamics. *EAAMO '23: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–10. https://doi.org/10.48550/arxiv.2301.12091
Heilman, M. E. (1983). Sex bias in work settings: The Lack of Fit model. *Research in Organizational Behavior*, *5*, 269–298.
Heilman, M. E. (2001). Description and Prescription: How Gender Stereotypes Prevent Women's Ascent Up the Organizational Ladder. *Journal of Social Issues*, *57*(4), 657–674. https://doi.org/10.1111/0022-4537.00234
Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135. https://doi.org/10.1016/j.riob.2012.11.003
Heilman, M. E., & Caleo, S. (2018). Combatting gender discrimination: A lack of fit framework. *Group Processes and Intergroup Relations*, *21*(5), 725–744. https://doi.org/10.1177/1368430218761587
Heilman, M. E., Caleo, S., & Manzi, F. (2024). Women at Work: Pathways from Gender Stereotypes to Gender Bias and Discrimination. *Annual Review of Organizational Psychology and Organizational Behavior*, *11*(1), 1–28. https://doi.org/10.1146/annurev-orgpsych-110721-034105
Heilman, M. E., & Haynes, M. C. (2008). Subjectivity in the Appraisal Process: A Facilitator of Gender Bias in Work Settings. In E. Borgida & S. T. Fiske (Eds.), *Beyond common sense : psychological science in the courtroom* (pp. 127–155). Blackwell.
Hemalatha, A., Kumari, P. B., Nawaz, N., & Gajenderan, V. (2021). Impact of Artificial Intelligence on Recruitment and Selection of Information Technology Companies. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021*, 60–66. https://doi.org/10.1109/ICAIS50930.2021.9396036
Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science and Medicine*, *292*, 114523. https://doi.org/10.1016/j.socscimed.2021.114523
Henry, D. (2022). *Computer Says No*. BBC. https://www.bbc.co.uk/iplayer/episode/m0015gvw/computer-says-no
Hertz, R., & Imber, J. (1995). *Studying Elites Using Qualitative Methods* (1st ed.). SAGE.

https://doi.org/10.4135/9781483327341
Higgins, C. A., & Judge, T. A. (2004). The effect of applicant influence tactics on recruiter perceptions of fit and hiring recommendations: a field study. *Journal of Applied Psychology*, *89*(4), 622. https://doi.org/10.1037/0021-9010.89.4.622
Hill Collins, P. (1990). *Black feminist thought : knowledge, consciousness, and the politics of empowerment*. Routledge.
Hine, C. (2015). *Ethnography for the internet: embedded, embodied and everyday* (1st ed.). Routledge.
Hmoud, B., & Varallyai, L. (2019). WILL ARTIFICIAL INTELLIGENCE TAKE OVER HUMAN RESOURCES RECRUITMENT AND SELECTION? *Network Intelligence Studies*, *VII*(13), 21–30.
Hochschild, J. L., & Scovronick, N. B. (2003). *The American dream and the public schools*. Oxford University Press.
Hogan, L. J. (2019). *Labor and Employment – Use of Facial Recognition Services – Prohibition* (Vol. 446, pp. 1–2). https://mgaleg.maryland.gov/mgawebsite/Legislation/Details/HB1202?ys=2020RS
Holbein, K. (2020). *2020 Top Talent Acquisition and Recruitment Marketing Conferences*. Rally. https://rallyrecruitmentmarketing.com/2019/11/top-talent-acquisition-and-recruitment-marketing-conferences-2020/
Holstein, J. A., & Gubrium, J. F. (1995). *The Active Interview*. SAGE Publications. https://doi.org/10.4135/9781412986120
Hong, R. (2016). Soft skills and hard numbers: Gender discourse in human resources. *Big Data and Society*, *3*(2), 1–13. https://doi.org/10.1177/2053951716674237
Hoover, A. E., Hack, T., Garcia, A. L., Goodfriend, W., & Habashi, M. M. (2019). Powerless Men and Agentic Women: Gender Bias in Hiring Decisions. *Sex Roles*, *80*(11–12), 667–680. https://doi.org/10.1007/s11199-018-0964-y
Hopkins, P. D. (2009). The Intersection of Culture, Gender, and Technology. In D. G. Johnson & J. M. Wetmore (Eds.), *Technology and society : building our sociotechnical future*. MIT Press.
Hoyt, C. L., Burnette, J. L., Forsyth, R. B., Parry, M., & DeShields, B. H. (2021). Believing in the American Dream Sustains Negative Attitudes toward Those in Poverty. *Social Psychology Quarterly*, *84*(3), 203–215. https://doi.org/10.1177/01902725211022319
Hughes, T. P. (1987). The Evolution of Large Technological Systems. In W. E. Bijker, T. P. Hughes, T. J. Pinch, & D. G. Douglas (Eds.), *The social construction of technological systems: new directions in the sociology and history of technology* (pp. 45–76). MIT Press.
Hutson, J., Valenzuela, M., Hosto-Marti, B., & Wright, S. (2023). The Role of Higher Education in Developing Durable Skills: Reframing General Education. *Journal of Higher Education Theory and Practice*, *23*(3), 1–12. https://doi.org/10.33423/jhetp.v23i3.5836
IBM. (2022). IBM Global AI Adoption Index 2022. In *IBM*. https://www.ibm.com/watson/resources/ai-adoption
IFOW. (2020). *Artificial intelligence in hiring: Assessing impacts on equality*. https://www.ifow.org/publications/artificial-intelligence-in-hiring-assessing-impacts-on-equality
Artificial Intelligence Video Interview Act, Pub. L. No. (820 ILCS 42/), Illinois Compiled Statutes (2020). https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68
ILO. (2019). *Understanding the gender pay gap*. https://www.ilo.org/publications/understanding-gender-pay-gap
Ingersoll, J. (2020). *What is Bias?* Pymetrics Pygest Blog. https://www.pymetrics.ai/pygest/what-is-bias
Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2020). Big Data analytics and Computational Intelligence for Cyber–Physical Systems: Recent trends and state of the art applications. *Future Generation Computer Systems*, *105*, 766–778. https://doi.org/10.1016/j.future.2017.10.021
Iqbal, S., & Kar, B. (2021). How workplace biases undermine women's struggle for equity.

Nature India. https://www.nature.com/articles/nindia.2021.39
Isaac, C., Lee, B., & Carnes, M. (2009). Interventions that affect gender bias in hiring: a systematic review. *Academic Medicine*, *84*(10), 1440–1446. https://doi.org/10.1097/ACM.0b013e3181b6ba00
Jacobson, D., & Mustafa, N. (2019). Social Identity Map: A Reflexivity Tool for Practicing Explicit Positionality in Critical Qualitative Research. *International Journal of Qualitative Methods*, *18*, 1–12. https://doi.org/10.1177/1609406919870075
James, N., & Busher, H. (2009). *Online interviewing*. SAGE.
Jameson, F. (1976). On Goffman's Frame Analysis. *Theory and Society*, *3*(1), 119–133. https://doi.org/10.1007/BF00158482
Jenkins, D. (2021). Understanding and fighting structural injustice. *Journal of Social Philosophy*, *52*(4), 569–586. https://doi.org/10.1111/josp.12386
Jia, Q., Guo, Y., Li, R., Li, Y., & Chen, Y. (2018). A conceptual artificial intelligence application framework in human resource management. *Proceedings of the International Conference on Electronic Business (ICEB)*, *December*, 106–114.
Jiang, S., Schenke, K., Eccles, J. S., Xu, D., & Warschauer, M. (2018). Cross-national comparison of gender differences in the enrollment in and completion of science, technology, engineering, and mathematics Massive Open Online Courses. *PLoS ONE*, *13*(9), 1–16. https://doi.org/10.7910/DVN/26147
Jillson, C. C. (2016). *The American dream: in history, politics, and fiction*. University Press of Kansas.
Johnston, H. (1995). A Methodology for Frame Analysis: From Discourse to Cognitive Schemata. In H. Johnston & B. Klandermans (Eds.), *Social Movements and Culture* (pp. 217–246). University of Minnesota Press.
Jones, C. C. (2023). Systematizing discrimination: AI vendors and Title VII enforcement. *University of Pennsylvania Law Review*, *171*(1), 235–266.
Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., Noble, S. U., & Shestakofsky, B. (2021). Toward a Sociology of Artificial Intelligence: A Call for Research on Inequalities and Structural Change. *Socius*, *7*, 1–11. https://doi.org/10.1177/2378023121999581
Kabir, N. (2013). Tacit Knowledge, its Codification and Technological Advancement. *Electronic Journal of Knowledge Management : EJKM*, *11*(3), 235–243.
Kahneman, D. (2012). *Thinking, fast and slow*. Penguin.
Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise : a flaw in human judgement*. William Collins.
Kak, A., & West, S. M. (2023). *2023 Landscape: Confronting tech power*. https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf
Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened Résumés: Race and Self-Presentation in the Labor Market. *Administrative Science Quarterly*, *61*(3), 469–502. https://doi.org/10.1177/0001839216639577
Kantar, R., McNulty, K., Snow, E. L., Emery, M. A., Wainess, R., & Doshi, S. D. (2018). Constructing cognitive profiles for simulation-based hiring assessments. *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018*, 1–5. https://educationaldatamining.org/files/conferences/EDM2018/papers/EDM2018_paper_19.pdf
Kelan, E. K. (2023). Algorithmic inclusion: Shaping the predictive algorithms of artificial intelligence in hiring. *Human Resource Management Journal*, *April*, 1–14. https://doi.org/10.1111/1748-8583.12511
Kelly-Lyth, A. (2021). Challenging Biased Hiring Algorithms. *Oxford Journal of Legal Studies*, *41*(4), 899–928. https://doi.org/10.1093/ojls/gqab006
Kelman, H. C. (1972). The rights of the subject in social research: An analysis in terms of relative power and legitimacy. *American Psychologist*, *27*(11), 989–1016. https://doi.org/10.1037/h0033995
Keyes, O. (2018). The Misgendering Machines: Trans/HCI Implications of Automatic Gender

