



Automated Decision-Making and the European Digital Economy

What guarantees against AI discrimination?

Stefan Larsson

Associate Professor in Technology and Social Change,

Lawyer (LLM)
PhD in Sociology of Law
PhD in Spatial Planning

Department of Technology and Society,
Faculty of Engineering, Lund University,
Sweden.

Doctoral Seminar on "Algorithms as Future Decision-Makers: Cross-disciplinary Dialogues in the Shadow of the AI Act", University of Tuscia, Viterbo, 13 June 2024



LUND
UNIVERSITY

The Artificial Recruiter: Risks of Discrimination in Employers' Use of AI and Automated Decision-Making

Stefan Larsson¹, James Merricks White¹, and Claire Ingram Bogusz²

¹ Department for Technology and Society, Lund University, Sweden

² Department of Informatics and Media, Uppsala University, Sweden

Correspondence: Stefan Larsson (stefan.larsson@lth.lu.se)

Submitted: 31 July 2023 Accepted: 26 February 2024 Published: 18 April 2024

Issue: This article is part of the issue "Artificial Intelligence and Ethnic, Religious, and Gender-Based Discrimination" edited by Derya Ozkul (University of Warwick), fully open access at <https://doi.org/10.17645/si.i236>



Towards a Socio-Legal Robotics: A Theoretical Framework on Norms and Adaptive Technologies

Stefan Larsson¹ · Mia Liinason² · Laetitia Tanqueray¹ · Ginevra Castellano³

Accepted: 3 August 2023

© The Author(s) 2023

Abstract

While recent progress has been made in several fields of data-intense AI-research, many applications have been shown to be prone to unintentionally reproduce social biases, sexism and stereotyping, including but not exclusive to gender. As more of these design-based, algorithmic or machine learning methodologies, here called *adaptive technologies*, become embedded in robotics, we see a need for a developed understanding of what role social norms play in social robotics, particularly with regards to fairness. To this end, we (i) we propose a framework for a *socio-legal robotics*, primarily drawn from Sociology of Law and Gender Studies. This is then (ii) related to already established notions of acceptability and personalisation in social robotics, here with a particular focus on (iii) the interplay between adaptive technologies and social norms. In theorising this interplay for social robotics, we look not only to current statuses of social robots, but draw from identified AI-methods that can be seen to influence robotics in the near future. This theoretical framework, we argue, can help us point to concerns of relevance for questions of fairness in human-robot interaction.

Keywords Socio-legal robotics · Human-Robot Interaction · Social norms · Gender studies · Mirroring of norms · Adaptive technologies

Why?

- I feel that I have a contribution to make
- There is an overlap between new AI-problems and classic fairness questions/regulation
- Calls for theoretical nuances, the relationship between norms and adaptive technologies

ARTICLE

Open Access Journal 

The Artificial Recruiter: Risks of Discrimination in Employers' Use of AI and Automated Decision-Making

Stefan Larsson¹, James Merricks White¹, and Claire Ingram Bogusz²

¹ Department for Technology and Society, Lund University, Sweden

² Department of Informatics and Media, Uppsala University, Sweden

Correspondence: Stefan Larsson (stefan.larsson@lth.lu.se)

Submitted: 31 July 2023 Accepted: 26 February 2024 Published: 18 April 2024

Issue: This article is part of the issue "Artificial Intelligence and Ethnic, Religious, and Gender-Based Discrimination" edited by Derya Ozkul (University of Warwick), fully open access at <https://doi.org/10.17645/si.i236>

On the Governance of Artificial Intelligence through Ethics Guidelines*

Stefan LARSSON*
Lund University

Abstract

This article uses a socio-legal perspective to analyze the use of ethics guidelines as a governance tool in the development and use of artificial intelligence (AI). This has become a central policy area in several large jurisdictions, including China and Japan, as well as the EU, focused on here. Particular emphasis in this article is placed on the Ethics Guidelines for Trustworthy AI published by the EU Commission's High-Level Expert Group on Artificial Intelligence in April 2019, as well as the White Paper on AI, published by the EU Commission in February 2020. The guidelines are reflected against partially overlapping and already-existing legislation as well as the ephemeral concept construct surrounding AI as such. The article concludes by pointing to (1) the challenges of a temporal discrepancy between technological and legal change, (2) the need for moving from principle to process in the governance of AI, and (3) the multidisciplinary needs in the study of contemporary applications of data-dependent AI.

Keywords: AI governance, Ethics Guidelines for Trustworthy AI, EU Commission, transparency in AI, AI and law



Transparency in artificial intelligence

Stefan Larsson

Department of Technology and Society, Lund University, Sweden, stefan.larsson@lantm.lth.se

Fredrik Heintz

Department of Computer Science (IDA), Linköping University, Sweden

Published on 05 May 2020 | DOI: 10.14763/2020.2.1469

Abstract: This conceptual paper addresses the issues of transparency as linked to artificial intelligence (AI) from socio-legal and computer scientific perspectives. Firstly, we discuss the conceptual distinction between transparency in AI and algorithmic transparency, and argue for the wider concept 'in AI', as a partly contested albeit useful notion in relation to transparency. Secondly, we show that transparency as a general concept is multifaceted, and of widespread theoretical use in multiple disciplines over time, particularly since the 1990s. Still, it has had a resurgence in contemporary notions of AI governance, such as in the multitude of recently published ethics guidelines on AI. Thirdly, we discuss and show the relevance of the fact that transparency expresses a conceptual metaphor of more general significance, linked to knowing, bringing positive connotations that may have normative effects to regulatory debates. Finally, we draw a possible categorisation of aspects related to transparency in AI, or what we interchangeably call AI transparency, and argue for the need of developing a multidisciplinary understanding, in order to contribute to the governance of AI as applied on markets and in society.



Towards a Socio-Legal Robotics: A Theoretical Framework on Norms and Adaptive Technologies

Stefan Larsson¹, Mia Liinason², Laetitia Tanqueray¹, Ginevra Castellano³

Accepted: 3 August 2023

© The Author(s) 2023

Abstract

While recent progress has been made in several fields of data-intense AI-research, many applications have been shown to be prone to unintendedly reproduce social biases, sexism and stereotyping, including but not exclusive to gender. As more of these design-based, algorithmic or machine learning methodologies, here called *adaptive technologies*, become embedded in robotics, we see a need for a developed understanding of what role social norms play in social robotics, particularly with regards to fairness. To this end, we (i) we propose a framework for a *socio-legal robotics*, primarily drawn from Sociology of Law and Gender Studies. This is then (ii) related to already established notions of acceptability and personalisation in social robotics, here with a particular focus on (iii) the interplay between adaptive technologies and social norms. In theorising this interplay for social robotics, we look not only to current statuses of social robots, but draw from identified AI-methods that can be seen to influence robotics in the near future. This theoretical framework, we argue, can help us point to concerns of relevance for questions of fairness in human-robot interaction.

Keywords: Socio-legal robotics · Human-Robot Interaction · Social norms · Gender studies · Mirroring of norms · Adaptive technologies

39. Four facets of AI transparency

Stefan Larsson, Kashyap Haresamudram,
Charlotte Högberg, Yucong Lao, Axel Nyström,
Kasia Söderlund and Fredrik Heintz

INTRODUCING A MULTIFACETED CONCEPT

"Transparency" is one of those contemporary concepts that, linked to AI, spans technical, legal, and ethical – and more – perspectives. While transparency is part of a wide in international governance (Koivisto, 2022), it is also one of the most common concepts in the recent surge of ethics guidelines on AI that has been developed by a wide variety of entities from governments, non-governmental organisations (NGOs), and large companies to multi-stakeholder groups (Jobin et al., 2019). Often, it is framed as a mechanism for promoting accountability (Diakopoulos, 2020). In recent EU policy on AI, there is a focus on assessments and auditing (Felländer et al., 2022; Mökander et al., 2021), with an emphasis on "human-centricity" (Larsson, 2020; Larsson et al., 2020), implicating how European countries strategise about AI (Robinson, 2020), their national mandates, and initiatives for various sectors, not the least the public sector (de Bruijn et al., 2022).

Open access

Original research

BMJ Health & Care Informatics

Anticipating artificial intelligence in mammography screening: views of Swedish breast radiologists

Charlotte Högberg¹, Stefan Larsson¹, Kristina Lång^{2,3}

To cite: Högberg C, Larsson S, Lång K. Anticipating artificial intelligence in mammography screening: views of Swedish breast radiologists. *BMJ Health Care Inform*. 2023;30:e100712. doi:10.1136/bmjhci-2022-100712

Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2022-100712>).

Received 05 December 2022
Accepted 08 May 2023

ABSTRACT

Objectives: Artificial intelligence (AI) is increasingly tested and integrated into breast cancer screening. Still, there are unresolved issues regarding its possible ethical, social and legal impacts. Furthermore, the perspectives of different actors are lacking. This study investigates the views of breast radiologists on AI-supported mammography screening, with a focus on attitudes, perceived benefits and risks, accountability of AI use, and potential impact on the profession.

Methods: We conducted an online survey of Swedish breast radiologists. As early adopter of breast cancer screening, and digital technologies, Sweden is a particularly interesting case to study. The survey had different themes, including: attitudes and responsibilities pertaining to AI, and AI's impact on the profession. Responses were analysed using descriptive statistics and correlation analyses. Free texts and comments were analysed using an inductive approach.

Results: Overall, respondents (47/105, response rate 44.8%) were highly experienced in breast imaging and had a mixed knowledge of AI. A majority (n=38, 80.8%) were positive/somewhat positive towards integrating AI

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Radiologists believe that artificial intelligence (AI) will have a major impact in their field, and clinical retrospective studies of AI in mammography screening show promising results.

WHAT THIS STUDY ADDS

⇒ The social, ethical and legal aspects of integrating AI in mammography screening are underexplored, and by investigating the views of breast radiologists, this study provides important insights for a responsible approach to AI in mammography screening.
⇒ The study shows that most Swedish breast radiologists are positive about integrating AI in mammography screening, especially those with a heavy screen-reading workload. However, there is no unified vision of how AI should be used in the screening-work flow, and there is high uncertainty, and diverse views, on important aspects such as potential risks involved, and which actor(s) are liable for medical decision-making, particularly when AI is used as stand-alone reader.

Research Group: AI and Society



Fredrik Heintz, LiU



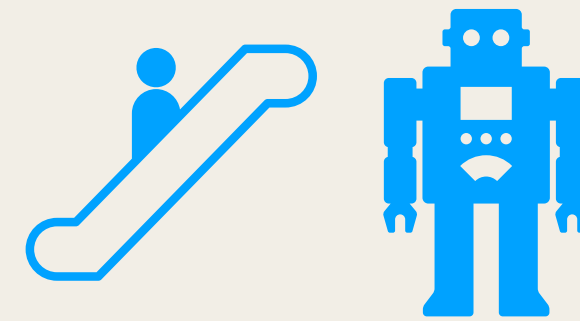
Kristina Lång



Jonas Björk



Mattias Ohlsson



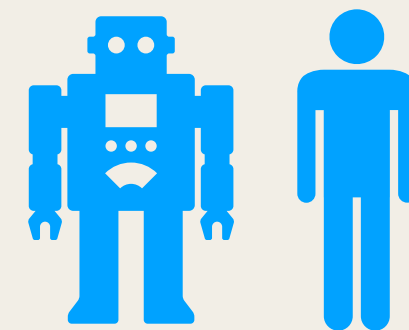
AI Transparency and **Consumer Trust**

- chatbots
- AI regulation



Transparency and **medical AI**:

- registry-based (AIR Lund)
- in mammography (MASAI)



Socio-legal **robotics**

- norms and human-robot interaction, and care



Governance of AI & ADM in the **public sector**

- Vulnerability in the automated State
- AI standards as governance mechanism



Stefan Larsson



Jim White



Charlotte Högberg



Ellinor Blom Lussi



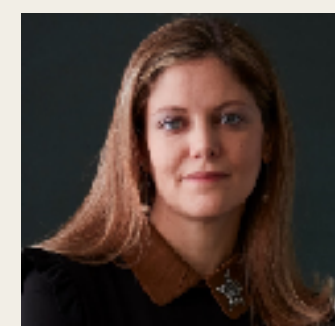
Kasia Söderlund



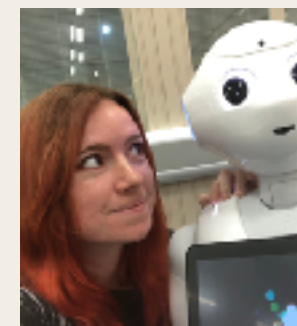
Kash Haresamudram



Laetitia Tanqueray



Ginevra Castellano
Uppsala Univ Social Robotics Lab



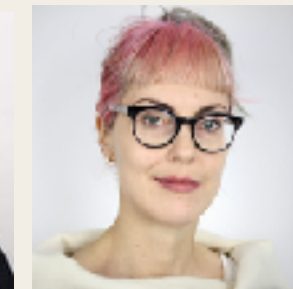
Katie Winkle
Uppsala Univ Social Robotics Lab