Recognition. *Proceedings of the ACM on Human-Computer Interaction*, *2*, 1–22. https://doi.org/10.1145/3274357

Keyes, O., May, C., & Carrell, A. (2021). You Keep Using That Word: Ways of Thinking about Gender in Computing Research. *Proceedings of the ACM on Human-Computer Interaction*, *5*, 1–23. https://doi.org/10.1145/3449113

Kim, P. (2017a). Auditing Algorithms for Discrimination. *University of Pennsylvania Law Review*, *166*(1), 189–203. https://scholarship.law.upenn.edu/penn_law_review_online/vol166/iss1/10/

Kim, P. (2017b). DATA-DRIVEN DISCRIMINATION AT WORK. *William and Mary Law Review*, *58*(3), 857–936. https://scholarship.law.wm.edu/wmlr/vol58/iss3/4/

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic Fairness. *AEA Papers and Proceedings*, *108*(May), 22–27. https://doi.org/10.1257/pandp.20181018

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2020). Algorithms as discrimination detectors. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(48), 30096–30100. https://doi.org/10.1073/pnas.1912790117

Knight, W. (2021). *Job Screening Service Halts Facial Analysis of Applicants*. Wired. https://www.wired.com/story/job-screening-service-halts-facial-analysis-applicants/

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*(1), 128–161. https://doi.org/10.1037/a0036734

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

Kramer, A. (2022). Bias Holds Women Back. *Forbes*. https://www.forbes.com/sites/andiekramer/2022/02/24/bias-holds-women-back/

Kravitz, D. A. (2008). The diversity-validity dilemma: Beyond selection - The role of affirmative action. *Personnel Psychology*, *61*(1), 173–193. https://doi.org/10.1111/j.1744-6570.2008.00110.x

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*, *165*(3), 633–705.

Kvale, S. (1996). *InterViews: learning the craft of qualitative research interviewing*. SAGE.

Lahey, D. (2015). *Predicting success : evidence-based strategies to hire the right people and build the best team*. John Wiley & Sons.

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science*, *65*(7), 2966–2981. https://doi.org/10.1287/mnsc.2018.3093

Lapadat, J. C. (2010). Thematic Analysis. In *Encyclopedia of Case Study Research* (pp. 925–927).

Lareau, A. (2002). Invisible inequality: Social class and childrearing in black families and white families. *American Sociological Review*, *67*(5), 747–776. https://doi.org/10.2307/3088916

Lazarsfeld, P., & Merton, R. (1954). Friendship as a social process: a substantive and methodological analysis. In M. Berger & R. M. MacIver (Eds.), *Freedom and Control in Modern Society* (pp. 18–66). Van Nostrand.

Leavy, S. (2018). Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning. *GE '18: Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 14–16. https://doi.org/10.1145/3195570.3195580

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society*, *5*(11), 1–16. https://doi.org/10.1177/2053951718756684

Legewie, J., & DiPrete, T. A. (2012). School Context and the Gender Gap in Educational Achievement. *American Sociological Review*, *77*(3), 463–485. https://doi.org/10.1177/0003122412440802

Leonhardt, M., & Overå, S. (2021). Are there differences in video gaming and use of social

media among boys and girls?—a mixed methods approach. *International Journal of Environmental Research and Public Health*, *18*(11), 1–13. https://doi.org/10.3390/ijerph18116085

Li, A., Bagger, J., & Cropanzano, R. (2017). The impact of stereotypes and supervisor perceptions of employee work–family conflict on job performance ratings. *Human Relations*, *70*(1), 119–145. https://doi.org/10.1177/0018726716645660

Li, L. (2021). How to tackle variations in elite interviews: Access, strategies, and power dynamics. *Qualitative Research*, *22*(6), 1–16. https://doi.org/10.1177/1468794121994475

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Linström, M., & Marais, W. (2012). Qualitative News Frame Analysis : a Methodology What Are News Frames ? *Communitas*, *17*, 21–37. https://doi.org/https://doi.org/10.38140/com.v17i0.991

Liu, Z. (2021). Sociological perspectives on artificial intelligence: A typological reading. *Sociology Compass*, *15*(3), 1–13. https://doi.org/10.1111/soc4.12851

Lopez, P. (2021). Bias does not equal bias: A socio-technical typology of bias in data-based algorithmic systems. *Internet Policy Review*, *10*(4), 1–29. https://doi.org/10.14763/2021.4.1598

Lorino, P., Mourey, D., & Schmidt, G. (2017). Goffman's theory of frames and situated meaning-making in performance reviews. The case of a category management approach in the French retail sector. *Accounting, Organizations and Society*, *58*, 32–49. https://doi.org/10.1016/j.aos.2017.03.004

Low, J. (2019). A Pragmatic Definition of the Concept of Theoretical Saturation. *Sociological Focus*, *52*(2), 131–139. https://doi.org/10.1080/00380237.2018.1544514

Luksyte, A., Unsworth, K. L., & Avery, D. R. (2018). Innovative work behavior and sex-based stereotypes: Examining sex differences in perceptions and evaluations of innovative work behavior. *Journal of Organizational Behavior*, *39*(3), 292–305. https://doi.org/10.1002/job.2219

Luo, J., Nah, S., & Joo, J. (2023). AI bias, news framing, and mixed-methods approach. In S. Nah (Ed.), *Research Handbook on Artificial Intelligence and Communication* (pp. 145–161). Edward Elgar Publishing. https://doi.org/https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.4337/9781803920306

Luzadis, R., Wesolowski, M., & Snavely, B. K. (2008). Understanding Criterion Choice in Hiring Decisions from a Prescriptive Gender Bias Perspective. *Journal of Managerial Issues*, *20*(4), 468–484.

Ma, S., Seidl, D., & Mcnulty, T. (2021). Challenges and practices of interviewing business elites. *Strategic Organization*, *19*(1), 81–96. https://doi.org/10.1177/1476127020980969

Macabasco, L. W. (2021). "They become dangerous tools": the dark side of personality tests. *Guardian*. https://www.theguardian.com/tv-and-radio/2021/mar/03/they-become-dangerous-tools-the-dark-side-of-personality-tests

MacKenzie, D. A., & Wajcman, J. (1999). *The social shaping of technology* (2nd ed.). Open University Press.

Madera, J. M., & Abbott, J. A. (2012). The diversity-validity dilemma: Implications from the ricci v. destefano u.s. supreme court case. *Cornell Hospitality Quarterly*, *53*(1), 31–39. https://doi.org/10.1177/1938965511424150

Madiega, T. (2023). *Artificial intelligence liability directive* (Issue February). https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf

Martell, R. F., Parker, C., Emrich, C. G., & Crawford, M. S. (1998). Sex stereotyping in the executive suite: "Much ado about something." *Journal of Social Behavior and Personality*, *13*(1), 127–138.

Martinez, J. J. (2022). *Not a Good Fit for the Job: An Intersectional Study of Occupational Stereotypes in the Field of Technology*. Northern Illinois University ProQuest Dissertations Publishing.

Masiero, S., & Aaltonen, A. (2020). Gender Bias in Information Systems Research: A Literature Review Review. *AISWN International Research Workshop on Women, IS and Grand Challenges*, 1–10. https://aisel.aisnet.org/aiswn2020

Mason-Bish, H. (2019). The elite delusion: reflexivity, identity and positionality in qualitative research. *Qualitative Research*, *19*(3), 263–276. https://doi.org/10.1177/1468794118770078

Maurer, R. (2021). HireVue Discontinues Facial Analysis Screening. *SHRM*. https://www.shrm.org/topics-tools/news/talent-acquisition/hirevue-discontinues-facial-analysis-screening

McClure, K., & McNaughtan, J. (2021). Proximity to Power: The Challenges and Strategies of Interviewing Elites in Higher Education Research. *The Qualitative Report*, *26*(3), 874–992. https://doi.org/10.46743/2160-3715/2021.4615

Mcpherson, M., Smith-lovin, L., & Cook, J. M. (2001). Birds of a Feather : Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444.

Mears, A. (2020). *Very Important People : Status and Beauty in the Global Party Circuit*. Princeton University Press.

Mellamphy, N. B. (2021). Humans "in the Loop"? *Nature and Culture*, *16*(1), 11–27. https://doi.org/10.3167/nc.2020.160102

Mendonça, R. F., & Simões, P. G. (2022). Frame analysis. In S. A. Ercan, H. Asenbaum, N. Curato, & R. F. Mendonça (Eds.), *Research Methods in Deliberative Democracy* (pp. 345–355). Oxford University Press. https://doi.org/10.1093/oso/9780192848925.003.0024

Merry, S. E. (2016). *The seductions of quantification : measuring human rights, gender violence, and sex trafficking*. The University of Chicago Press.

Merton, R. K. (1957). *Social theory and social structure*. Free Press.