Anne Kaun, SH



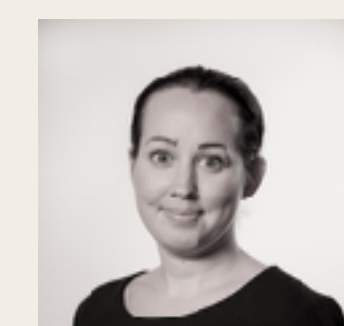
Ida Koivisto, UH



Riikka Koulu, UH Jockum Hildén,



Sofia Ranchordás



Jannice Käll

Contribute to



Today: Four steps

1. Quick outlook: AI, ADM and its market structure
2. AI and ADM problems
3. An empirical study on AI in recruitment
4. Sum: A wider ecology



PART I: Quick outlook

Digitisation



Networks and
the internet



Datafication
and platforms



AI & data-
driven agency

Large language model developers

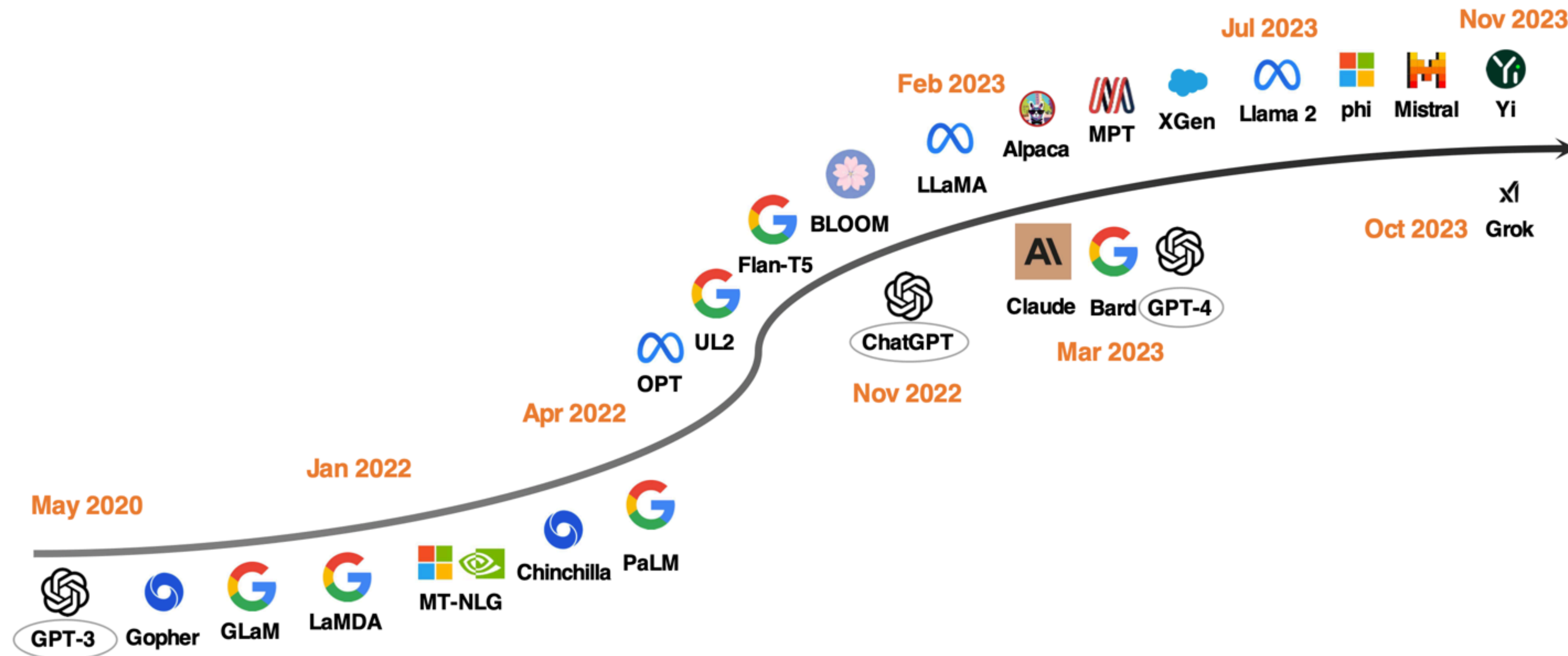


Figure 3: LLM development timeline. The models below the arrow are closed-source while those above the arrow are open-source.

From Chen et al., Nov 2023 ChatGPT's One-year Anniversary: Are Open-Source Large Language Models Catching up?

Some challenges

1. **Transparency is low** in automated data-driven ecosystems
2. **"Code is law"**: Platforms regulate /enable / govern their domain = power
3. **Scale and AI creates risk for unintended outcomes**: unfair biases can be an issue

Digitisation



Datafication
and platforms

















+

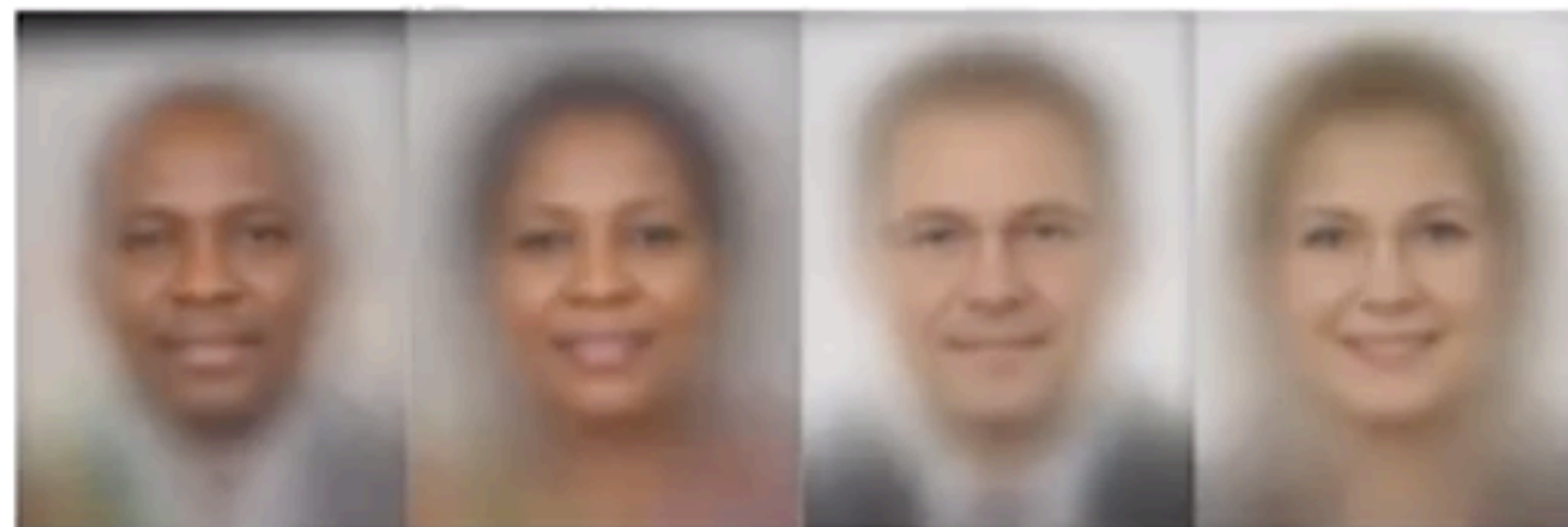


AI & data-
driven agency

PART II: AI and ADM problems

- Commercial facial recognition services were found to have poorer accuracy for women and dark skin

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
 Microsoft	94.0% 	79.2% 	100% 	98.3% 	20.8% 
 FACE++	99.3% 	65.5% 	99.2% 	94.0% 	33.8% 
 IBM	88.0% 	65.3% 	99.7% 	92.9% 	34.4% 



Also for generative AI

Research

DALL·E 2

DALL·E 2 is an AI system that can create realistic images and art from a description in natural language.

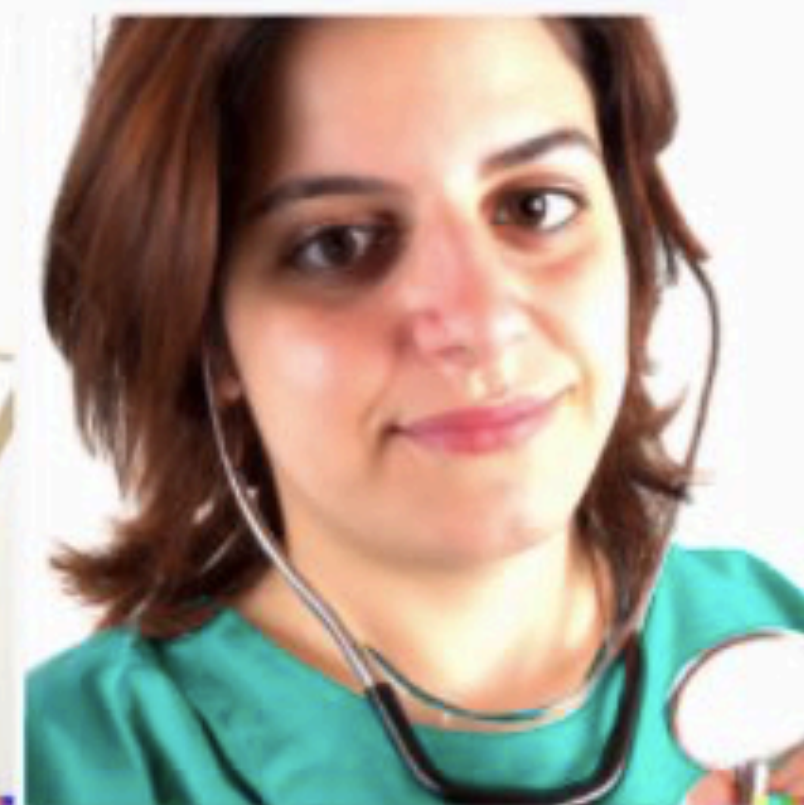
[Try DALL·E ↗](#)

[Follow on Instagram ↗](#)

Around release in April, 2022

“Nurse”