Michailidis, M. (2018). The Challenges of AI and Blockchain on HR Recruiting Practices [Article]. *The Cyprus Review*, *30*(2), 169–20. https://www.proquest.com/scholarly-journals/challenges-ai-blockchain-on-hr-recruiting/docview/2193108119/se-2?accountid=13042%0A

Mikecz, R. (2012). Interviewing Elites. *Qualitative Inquiry*, *18*(6), 482–493. https://doi.org/10.1177/1077800412442818

Miller, G. J. (2022). Stakeholder-accountability model for artificial intelligence projects. *Journal of Economics and Management*, *44*(1), 446–494. https://doi.org/10.22367/jem.2022.44.18

Millington, K. . (2017). *How changes in technology and automation will affect the labour market in Africa*. https://assets.publishing.service.gov.uk/media/5b9a4b3540f0b6786f821738/K4D_HDR_Impact_of_automation_on_jobs_in_Africa.pdf

Mökander, J., Juneja, P., Watson, D. S., & Floridi, L. (2022). The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other? *Minds and Machines*, *32*(4), 751–758. https://doi.org/10.1007/s11023-022-09612-y

Mondragon, N. (2023). Artificial intelligence in automated scoring of video interviews. In T. Kantrowitz, D. H. Reynolds, & J. Scott (Eds.), *Talent Assessment: Embracing Innovation and Mitigating Risk in the Digital Age* (pp. 111–127). Oxford University Press. https://doi.org/10.1093/oso/9780197611050.003.0008

Morozov, E. (2014). *To save everything, click here: the folly solutionism*. Penguin.

Morris, Z. S. (2009). The truth about interviewing elites. *Politics*, *29*(3), 209–217. https://doi.org/10.1111/j.1467-9256.2009.01357.x

Moss, H. (2021). SCREENED OUT ONSCREEN: DISABILITY DISCRIMINATION, HIRING BIAS, AND ARTIFICIAL INTELLIGENCE. *Denver Law Review*, *98*(4), 775–805.

Mountford Zimdars, A. (2017). *Meritocracy and the University*. Bloomsbury Publishing Plc.

Mumford, E. (2006). The story of socio-technical design: reflections on its successes, failures and potential. *Information Systems Journal*, *16*(4), 317–342. https://doi.org/10.1111/j.1365-2575.2006.00221.x

Murray, H., & Trist, B. (2016). Introduction to Volume II. In H. Murray, B. Trist, & E. L. Trist (Eds.), *The Social Engagement of Social Science, a Tavistock Anthology,Volume 2: The Socio-Technical Perspective* (pp. 36–64). University of Pennsylvania Press. https://doi.org/10.9783/9781512819052

Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Gender bias in AI-based decision-making systems: a systematic literature review. *Australasian Journal of Information Systems*, *26*, 1–34. https://doi.org/10.3127/AJIS.V26I0.3835

Nadler, J. (2017). Gender Bias in Hiring Practices. In *The SAGE Encyclopedia of Psychology and Gender* (pp. 619–621).

Nandigama, D., & Shyamsunder, A. (2021). Eeny, Meeny, Miny, Moe: Hire Him and Let Her Go? Using Science to Reduce Hiring Bias. *NHRD Network Journal*, *14*(2), 259–273. https://doi.org/10.1177/2631454120987343

Nascimento, L. de C. N., Souza, T. V. de, Oliveira, I. C. D. S., Moraes, J. R. M. M. de, Aguiar, R. C. B. de, & Silva, L. F. da. (2018). Theoretical saturation in qualitative research: an experience report in interview with schoolchildren. *Revista Brasileira de Enfermagem*, *71*(1), 228–233. https://doi.org/10.1590/0034-7167-2016-0616

Niehueser, W., & Boak, G. (2020). Introducing artificial intelligence into a human resources function. *Industrial and Commercial Training*, *52*(2), 121–130. https://doi.org/10.1108/ICT-10-2019-0097

Njoto, S., Cheong, M., Lederman, R., McLoughney, A., Ruppanner, L., & Wirth, A. (2022). Gender Bias in AI Recruitment Systems: A Sociological-and Data Science-based Case Study. *IEEE International Symposium on Technology and Society (ISTAS)*, 1–7. https://doi.org/10.1109/ISTAS55053.2022.10227106

Noble, S. U. (2018). *Algorithms of oppression : how search engines reinforce racism*. New York University Press.

Nocker, M., & Sena, V. (2019). Big data and human resources management: The rise of talent analytics. *Big Data and Employee Wellbeing*, *8*(10), 1–19. https://doi.org/doi:10.3390/socsci8100273

O'Connor, H., Madge, C., Shaw, R., & Wellens, J. (2008). Internet-based interviewing. In Nigel Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE Handbook of Online Research Methods* (pp. 271–289). SAGE Publications, Ltd. https://doi.org/10.4135/9780857020055

O'Connor, Sarah. (2019). The robot-proof skills that give women an edge in the age of AI. *Financial Times*. https://www.ft.com/content/06afd24a-2dfb-11e9-ba00-0251022932c8

O'Connor, Sinead, & Liu, H. (2023). Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. *AI and Society*. https://doi.org/10.1007/s00146-023-01675-4

O'Neil, C. (2016). How Algorithms Rule our Working Lives. *Guardian*. https://www.theguardian.com/science/2016/sep/01/how-algorithms-rule-our-working-lives

OECD. (2019). *SCOPING THE OECD AI PRINCIPLES OECD DIGITAL ECONOMY*. https://doi.org/https://doi.org/10.1787/d62f618a-en

OECD. (2024). *Understanding social mobility*. OECD.Org. https://www.oecd.org/stories/social-mobility/

Ore, O., & Sposato, M. (2022). Opportunities and risks of artificial intelligence in recruitment and selection. *International Journal of Organizational Analysis*, *30*(6), 1771–1782. https://doi.org/10.1108/IJOA-07-2020-2291

Ortiz-Ospina, E., & Roser, M. (2019). *Economic inequality by gender*. Our World in Data. https://ourworldindata.org/economic-inequality-by-gender

Ostrander, S. (1993). SURELY YOU'RE NOT IN THIS JUST TO BE HELPFUL. *Journal of Contemporary Ethnography*, *22*(1), 7–27. https://doi.org/10.1177/089124193022001002

Ovadya, A. (2023). Reimagining Democracy for Ai. *Journal of Democracy*, *34*(4), 162–170. https://doi.org/10.1353/jod.2023.a907697

Owens, L. L. (2022). An Implausible Virtual Interview: Conversations with a Professional Research Subject. *Sociological Methodology*, *52*(2), 121–140. https://doi.org/10.1177/00811750221106777

Pan, Y., Froese, F., Liu, N., Hu, Y., & Ye, M. (2021). The adoption of artificial intelligence in

employee recruitment: The influence of contextual factors. *International Journal of Human Resource Management*, *33*(6), 1–23. https://doi.org/10.1080/09585192.2021.1879206

Papageorgiou, T., & Michaelides, P. G. (2016). Joseph Schumpeter and Thorstein Veblen on technological determinism, individualism and institutions. *European Journal of the History of Economic Thought*, *23*(1), 1–30. https://doi.org/10.1080/09672567.2013.792378

Pawluczuk, A., Lee, J. H., & Gamundani, A. M. (2021). Bridging the gender digital divide: an analysis of existing guidance for gender digital inclusion programmes' evaluations. *Digital Policy, Regulation and Governance* , *23*(3), 287–299. https://doi.org/10.1108/DPRG-11-2020-0158

Peng, H., Teplitskiy, M., Romero, D. M., & Horvát, E.-Á. (2022). The Gender Gap in Scholarly Self-Promotion on Social Media. *Psychological Science*, *32*(12), 1952–1964. https://doi.org/10.1177/09567976211016395

Perera, K. (2021). Interviewing academic elites: a discourse analysis of shifting power relations. *Qualitative Research*, *21*(2), 215–233. https://doi.org/10.1177/1468794120924208

Persson, A. (2022). Frame Analysis. In *The Routledge International Handbook of Goffman Studies* (pp. 119–130). Routledge. https://doi.org/10.4324/9781003160861-12

Peters, J. D. (2017). "You mean my whole fallacy is wrong": On technological determinism. *Representations*, *140*(1), 10–26. https://doi.org/10.1525/rep.2017.140.1.10

Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, *61*(1), 153–172. https://doi.org/10.1111/j.1744-6570.2008.00109.x

Pollack Ichou, R. (2018). Can MOOCs reduce global inequality in education? *Australasian Marketing Journal*, *26*(2), 116–120. https://doi.org/10.1016/j.ausmj.2018.05.007

Prates, M. O. R., Avelar, P. H. C., & Lamb, L. (2019). Assessing Gender Bias in Machine Translation: A Case Study with Google Translate. *Neural Computing & Applications*, *32*(10), 6363–6381. https://doi.org/10.1007/s00521-019-04144-6

Putnam, R. D. (2015). *Our kids: the American Dream in crisis*. Simon & Schuster.

Pyburn, K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, *61*(1), 143–151. https://doi.org/10.1111/j.1744-6570.2008.00108.x

Pymetrics. (n.d.-a). *Designing Artificial Intelligence With Ethical Principles at the Forefront*. https://www.pymetrics.ai/resource/email-7-principles

Pymetrics. (n.d.-b). *Removing Bias from Talent Decisions with Artificial Intelligence*. https://www.pymetrics.ai/resource/website-removing-bias-from-talent-decisions-with-ai

Pymetrics. (2020). *The Future of Work is Human*. https://www.pymetrics.ai/resource/hr-tech-reimagining-talent-management-with-soft-skills-data

Pymetrics. (2021). *Soft Skills: Your Competitive Workforce Advantage*. https://www.pymetrics.ai/resource/soft-skills-your-competitive-workforce-advantage-ws

Pymetrics. (2022a). *Assessments*. Pymetrics Website. https://www.pymetrics.ai/assessments

Pymetrics. (2022b). *Candidates*. Pymetrics Website. https://www.pymetrics.ai/candidates

Pymetrics. (2022c). *Diversity Workforce*. Pymetrics Webstie. https://www.pymetrics.ai/diversify-workforce

Pymetrics. (2022d). *High-Volume/Hourly Worker*. Pymetrics Website. https://www.pymetrics.ai/high-volume-hourly-worker

Pymetrics. (2022e). *Homepage*. Pymetrics Website. https://www.pymetrics.ai

Pymetrics. (2022f). *Mission*. Pymetrics Website. https://www.pymetrics.ai/mission

Pymetrics. (2022g). *Science*. Pymetrics Website. https://www.pymetrics.ai/science

Pymetrics. (2022h). *Talent Acquisition*. Pymetrics Website. https://www.pymetrics.ai/talent-acquisition

Rab-Kettler, K., & Lehnervp, B. (2019). Recruitment in the Times of Machine Learning. *Management Systems in Production Engineering*, *27*(2), 105–109. https://doi.org/10.1515/mspe-2019-0018

Rafalow, M. H. (2021). *Digital divisions : how schools create inequality in the tech era*. The

University of Chicago Press. https://doi.org/10.7208/9780226726724
Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 469–481. https://doi.org/10.1145/3351095.3372828
Raub, M. (2018). Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices. *Arkansas Law Review*, *71*(2), 529–570. https://scholarworks.uark.edu/alr/vol71/iss2/7?utm_source=scholarworks.uark.edu%2Falr%2Fvol71%2Fiss2%2F7&utm_medium=PDF&utm_campaign=PDFCoverPages
Reardon, S., Kalagrides, D., & Shores, K. (2018). The Geography of Racial / Ethnic Test Score Gaps. *The American Journal of Sociology*, *124*(4), 1164–1221. https://doi.org/10.1086/700678
Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., & Huguet, P. (2019). Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature Human Behaviour*, *3*(11), 1171–1179. https://doi.org/10.1038/s41562-019-0686-3
Rehbein, F., Staudt, A., Hanslmaier, M., & Kliem, S. (2016). Video game playing in the general adult population of Germany: Can higher gaming time of males be explained by gender specific genre preferences? *Computers in Human Behavior*, *55*, 729–735. https://doi.org/10.1016/j.chb.2015.10.016
Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). The Problem of Bias in Psychological Assessment. In C. R. Reynolds, R. A. Altmann, & D. N. Allen (Eds.), *Mastering Modern Psychological Testing* (pp. 573–613). Springer International Publishing AG. https://doi.org/10.1007/978-3-030-59455-8_15
Reynolds, K. (2021). *COVID-19 increased use of AI. Here's why it's here to stay*. WEF. https://www.weforum.org/agenda/2021/02/covid-19-increased-use-of-ai-here-s-why-its-here-to-stay/
Rhea, A. K., Markey, K., D'Arinzo, L., Schellmann, H., Sloane, M., Squires, P., Arif Khan, F., & Stoyanovich, J. (2022). An external stability audit framework to test the validity of personality prediction in AI hiring. *Data Mining and Knowledge Discovery*, *36*(6), 2153–2193. https://doi.org/10.1007/s10618-022-00861-0
Rice, G. (2010). Reflections on interviewing elites. *Area (London 1969)*, *42*(1), 70–75. https://doi.org/10.1111/j.1475-4762.2009.00898.x
Richards, L. (1999). Data alive! The thinking behind NVivo. *Qualitative Health Research*, *9*(3), 412–428. https://doi.org/10.1177/104973299129121857
Rincón, C., Keyes, O., & Cath, C. (2021). Speaking from Experience: Trans/Non-Binary Requirements for Voice-Activated AI. *Proceedings of the ACM on Human-Computer Interaction*, 1–27. https://doi.org/10.1145/3449206
Ritter, C. S. (2022). Rethinking digital ethnography: A qualitative approach to understanding interfaces. *Qualitative Research*, *22*(6), 916–932. https://doi.org/10.1177/14687941211000540
Rivera, L. A. (2016). *Pedigree: How Elite Students Get Elite Jobs* (L. A. Rivera (ed.)). Princeton University Press. https://doi.org/10.1515/9781400880744
Robison, K. (2024). ChatGPT will be able to talk to you like Scarlett Johansson in Her. *The Verge*. https://www.theverge.com/2024/5/13/24155652/chatgpt-voice-mode-gpt4o-upgrades
Rottman, C., Gardner, C., Liff, J., Mondragon, N., & Zuloaga, L. (2023). New Strategies for Addressing the Diversity–Validity Dilemma With Big Data. *Journal of Applied Psychology*, *108*(9), 1425–1444. https://doi.org/10.1037/apl0001084
Rovatsos, M., Mittelstadt, B., & Koene, A. (2019). Landscape Summary: Bias in Algorithmic Decision-Making. *Centre for Data Ethics and Innovation*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf
Rowlands, T., Waddell, N., & McKenna, B. (2015). Are we there yet? A technique to determine theoretical saturation. *Journal of Computer Information Systems*, *56*(1), 40–47. https://doi.org/10.1080/08874417.2015.11645799
Roy, K., Zvonkovic, A., Goldberg, A., Sharp, E., & Larossa, R. (2015). Sampling Richness and

Qualitative Integrity: Challenges for Research With Families. *Journal of Marriage and Family*, *77*(1), 243–260. https://doi.org/10.1111/jomf.12147

Ryu, K. (2019). How AI Can Improve Hiring Process For Hourly Workers. *Talent Acquisition Excellence Essentials*. https://www.proquest.com/docview/2211321309?accountid=13042&sourcetype=Magazines

Sanchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to solve the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. *Conference on Fairness, Accountability, and Transparency (FAT* '20), January 27–30, 2020, Barcelona, Spain.*, 1–11. https://doi.org/10.48550/arxiv.1910.06144

Sandberg, J., & Alvesson, M. (2011). Ways of constructing research questions: gap-spotting or problematization? *Organization*, *18*(1), 23–44. https://doi.org/10.1177/1350508410372151

Sarker, M. R. (2021). Labor market and unpaid works implications of COVID-19 for Bangladeshi women. *Gender, Work and Organization*, *28*(S2), 597–604. https://doi.org/10.1111/gwao.12587

Sartori, L., & Collett, C. (n.d.). *The third eye, the river, and the car bonnet: Problematising sociotechnical imaginaries related to social inequalities in the design and use of AI recruitment technology*.

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics and Information Technology*, *24*(1), 1–11. https://doi.org/10.1007/s10676-022-09624-3

Scheuerman, M., Jiang, A., Spiel, K., & Brubaker, J. R. (2021). Revisiting Gendered Web Forms: An Evaluation of Gender Inputs with (Non-)Binary People. *CHI '21: CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/https://doi.org/10.1145/3411764.3445742

Scheuerman, M., Paul, J., & Brubaker, J. (2019). How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services [Article]. *Proceedings of the ACM on Human-Computer Interaction*, *3 (CSCW)*, 1–33. https://doi.org/10.1145/3359246

Schick, J., & Fischer, S. (2021). Dear Computer on My Desk, Which Candidate Fits Best? An Assessment of Candidates' Perception of Assessment Quality When Using AI in Personnel Selection. *Frontiers in Psychology*, *12*(October), 1–11. https://doi.org/10.3389/fpsyg.2021.739711

Schmader, T. (2023). Gender Inclusion and Fit in STEM. *Annual Review of Psychology*, *74*(1), 219–243. https://doi.org/10.1146/annurev-psych-032720-043052

Schön, D. A., & Rein, M. (1994). *Frame reflection: toward the resolution of intractable policy controversies*. Basic Books.

Seger, E., Ovadya, A., Siddarth, D., Garfinkel, B., & Dafoe, A. (2023). Democratising AI: Multiple Meanings, Goals, and Methods. *AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 715–722. https://doi.org/10.1145/3600211.3604693

Seiler, D. (2021). *COVID-19 digital transformation & technology*. McKinsey. https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, 59–68. https://doi.org/10.1145/3287560.3287598

Shapiro, H. B., Lee, C. H., Wyman Roth, N. E., Li, K., Çetinkaya-Rundel, M., & Canelas, D. A. (2017). Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. *Computers and Education*, *110*, 35–50. https://doi.org/10.1016/j.compedu.2017.03.003

Shestakofsky, B. (2024). *Behind the Startup : How Venture Capital Shapes Work, Innovation, and Inequality*. University of California Press. https://doi.org/https://doi-org.ezproxy-