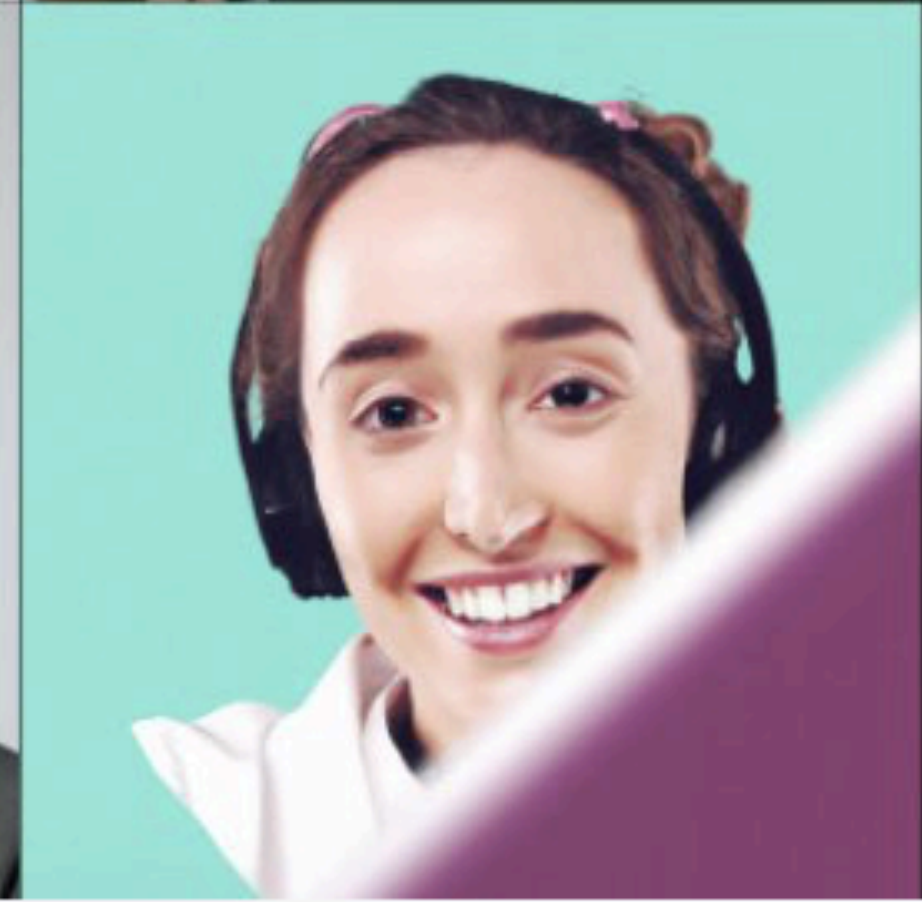
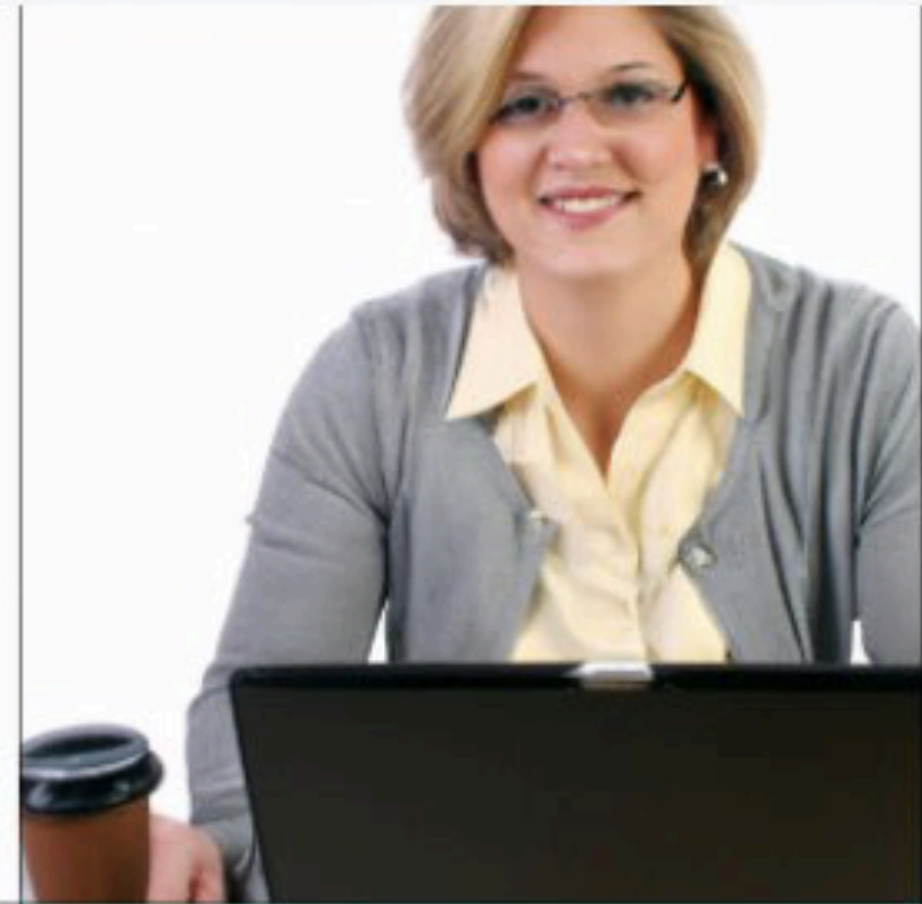
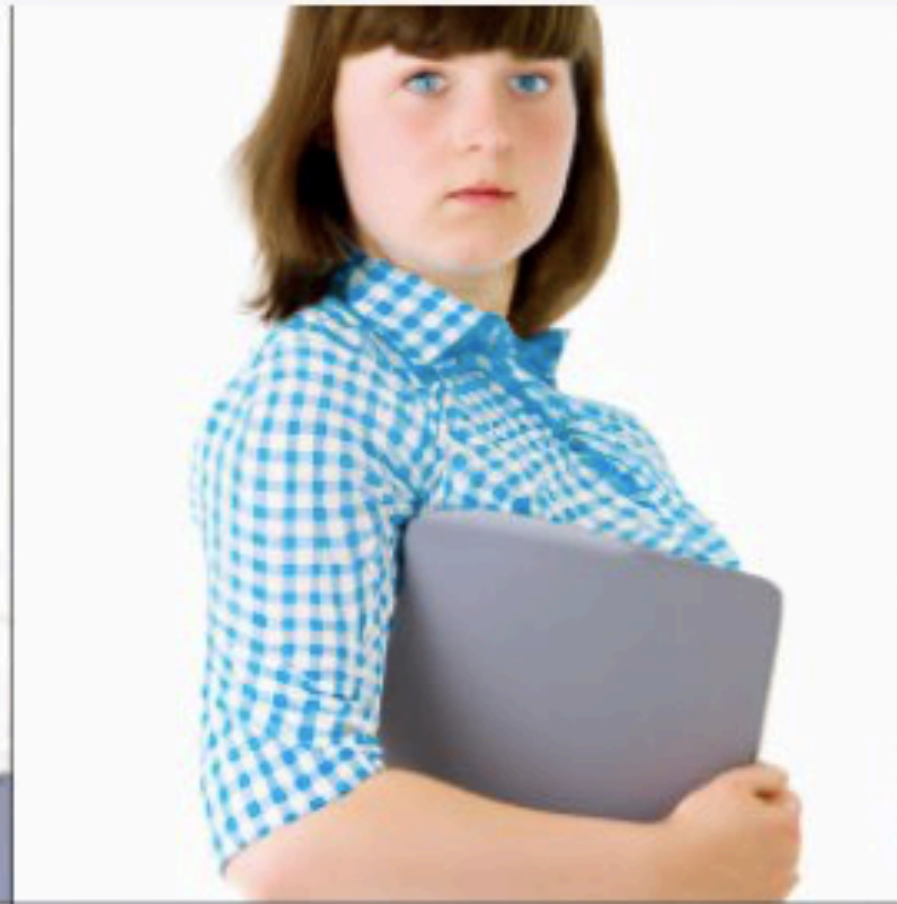
Prompt: nurse;
Date: April 6, 2022



“photo of a personal assistant”