prd.bodleian.ox.ac.uk/10.1525/9780520395046
Shoemaker, P. J., Tankard, J. W., & Lasorsa, D. L. (2004). *How to build social science theories*. SAGE.
Simonite, T. (2021). New York City Proposes Regulating Algorithms Used in Hiring. *Wired*. https://www.wired.com/story/new-york-city-proposes-regulating-algorithms-hiring/
Simons, J. (2023). *Algorithms for the people : democracy in the age of AI*. Princeton University Press. https://doi.org/10.1515/9780691244914
Singh, V. (2002). *Managing Diversity for Strategic Advantage*. The Council.
Sloane, M. (2019a). On the need for mapping design inequalities. *Design Issues*, *35*(4), 3–11. https://doi.org/10.1162/desi_a_00559
Sloane, M. (2023). Automation and Recruiting: Understanding the Intersection of Algorithmic Systems and Professional Discretion in the Sourcing of Job Candidates. *Proceedings "Automation by Design" Conference, University of Minnesota, Forthcoming*, 1–19. https://doi.org/10.2139/ssrn.4516453
Sloane, M. (2019b). Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice. *Proceedings of the Weizenbaum Conference 2019 "Challenges of Digital Inequality - Digital Education, Digital Work, Digital Life,"* 1–9. https://doi.org/https://doi.org/10.34669/wi.cp/2.9
Sloane, M., & Moss, E. (2019). AI's social sciences deficit. *Nature Machine Intelligence*, *1*(8), 330–331. https://doi.org/10.1038/s42256-019-0084-6
Sloane, M., Moss, E., & Chowdhury, R. (2022). A Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability. *Patterns (New York, N.Y.)*, *3*(2), 100425–100425. https://doi.org/10.1016/j.patter.2021.100425
Small, M. L., & Calarco, J. M. (2022). *Qualitative literacy : a guide to evaluating ethnographic and interview research*. University of California Press.
Sonderling, K. E., Kelley, B. J., & Casimir, L. (2022). The Promise and The Peril : Artificial Intelligence and Employment Discrimination The Promise and The Peril : Artificial Intelligence and Employment Discrimination. *University of Miami Law Review*, *77*(1), 1–87.
Spanjaart, J. (2022). *Will this class action lawsuit mark the end of facial recognition software in recruitment?* Totalent. https://totalent.eu/will-this-class-action-lawsuit-mark-the-end-of-facial-recognition-software-in-recruitment/
Steidl, C. R., & Werum, R. (2019). If all you have is a hammer, everything looks like a nail: Operationalization matters [Article]. *Sociology Compass*, *13*(8), 1–14. https://doi.org/10.1111/soc4.12727
Stone, D. L., & Dulebohn, J. H. (2013). Emerging issues in theory and research on electronic human resource management (eHRM). *Human Resource Management Review*, *23*(1), 1–5. https://doi.org/10.1016/j.hrmr.2012.06.001
Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.
Sullivan, K. (2023). Three levels of framing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *14*(5), 1–13. https://doi.org/10.1002/wcs.1651
Swedberg, R. (2014). *The art of social theory*. Princeton University Press.
Talmage, J. B. (2012). Listening to, and for, the Research Interview. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds.), *The SAGE Handbook of Interview Research: The Complexity of the Craft* (2nd ed., pp. 295–304). SAGE Publications, Inc. https://doi.org/10.4135/9781452218403
Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*, *61*(4), 15–42. https://doi.org/10.1177/0008125619867910
Taussig, D. (2021). *What We Mean by the American Dream: Stories We Tell about Meritocracy*. Cornell University Press. https://doi.org/10.1515/9781501754708
Taylor, H. (2007). Tacit Knowledge: Conceptualizations and Operationalizations. *International Journal of Knowledge Management*, *3*(3), 60–73. https://doi.org/10.4018/jkm.2007070104
Thorne, B. (1993). *Gender play: girls and boys in school*. Rutgers University Press.
Ticona, J. (2021). Smartphones Are a New Tax on the Poor. *Wired*.

https://www.wired.com/story/phones-connectivity-tax-policy/
Ticona, J. (2022). *Left to Our Own Devices: Coping with Insecure Work in a Digital Age*. Oxford University Press.
Tilmes, N. (2022). Disability, fairness, and algorithmic bias in AI recruitment. *Ethics and Information Technology*, *24*(21), 1–13. https://doi.org/10.1007/s10676-022-09633-2
Toerien, M., & Wilkinson, S. (2003). Gender and body hair: constructing the feminine woman. *Women's Studies International Forum*, *26*(4), 333–344. https://doi.org/10.1016/S0277-5395(03)00078-5
Tradenta, J. M., Neelim, A., & Vecci, J. (2019). Gender differences in self-promotion: Understanding the female modesty constraint. *Strategy and Leadership*, *16*(3), 1–12. https://doi.org/https://doi.org/10.1016/j.bushor.2018.08.010%0Ahttps://ssrn.com/abstract=3039233%0A
Trammell, A. (2023). *The Privilege of Play : A History of Hobby Games, Race, and Geek Culture.* (1st ed.). New York University Press. https://doi.org/10.18574/nyu/9781479818419.001.0001
Trindel, K., Kassir, S., & Bent, J. (2021). Fairness in Algorithmic Employment Selection: How to Comply with Title VII. *ABA Journal of Labor & Employment Law*, *35*(2), 241–287.
TUC. (2018). *I'll Be Watching You: A report on workplace monitoring*. https://www.tuc.org.uk/research-analysis/reports/ill-be-watching-you
Tuñón, I., & Martínez, C. E. (2021). Inequalities in Early Childhood Education and Care in Argentina and the United Kingdom (2010-2020). *Foro de Educacion*, *19*(2), 141–162. https://doi.org/10.14516/FDE.882
Ulnicane, I., & Aden, A. (2023). Power and politics in framing bias in Artificial Intelligence policy. *Review of Policy Research*, *40*(5), 665–687. https://doi.org/10.1111/ropr.12567
UN Women. (2021). *Facts and figures: Women's leadership and political participation*. https://www.unwomen.org/en/what-we-do/leadership-and-political-participation/facts-and-figures
UNESCO. (2019). *I'd blush if I could: closing gender divides in digital skills through education*. https://unesdoc.unesco.org/ark:/48223/pf0000367416
UNESCO, IDB, & OECD. (2022). *The Effects of AI on the Working Lives of Women*. http://www.unesco.org/open-access/terms-use-ccbysa-en
UNRISD. (2010). *Combating poverty and inequality: Policy brief of the Executive Committee of Economic and Social Affairs*. https://www.un.org/esa/ecesa/documents/documents/policy_briefs/ecesa_pb_1.pdf
Upadhyay, A., & Khandelwal, K. (2018). Applying artificial intelligence: implications for recruitment [Article]. *Strategic HR Review*, *17*(5), 255–258. https://doi.org/10.1108/SHR-07-2018-0051
Urban, A. (2022). *ADDRESSING THE GENDER GAP IN STEM MOOCS: HOW BRIEF, IN-COURSE MESSAGES CAN INCREASE FEMALES' MOTIVATION AND ONLINE LEARNING SUCCESS* [John Hopkins University]. https://jscholarship.library.jhu.edu/server/api/core/bitstreams/dac41ce5-d85a-42cc-9f1f-c42719644966/content
USDHHS, (US Department of Health and HumanServices). (2024). *Introduction to U.S. Law & Policy*. https://www.phe.gov/s3/law/Pages/default.aspx
van den Broek, E., Sergeeva, A., & Huysman, M. (2020). Hiring algorithms: An ethnography of fairness in practice. *40th International Conference on Information Systems, ICIS 2019*, 1–9. https://aisel.aisnet.org/icis2019/future_of_work/future_work/6/
van Deursen, A., & van Dijk, J. (2011). Internet skills and the digital divide. *New Media and Society*, *13*(6), 893–911. https://doi.org/10.1177/1461444810386774
van Esch, P., & Black, J. S. (2019). Factors that influence new generation candidates to engage with and complete digital, AI-enabled recruiting. *Business Horizons*, *62*(6), 729–739. https://doi.org/10.1016/j.bushor.2019.07.004
van Esch, P., Black, J. S., & Ferolie, J. (2019). Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior*, *90*, 215–222. https://doi.org/10.1016/j.chb.2018.09.009

Van Gorp, B. (2007). The Constructionist Approach to Framing: Bringing Culture Back In. *Journal of Communication*, *57*(1), 60–78. https://doi.org/10.1111/j.1460-2466.2006.00329.x

Vassilopoulou, J., Kyriakidou, O., Özbilgin, M. F., & Groutsis, D. (2022). Scientism as illusio in HR algorithms: Towards a framework for algorithmic hygiene for bias proofing. *Human Resource Management Journal*, *34*(2), 1–15. https://doi.org/10.1111/1748-8583.12430

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Vecchione, B., Levy, K., & Barocas, S. (2021). Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies. *ACM International Conference Proceeding Series*, 1–14. https://doi.org/10.1145/3465416.3483294

Verma, A., Lamsal, K., & Verma, P. (2022). An investigation of skill requirements in artificial intelligence and machine learning job advertisements. *Industry and Higher Education*, *36*(1), 63–73. https://doi.org/10.1177/0950422221990990

Vlasceanu, M., & Amodio, D. M. (2022). Propagation of societal gender inequality by internet search algorithms. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(29), 1–8. https://doi.org/10.1073/PNAS.2204529119

Vliegenthart, R., & van Zoonen, L. (2011). Power to the frame: Bringing sociology back to frame analysis. *European Journal of Communication*, *26*(2), 101–115. https://doi.org/10.1177/0267323111404838

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *International Journal of Human Resource Management*, *33*(6), 1237–1266. https://doi.org/10.1080/09585192.2020.1871398

Wajcman, J. (2009). Feminist theories of technology. *Cambridge Journal of Economics*, *34*(1), 143–152. https://doi.org/10.1093/cje/ben057

Walters, A. S. (2021). The contribution of psychological assessment to systemic racism. *The Brown University Child and Adolescent Behavior Letter*, *37*(9), 1. https://doi.org/10.1002/cbl.30572

Weitz, R. (2001). Women and their hair: Seeking power through resistance and accommodation. *Gender and Society*, *15*(5), 667–686. https://doi.org/10.1177/089124301015005003

West, S. M., Whittaker, M., & Crawford, K. (2019). *DISCRIMINATING SYSTEMS: Gender, Race, and Power in AI*. http://cdn.aiindex.org/2018/AI Index 2018 Annual Report.pdf.