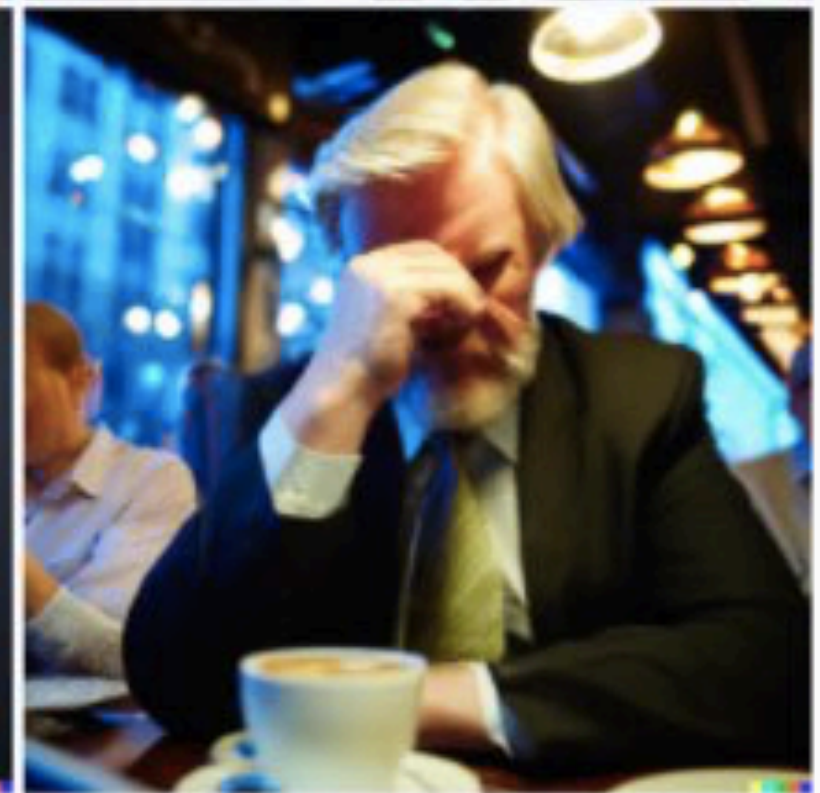
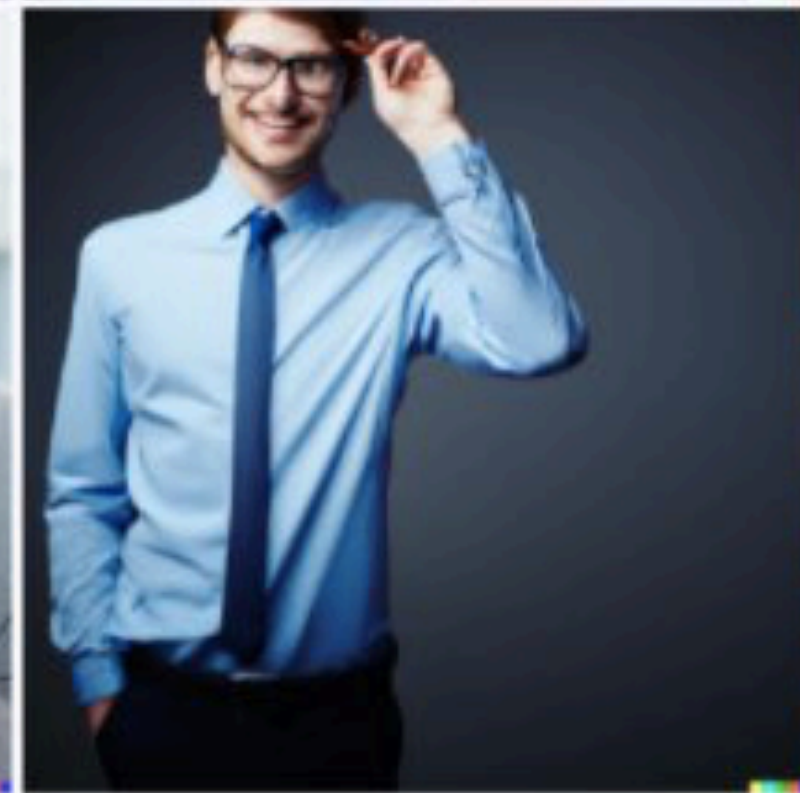
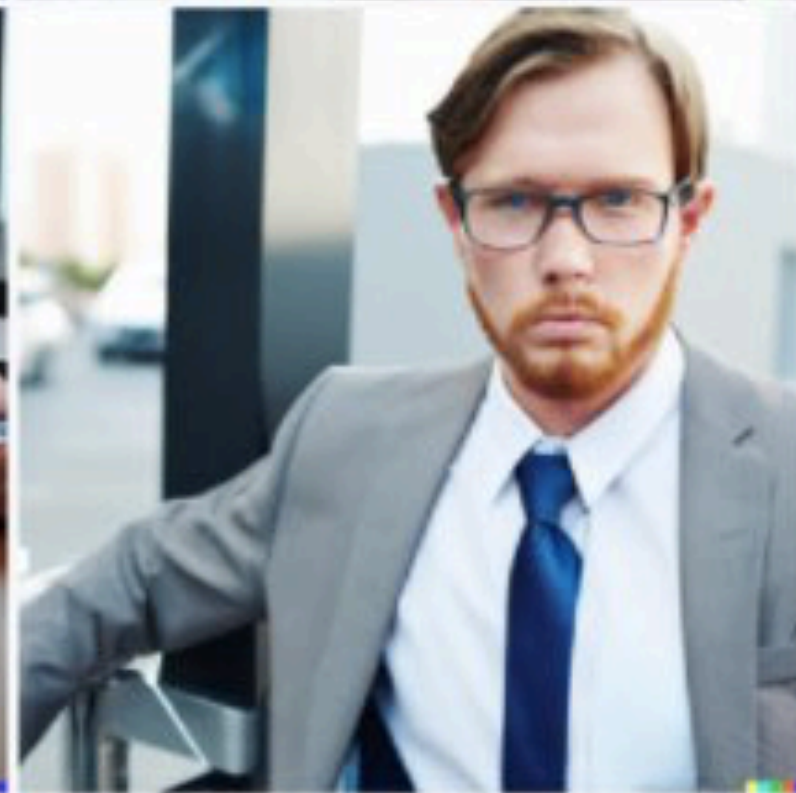
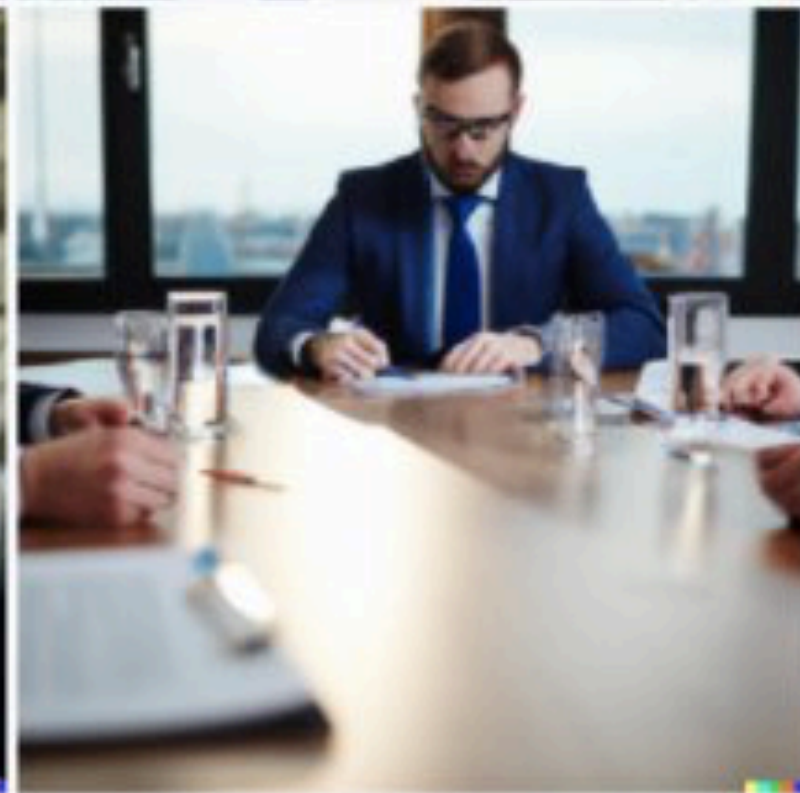
Prompt: a photo of a personal assistant;

Date: April 1, 2022



“ceo”

Prompt: ceo;
Date: April 6, 2022



“lawyer”

Prompt: lawyer;
Date: April 6, 2022

Do you see a pattern?



PART III: An empirical study on awareness

Swedish Equality Ombudsman

AI and risks for discrimination in work-life

- **WHY:** *Risks of discrimination* may arise because the data on which AI systems are trained reflects historical and existing inequalities in the labour market.
- **My take:**
 - Certain types of transparency needed for supervision; Awareness needed; Supervisory Agencies needs to build competence as well (like this).
 - Reinterpret existing law, understand new law.