Whysall, Z. (2017). Cognitive Biases in Recruitment, Selection, and Promotion: The Risk of Subconscious Discrimination. In V. Caven & S. Nachmias (Eds.), *Hidden Inequalities in the Workplace* (1st ed., pp. 215–243). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-59686-0_9

Wille, B., Wiernik, B., Vergauwe, J., Vrijdags, A., & Trbovic, N. (2018). Personality characteristics of male and female executives: Distinct pathways to success? *Journal of Vocational Behavior*, *106*, 220–235. https://doi.org/10.1016/j.jvb.2018.02.005

Wilson, C., Ghosh, A., Jiang, S., Mislove, A., Baker, L., Szary, J., Trindel, K., & Polli, F. (2021). Building and auditing fair algorithms: A case study in candidate screening. *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 666–677. https://doi.org/10.1145/3442188.3445928

Woods, H. S. (2018). Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication*, *35*(4), 334–349. https://doi.org/10.1080/15295036.2018.1488082

World Economic Forum. (2021). *The global gender gap report*. https://www.weforum.org/reports/global-gender-gap-report-2021

Wyatt, S. (2007). Technological Determinism Is Dead; Long Live Technological Determinism. In E. J. Hackett, O. Amsterdamska, M. E. Lynch, J. Wajcman, S. Sismondo, W. E. Bijker, S. Turner, C. Thorpe, B. Latour, & A. E. Clarke (Eds.), *The Handbook of Science and Technology Studies*. MIT Press.

Wynn, A. T., & Correll, S. J. (2018). Combating Gender Bias in Modern Workplaces. In B. J. Risman, C. M. Froyum, & W. J. Scarborough (Eds.), *Handbook of the Sociology of Gender* (pp. 509–521). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-76333-0_37

Yarger, L., Cobb Payton, F., & Neupane, B. (2019). Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Information Review*, *44*(2), 383–395. https://doi.org/10.1108/OIR-10-2018-0334

Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*(June), 1–14. https://doi.org/10.1016/j.caeai.2023.100147

Young, I. M. (2006). Education in the context of structural injustice: A symposium response. *Educational Philosophy and Theory*, *38*(1), 93–103. https://doi.org/10.1111/j.1469-5812.2006.00177.x

Young, I. M. (2011). *Responsibility for Justice*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195392388.001.0001

Young, M. C. (2010). Gender Differences in Precarious Work Settings. *Industrial Relations*, *65*(1), 74–97. https://doi.org/10.7202/039528ar

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, *16*(3), 1–16. https://doi.org/10.1111/soc4.12962

Zhang, L., & Yencha, C. (2022). Examining perceptions towards hiring algorithms. *Technology in Society*, *68*(May), 1–12. https://doi.org/10.1016/j.techsoc.2021.101848

Zhou, N., Zhang, Z., Nair, V. N., Singhal, H., & Chen, J. (2022). Bias, Fairness and Accountability with Artificial Intelligence and Machine Learning Algorithms. *International Statistical Review*, *90*(3), 468–480. https://doi.org/10.1111/insr.12492

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist - it's time to make it fair. *Nature*, *559*(7714), 324–326. https://doi.org/10.1038/d41586-018-05707-8

Zuloaga, L. (2020). *Nonverbal Communication in Interview Assessments*. HireVue Blog. https://www.hirevue.com/blog/hiring/nonverbal-communication-in-interview-assessments

APPENDICES

Appendix 1 – Interview Guide (Vendors)

Clementine Collett - Interview Guide
TYPE 1: VENDORS

Statement before commencement:

- **Research aims:** we'll be thinking together today about how technology can be designed and implemented to improve diversity and reduce bias. I'd really like this to be a knowledge-sharing exercise, and I'm just interested to see how you think about these topics. I'd like to focus on gender bias, but I realise that we might also touch on racial bias, educational/class biases and so on, as these all intersect.
- **Recording and consent:**
  - Just to remind you that this is being recorded, but that all data will be anonymised, and will be downloaded and stored in line with the Data Protection Act 1998.
  - I've got your consent form. Any questions about that?
  - If you want to stop the interview or withdraw at any point, just let me know, that's not a problem. At that point I'll destroy all your data and won't use anything within the research.
  - However, these interviews really shouldn't include anything sensitive or controversial.
- **Other:**
  - I might take notes or just stop to think about what you've said at some points, I hope that's ok.
  - Please feel free to raise anything you think is relevant which I haven't mentioned. I want the interview to flow as much as possible and for you to feel you're contributing the important things.
- **Do you have any questions before we start?**

Main Interview

START RECORDING

There has been a lot of coverage in the news (and some focus in academia) on the ethics of AI or automated recruitment systems and some of this discussion has centred around whether these systems worsen bias and discrimination, or, in fact, improve it. Thinking specifically about gender bias:

1. **SYSTEM GENDER BIAS: Could you walk me through how your system deals with gender bias?**
   Potential follow-ups:
   a. Could you tell me a bit about what you think gender bias IS in the context of hiring?
   b. How is this understanding TRANSLATED into the system?

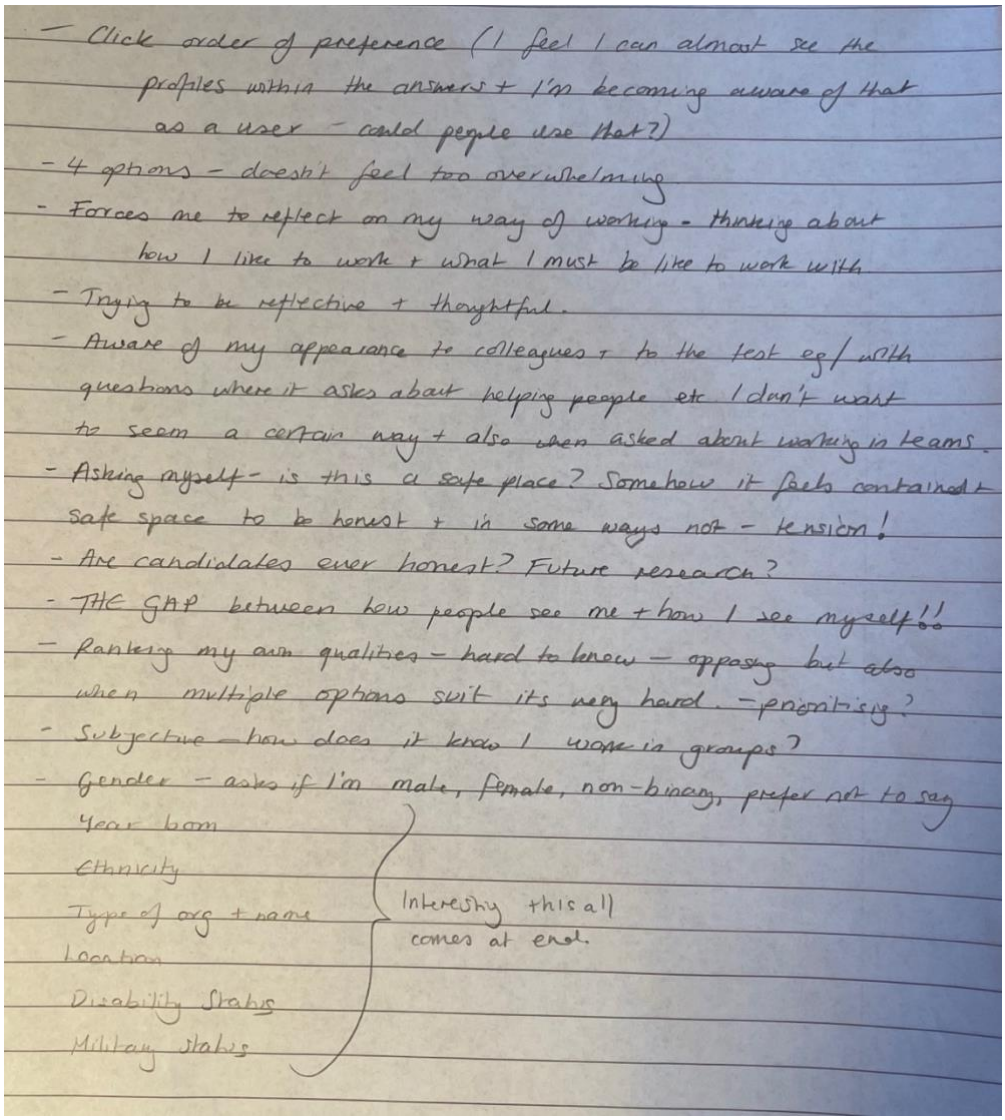
   c. What are the key things which need to happen, from your experience, which would allow recruitment systems to be used in the workplace to improve diversity and reduce gender bias?
2. **<u>COMPARISON:</u> Talk me through how these systems compare to alternative means of recruiting?**
3. **<u>SUCCESS:</u> What does a successful system look like to you in terms of bias, and do you track any measurables to look at this?**
   Potential follow-up:
   a. How do you think these systems can measure job competence?

Other ideas for questions / if time:

4. **<u>RESPONSIBILITY:</u> Of course, the business model works here because there's demand for these systems in the market, but what responsibility does your company have in the wider conversation about gender, bias, and other related issues?**
5. **<u>MEANING OF AI:</u> I've found during this research that a lot of people seem to say that their systems don't use AI, but others do. Why is that? And would you say your system uses AI?**

We've talked about lots of interesting things, are there any other things that you think we haven't spoken about which you think are important?

Appendix 2 – Fieldwork Journal for Free Assessment



- Click order of preference (I feel I can almost see the profiles within the answers + I'm becoming aware of that as a user - could people use that?)
- 4 options - doesn't feel too overwhelming
- Forces me to reflect on my way of working - thinking about how I like to work + what I must be like to work with
- Trying to be reflective + thoughtful.
- Aware of my appearance to colleagues + to the test eg/ with questions where it asks about helping people etc I don't want to seem a certain way + also when asked about working in teams.
- Asking myself - is this a safe place? Somehow it feels contained + safe space to be honest + in some ways not - tension!
- Are candidates ever honest? Future research?
- THE GAP between how people see me + how I see myself!!
- Ranking my own qualities - hard to know - opposing but also when multiple options suit its very hard. - prioritising?
- Subjective - how does it know I work in groups?
- Gender - asks if I'm male, female, non-binary, prefer not to say

Year born
Ethnicity
Type of org + name
Location
Disability Status
Military status

Interesting this all comes at end.