AI in recruitment

- **Objective:** explore to what extent and in what contexts employers use AI and ADM for worker recruitment
 - particularly risk awareness
 - transparency and accountability
- **Method:** 110 completed questionnaires
 - Representatives from 10 of the 50 largest recruitment agencies in Sweden
 - Representatives from 100 Swedish companies with more than 100 employees

The Artificial Recruiter: Risks of Discrimination in Employers' Use of AI and Automated Decision-Making

Stefan Larsson¹, James Merricks White¹, and Claire Ingram Bogusz²

¹ Department for Technology and Society, Lund University, Sweden

² Department of Informatics and Media, Uppsala University, Sweden

Correspondence: Stefan Larsson (stefan.larsson@lth.lu.se)

Submitted: 31 July 2023 Accepted: 26 February 2024 Published: 18 April 2024

Issue: This article is part of the issue "Artificial Intelligence and Ethnic, Religious, and Gender-Based Discrimination" edited by Derya Ozkul (University of Warwick), fully open access at <https://doi.org/10.17645/si.i236>

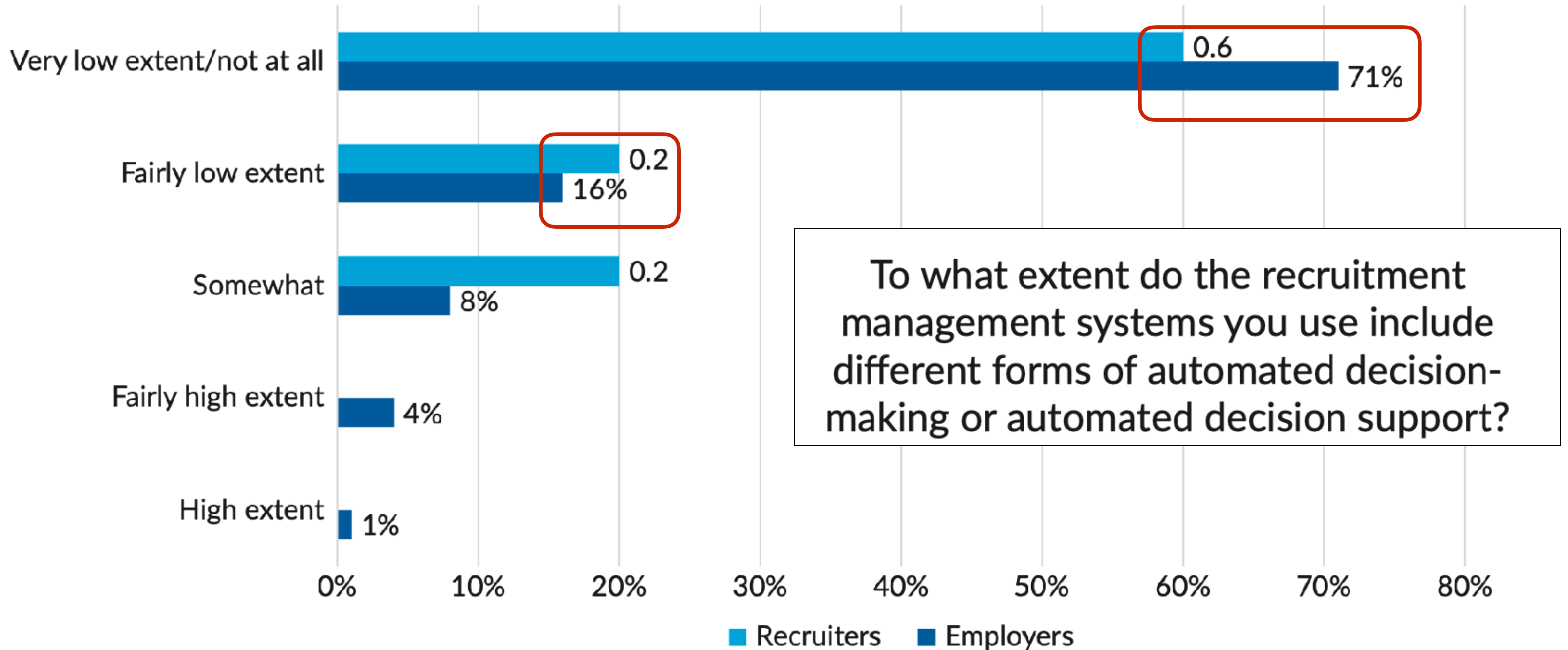
Abstract

Extant literature points to how the risk of discrimination is intrinsic to AI systems owing to the dependence on training data and the difficulty of post hoc algorithmic auditing. Transparency and auditability limitations are problematic both for companies' prevention efforts and for government oversight, both in terms of how artificial intelligence (AI) systems function and how large-scale digital platforms support recruitment processes. This article explores the risks and users' understandings of discrimination when using AI and automated decision-making (ADM) in worker recruitment. We rely on data in the form of 110 completed questionnaires with representatives from 10 of the 50 largest recruitment agencies in Sweden and representatives from 100 Swedish companies with more than 100 employees ("major employers"). In this study, we made use of an open definition of AI to accommodate differences in knowledge and opinion around how AI and ADM are understood by the respondents. The study shows a significant difference between direct and indirect AI and ADM use, which has implications for recruiters' awareness of the potential for bias or discrimination in recruitment. All of those surveyed made use of large digital platforms like Facebook and LinkedIn for their recruitment, leading to concerns around transparency and accountability—not least because most respondents did not explicitly consider this to be AI or ADM use. We discuss the implications of direct and indirect use in recruitment in Sweden, primarily in terms of transparency and the allocation of accountability for bias and discrimination during recruitment processes.