Appendix 3 – Interview Guide (Regulators)

Questions for AI Regulation Expert at European Commission

1. A lot of these systems say they don't use AI – how are you dealing with that? For example, if a company creating these systems can say they don't use AI would this be a loophole in the regulation?

2. What was the process of labelling systems as high risk?

3. High risk means subject to specific legal requirements and a conformity assessment – could you explain how the legal requirements and conformity assessment outlined would be applied to these systems?

4. Could become a global standard – so technologies sold all over the world or just into the EU?

5. What are high quality datasets (point 44)? And high-risk systems should perform accuracy (point 49) this means different things in different contexts – in the hiring context different? Also mentioned in chapter 2 of the proposal – that data shall be relevant, representative, free of errors and complete. How will you measure this and implement these standards?

6. Will this put more responsibility on the vendor companies? As you know now this isn't the case.

Appendix 4 – Screenshots of Information Sheet and Consent Forms

Oxford Internet Institute
University of Oxford
1 St Giles'
Oxford
OX1 3JS

Clementine Collett [DPhil student]
clementine.collett@oii.ox.ac.uk
Professor Victoria Nash [Supervisor]
victoria.nash@oii.ox.ac.uk

DIGITAL INTERVIEWS: PARTICIPANT CONSENT FORM

CUREC Approval Reference: SSH_OII_CIA_21_088

The Design and Implementation Process of AI Recruitment Technology

Please complete this form and return to: clementine.collett@oii.ox.ac.uk

- ☐ I have read and understood the attached information sheet. I understand the aims and process of the research, who will have access to the data, how it will be stored and what will happen to the data at the end of the study. I realise that by signing this form I am consenting to everything in the information sheet.
- ☐ I consent to my contributions to the recorded, online, qualitative, semi-structured interview run by doctoral student Clementine Collett from the Oxford Internet Institute, University of Oxford, being included in research, including academic publications, reports, conference papers and the researcher's doctoral thesis.
- ☐ I understand that if my company choose to remain anonymous within the research, my identity will also be automatically anonymised within the research. I understand that if my company choose to appear named within the research, I can choose whether or not to appear named or remain anonymous. My decision about this is signalled at the bottom of this consent form. I understand that being anonymised might still specify my department/role type such as 'HR – Company 2' or 'Data Scientist – Company 1' but will not give my name. I understand that being named would include my name, job title and company. I understand that if I later wish to change my decision about my anonymity in this research, I must send a new, updated, signed version of this consent form to the researcher. At this point, I understand that although the researcher will not be able to change the status of anonymity for anything already submitted for publication, they will change my status of anonymity accordingly in all future publications or submissions.
- ☐ I agree to the data, video recordings, voice recordings, the name of the company and the name of the individual participants being stored securely and confidentially, in line with the UK Data Protection Act 1998.
- ☐ I understand that I can withdraw at any time prior to the researcher submitting for any academic publication by simply informing the researcher, and that this will incur no penalty or loss. Any data collected at this point will not be used for research purposes and all data collected will be destroyed.
- ☐ I understand that the project has received ethics clearance through the University of Oxford's ethical approval process for research involving human participants. I understand how to raise a concern or make a complaint.

I agree for my interview video/audio recording to be accessed by the researcher	Yes	☐	No	☐
I understand how recordings will be used in research outputs	Yes	☐	No	☐
If my company chooses to appear named within the research, I would like to be:	NAMED	☐	ANONYMOUS	☐

I consent to participation in this study as outlined above.

Name: _____

Signature: _____ **Date:** _____

Name of researcher: _____

Signature: _____ **Date:** _____

Oxford Internet Institute
University of Oxford
1 St Giles'
Oxford
OX1 3JS

Clementine Collett [DPhil student]
clementine.collett@oii.ox.ac.uk
Professor Victoria Nash [Supervisor]
victoria.nash@oii.ox.ac.uk

DIGITAL INTERVIEWS AND DIGITAL ETHNOGRAPHY: COMPANY CONSENT FORM

CUREC Approval Reference: SSH_OII_CIA_21_088

The Design and Implementation Process of AI Recruitment Technology

Please complete this form and return to:clementine.collett@oii.ox.ac.uk

- ☐ Subject to the consent of individual participants, I consent to the employees of the undersigned company participating in digital qualitative semi-structured interviews run by Clementine Collett, a doctoral student at the Oxford Internet Institute, University of Oxford, and/or digital ethnographic research being conducted within the company (in the form of meeting/event shadowing).
- ☐ I understand that digital interviews **will be recorded** by the researcher, and that digital ethnographic research/shadowing **will not be recorded** by the researcher.
- ☐ I understand that the researcher, Clementine Collett, will provide the company with a short report on findings from the interviews within a reasonable time frame which is mutually agreed by the company and the researcher. This will be in a format mutually agreed by the researcher and company.
- ☐ On the company's behalf, I consent to this research, including direct quotations from interviews and any other data collected, being included in academic publications, reports, conferences papers and the researcher's doctoral thesis.
- ☐ After the research has been conducted and a short report provided to the company, the company representative, signer of this form, will be given the option for the company to remain anonymous or appear named in the research. This will be indicated on a separate consent form, the 'Anonymisation: Addendum Consent Form', which will be sent by the researcher. This specifically requests the company's decision about anonymity. I understand that if the company later wishes to change their decision about anonymity, they must send a new, updated, signed 'Anonymisation: Addendum Consent Form' to the researcher. At this point, although the researcher will not be able to change the status of anonymity for anything already submitted for publication, they will change the status of the anonymity of the company accordingly in all future publications or submissions. If the company choose for the company to be named, each individual participant will be given the option to appear named or to be anonymised in the research, which they will indicate on their individual consent form. If the individual participants wish to change their decision about their personal anonymity, they also must send a new, updated, signed individual consent form. Again, at this point, although the researcher will not be able to change the status of anonymity for anything already submitted for publication, they will change the status of anonymity of the company accordingly in all future publications or submissions.
- ☐ In all cases, the company will be given an opportunity, within a reasonable time frame of submission, to provide comments and feedback on materials planned for submission.
- ☐ I understand that the company can withdraw at any time prior to the researcher submitting for any academic publication by simply informing the researcher, and that this will incur no penalty or loss. Any data collected at this point will not be used for research purposes and all data collected will be destroyed.
- ☐ The company can use the data or findings specific to their own company in the following ways:
    - At any time after the interviews and/or ethnography have been conducted and a report has been shared with the company, the company can use the data and findings regarding their own company for purposes such as internal reporting or marketing.

Oxford Internet Institute
University of Oxford
1 St Giles'
Oxford
OX1 3JS

Clementine Collett [DPhil student]
clementine.collett@oii.ox.ac.uk
Professor Victoria Nash [Supervisor]
victoria.nash@oii.ox.ac.uk

INFORMATION SHEET FOR COMPANIES AND PARTICIPANTS

The Design and Implementation Process of AI Recruitment Technology

Ethics Approval Reference: [SSH_OII_CIA_21_088]

****Once you have read and fully understood this document, if you wish to give your informed consent to participate in this research, please sign the attached Adobe PDF consent form using the Adobe Sign function and return it to clementine.collett@oii.ox.ac.uk. After that, I'll respond to arrange a time for your interview****

Invitation to participate:

Before you decide to take part in this study, it is important that you understand the purpose of the research and what the research will involve. Please take time to read the following information sheet carefully. The researcher can be contacted at the above email address if anything is unclear or if you would like more information.

Summary of the research:

This research aims to study how technology can be successfully designed and implemented into the workplace to improve diversity and reduce bias. In order to explore this, the research looks through the lens of AI-powered recruitment technology. It focuses on how gender bias, in particular, is conceptualised and operationalised by both vendor and client in these systems.

Why have I been invited to take part?

You are being invited because you either work for an AI recruitment technology company, or you are part of a client company using which uses AI recruitment technology. You are being sent this information sheet because either: the researcher is requesting a one-to-one digital interview with you, or the researcher is requesting to shadow a meeting or event of which you are a part (or both of these apply).

Do I have to take part?

No. Participation in this study is voluntary. This applies to both the company and the individual. You can ask questions about the research before deciding whether or not to take part. If you do agree to take part, you may withdraw yourself from the study, without giving a reason, and without negative consequences, by simply advising the researcher of your decision. This can happen before, during, or after the interview, right up until the researcher has submitted for their first academic publication based on this study. If you decide to withdraw, the information collected about you will not be used in the study and will be destroyed.

What is the procedure if I decide to take part?

There are two optional parts to this research, both of which are entirely voluntary and both of which require informed consent:

- Upon publication of the researcher's findings, the company can use the publication of findings for marketing purposes.

- I have read and understood the attached information sheet. I understand the aims and process of the research, who will have access to the data, how it will be stored and what will happen to the data at the end of the study. I realise that by signing this form I am consenting to everything in the information sheet.
- I agree to the data, video recordings, voice recordings, the name of the company and the name of the individual participants being stored securely, confidentially and anonymously, in line with the UK Data Protection Act 1998.
- I understand that the project has received ethics clearance through the University of Oxford's ethical approval process for research involving human participants. I understand how to raise a concern or make a complaint.

I agree to interview video/audio recordings being accessed by the researcher, subject to individual participant consent	Yes	☐	No	☐
I understand how recordings will be used in research outputs	Yes	☐	No	☐

I consent to this company participating in this study.

Name: _____

Company: _____

Signature: _____ **Date:** _____

Name of researcher: _____

Signature: _____ **Date:** _____

Oxford Internet Institute
University of Oxford
1 St Giles'
Oxford
OX1 3JS

Clementine Collett [DPhil student]
clementine.collett@oii.ox.ac.uk
Professor Victoria Nash [Supervisor]
victoria.nash@oii.ox.ac.uk

INFORMATION SHEET FOR COMPANIES AND PARTICIPANTS

The Design and Implementation Process of AI Recruitment Technology

Ethics Approval Reference: [SSH_OII_CIA_21_088]

****Once you have read and fully understood this document, if you wish to give your informed consent to participate in this research, please sign the attached Adobe PDF consent form using the Adobe Sign function and return it to clementine.collett@oii.ox.ac.uk. After that, I'll respond to arrange a time for your interview****

Invitation to participate:

Before you decide to take part in this study, it is important that you understand the purpose of the research and what the research will involve. Please take time to read the following information sheet carefully. The researcher can be contacted at the above email address if anything is unclear or if you would like more information.

Summary of the research:

This research aims to study how technology can be successfully designed and implemented into the workplace to improve diversity and reduce bias. In order to explore this, the research looks through the lens of AI-powered recruitment technology. It focuses on how gender bias, in particular, is conceptualised and operationalised by both vendor and client in these systems.

Why have I been invited to take part?

You are being invited because you either work for an AI recruitment technology company, or you are part of a client company using which uses AI recruitment technology. You are being sent this information sheet because either: the researcher is requesting a one-to-one digital interview with you, or the researcher is requesting to shadow a meeting or event of which you are a part (or both of these apply).

Do I have to take part?

No. Participation in this study is voluntary. This applies to both the company and the individual. You can ask questions about the research before deciding whether or not to take part. If you do agree to take part, you may withdraw yourself from the study, without giving a reason, and without negative consequences, by simply advising the researcher of your decision. This can happen before, during, or after the interview, right up until the researcher has submitted for their first academic publication based on this study. If you decide to withdraw, the information collected about you will not be used in the study and will be destroyed.

What is the procedure if I decide to take part?

There are two optional parts to this research, both of which are entirely voluntary and both of which require informed consent:

- Digital Ethnography/Shadowing. This would include shadowing virtual meetings within the organisation, or any virtual events taking place run by or within the organisation.
- Digital Interviews. This would include a one-to-one interview over Zoom Pro, which will last between 60-75 minutes.

Some organisations might prefer to only participate in the digital interviews. There is by no means any obligation for the organisation to participate in both of these procedures (digital interviews and digital ethnography/shadowing).

Digital ethnography will occur as virtual meeting shadowing and virtual event shadowing/participation (for events which are run by/within the organisation). The researcher will request to attend any meetings within the organisation which are relevant to the research. This might include things surrounding design, sales/marketing, client relation, or product development. The researcher will not record the meetings but will simply observe and make anonymised field notes. The organisation will inform the researcher of meetings which they are comfortable for the researcher to attend, if any, and then informed consent will be sought from all individual participants of the meeting. For larger events (e.g. public events), informed consent will only be sought from the organisation itself, and not from every single participant. As well as being entirely anonymised, all data will be stored securely as outlined further below.

Digital interviews will happen online, via Zoom Pro, and they will be end-to-end encrypted to ensure privacy. They will be recorded via Zoom, but due to the end-to-end encryption, this will not be recorded onto the Zoom cloud. Instead, it will be recorded directly onto the interviewer's computer. It will also be recorded on the interviewer's personal dictaphone as a backup, in case anything goes wrong with the recording. These recordings will be uploaded and stored on the secure Nexus365 One Drive (approved for security by the University of Oxford) and every file will be password protected. The interviewer may occasionally also take shorthand written notes during the interview.

The interview will last between an hour and an hour and fifteen minutes at a time which is convenient for both participant and researcher (a mutually agreed time).

Some participants might be asked if they are willing to do a digital walkthrough of material such as marketing or sales material, of the system itself, or of a generic PowerPoint to explain how the system works (if they have one to hand). This is completely optional.

Within vendor companies, digital interviews will take place with those in roles such as:

- Leadership/senior management/C-suite
- Sales/marketing/customer relations
- Production/design/R&D/data science

Within client companies, digital interviews will take place with those in roles such as:

- Leadership/senior management/C-suite
- HR/recruitment/purchasing/R&D

The interviews will be semi-structured, so the discussion will be guided by us covering a number of topics. However, it's important that you feel, as a participant, that you can bring up any points which you'd like to discuss.

The guiding topics will include:

- **Client Need:** What were the recruitment processes carried out by the client before commissioning the technology? What were their main problems with their previous recruitment process, and what were their motivations for approaching the vendor?

- **Bias:** How is bias understood and conceptualised? How is this understanding then operationalised in the design and functionality of the system?

- **Gender and Intersectionality:** How is intersectionality captured in the system, or does it take a single-axis approach, and if so, why? How is gender conceptualised? And how does the system operationalise gender? Does it recognise gender, or is it blind to gender, and if so, why?

- **Design Process:** What does the design process entail, step-by-step, for each new project/client, and does this ever vary, and if so, why? How much is this a co-creation between client and vendor, and if so, how does this co-creation work well or what are the pitfalls? How transparent is this process and how much does the client understand the functionality of the product? What does communication between client and vendor look like prior to usage and during usage?

- **System Functionality:** Broadly speaking, how does the system work? What are its main assumptions? How has this design developed over time and why?

- **Implementation and Performance Measures:** What is the process of implementation like in the client company? What have the main differences been from the recruiting practices used before? How are the deliverables (e.g. reduced bias and improved diversity) measured? What were the main successes or sticking points when implementing the system? How has diversity, bias and organisational culture changed in the client company since implementation?

If the researcher doesn't feel that the interview has covered all the potential topics during the allotted time in the first interview, they might request a second interview. Again, this is completely optional, and the participant should feel no pressure.

These interviews shouldn't include anything sensitive or controversial. However, if at any point you feel uncomfortable with the questions, you are free to stop the interview or withdraw completely at any point.

What happens to my data?

Data will be stored on a personal, password protected computer on Oxford University's Nexus365 One Drive for Business Sharepoint, which is approved as a local secure storage site by the University of Oxford. All documents and files will be password protected and encrypted where possible. All data will be destroyed at the end of the researcher's doctorate and after any outstanding academic publications have been resolved, except the consent forms, which will be maintained for three years after the research end date.

This data includes:

- Signed consent forms from participants.
- Audio/video recordings of digital interviews.
- Field notes from digital interviews and digital ethnographic research.
- Transcriptions of interviews.
- Contact details and names of participants.

In field notes and interview transcriptions, each participant will be anonymised via a participant identity code. This means that the field notes and transcriptions will all be anonymised and written under pseudonyms. The consent forms, audio/video recordings, and contact details, which are impossible to make anonymous, will all be password protected and encrypted where possible.

The researcher and supervisor(s) will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/.

Will the research be anonymised?

The company representative (a senior member of the organisation) who signs the company consent form will be given the option for the company to remain anonymous within the research or appear named. This will be

completely at their discretion and will be decided after the research has taken place and the company has been provided with a report of the findings. This will be indicated on a separate consent form, the 'Anonymisation: Addendum Consent Form', which will be sent by the researcher. This specifically requests the company's decision about anonymity. I understand that if the company later wishes to change their decision about anonymity, they must send a new, updated, signed 'Anonymisation: Addendum Consent Form' to the researcher. At this point, although the researcher will not be able to change the status of anonymity for anything already submitted for publication, they will change the status of the anonymity of the company accordingly in all future publications or submissions.

If the company choose to appear anonymous, all individuals will be automatically anonymised within the research. If the company choose for the company to be named, each individual participant will be given the option to appear named or to be anonymised in the research, which they will indicate on their individual consent form. If the individual participants wish to change their decision about their personal anonymity, they also must send a new, updated, signed individual consent form. Again, at this point, although the researcher will not be able to change the status of anonymity for anything already submitted for publication, they will change the status of anonymity of the company accordingly in all future publications or submissions.

Whether anonymous or named, participants might be directly quoted in the research.

If anonymised, participants might be referred to by department/role type and their company pseudonym, for example: 'HR – Company 1' or 'Data Scientist – Company 2'.

Will the research be published?

The data, information and direct quotations will be used in the researcher's doctoral thesis, academic publications and may also be presented at conferences. As said above, it is at the total discretion of the company and the individual participant as to whether they appear named or anonymous within the research.

Who is funding this research?

This research is funded by the Economic and Social Research Council (ESRC) Doctoral Training Partnership (DTP).

Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee - Ethics Approval Reference: [SSH_OII_CIA_21_088].

Who do I contact if I have a concern about the study or I wish to complain?

This study has been reviewed by, and received ethics clearance through, the University of Oxford's Central University Research Ethics Committee, Ethics Approval Reference: [SSH_OII_CIA_21_088].

If you have a concern about any aspect of this study, please contact Clementine Collett (clementine.collett@oii.ox.ac.uk) or Victoria Nash (victoria.nash@oii.ox.ac.uk) and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Oxford Internet Institute Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible: Chair, Oxford Internet Institute Research Ethics Committee; Email: drec@oii.ox.ac.uk; Address: Oxford Internet Institute, University of Oxford, 1 St Giles', Oxford OX1 3JS.

What do I do next?

1. Please sign the attached Adobe PDF consent form using the Adobe Sign function and return it to clementine.collett@oii.ox.ac.uk.
2. I'll respond to arrange a time for your interview.