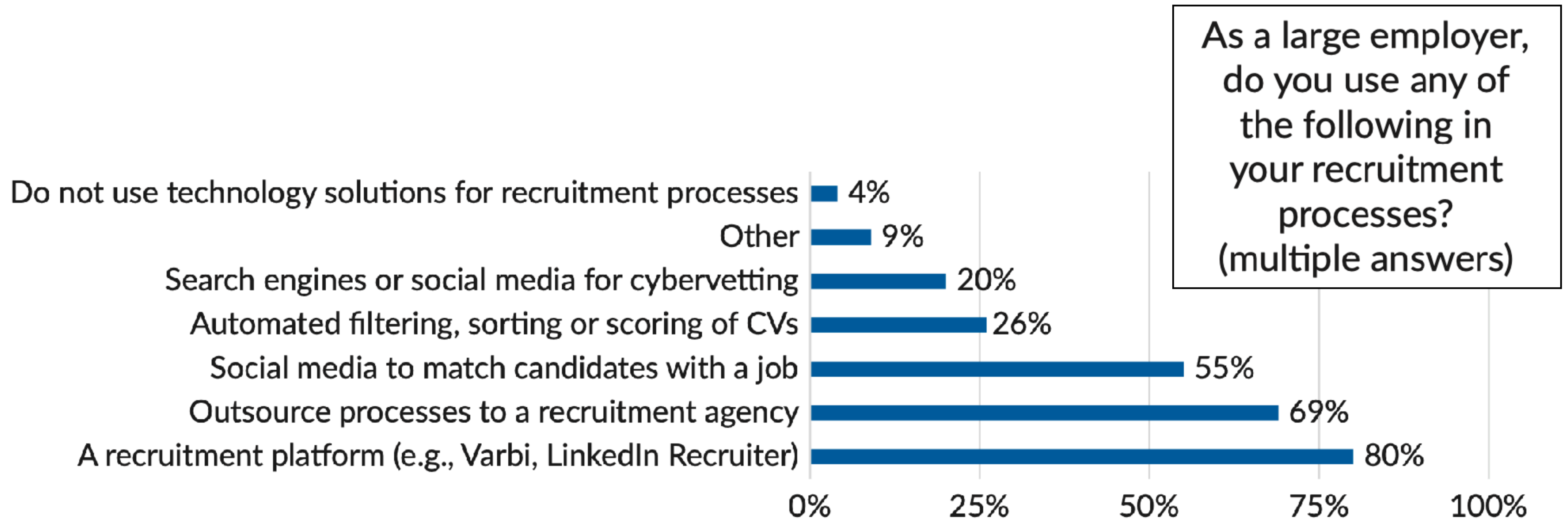
Keywords

ADM and risks of discrimination; AI and accountability; AI and risks of discrimination; AI and transparency; artificial intelligence; automated decision-making; discrimination in recruitment; indirect AI use; platforms and discrimination

Recruitment practices: ADM

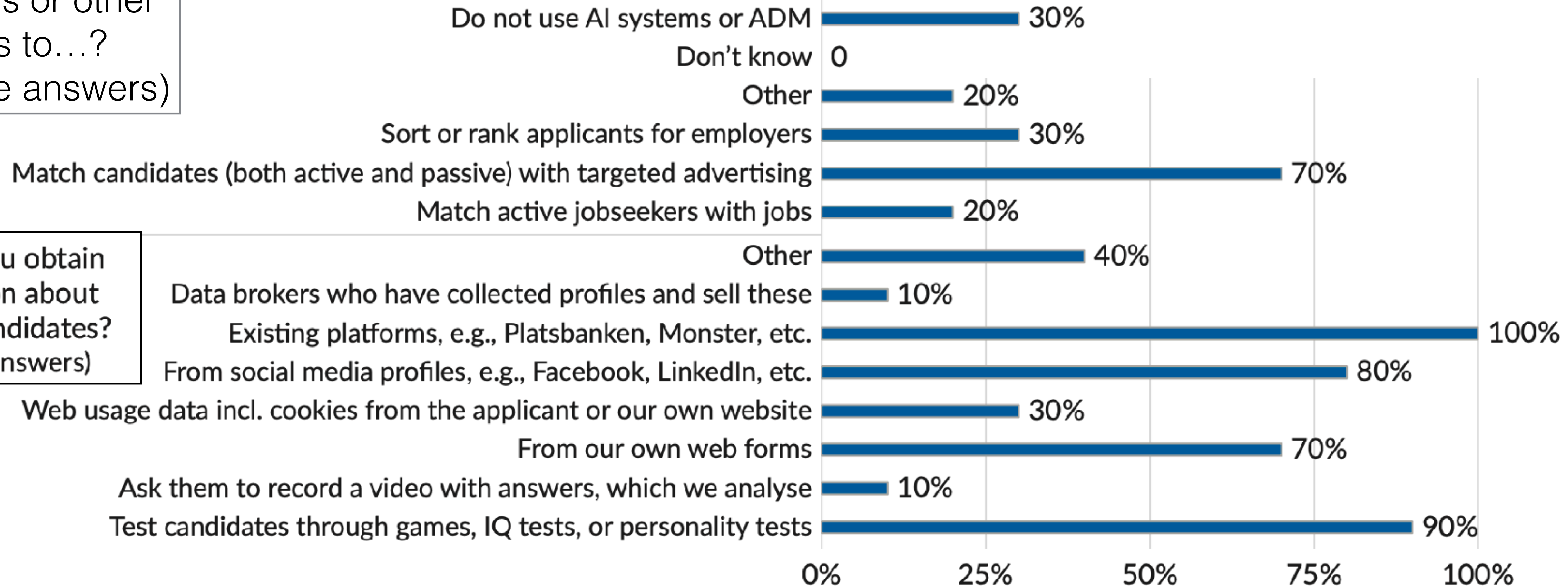


Which tools do major employers use in their recruitment processes?



Recruitment agencies

Do you use AI systems or other tools to...?
(multiple answers)



How do you obtain information about suitable candidates?
(multiple answers)

PART IV: A wider ecology

AI/ADM in recruitment: A wider ecology and low awareness

1. **AI-models can be riddled with unfair bias:** built on social structures as a socio-technical set of artefacts (cf. Larsson et al., 2023); **Proprietary platforms** are non-transparent; **Dominant firms** become infrastructure

Results

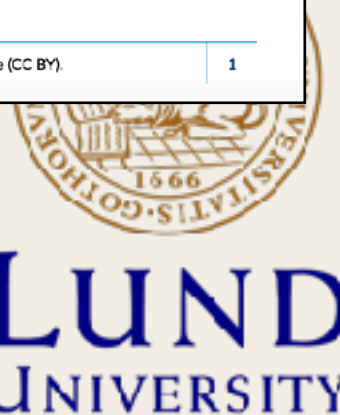
2. **Low awareness:** Direct question indicates *low* AI/ADM use; answers about tools and practices indicate *high use* in automation in decision support.

3. **AI is perceived** as “in the near future”, the platforms’ AI tools are not perceived as AI.

4. AI/ADM in recruitment is used in an “ecology” of subcontractors or external platform services. **Implicates accountability** and perceptions of what “we” do.

Reflection

5. **The AI Act** will ask for more; the Supervisory authorities should do more.





Thank you!

Stefan Larsson

Associate Professor in Technology and
Social Change,

Lawyer (LLM)
PhD in Sociology of Law
PhD in Spatial Planning

Department of Technology and Society,
Faculty of Engineering, Lund University,
Sweden.

Doctoral Seminar on "Algorithms as Future Decision-
Makers: Cross-disciplinary Dialogues in the Shadow of
the AI Act", University of Tuscia, Viterbo, 13 June 2024



LUND
UNIVERSITY