Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action

Ulnicane, Inga

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Social Inclusion
2024 • Volume 12 • Article 7543
https://doi.org/10.17645/si.v12.7543

cogitatio

ARTICLE

Open Access Journal

Intersectionality in Artificial Intelligence: Framing Concerns and Recommendations for Action

Inga Ulnicane

University of Birmingham, UK

Correspondence: Inga Ulnicane (i.ulnicane@bham.ac.uk)

Submitted: 18 August 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

While artificial intelligence (AI) is often presented as a neutral tool, growing evidence suggests that it exacerbates gender, racial, and other biases leading to discrimination and marginalization. This study analyzes the emerging agenda on intersectionality in AI. It examines four high-profile reports dedicated to this topic to interrogate how they frame problems and outline recommendations to address inequalities. These four reports play an important role in putting problematic intersectionality issues on the political agenda of AI, which is typically dominated by questions about AI's potential social and economic benefits. The documents highlight the systemic nature of problems that operate like a negative feedback loop or vicious cycle with the diversity crisis in the AI workforce leading to the development of biased AI tools when a largely homogenous group of white male developers and tech founders build their own biases into AI systems. Typical examples include gender and racial biases embedded into voice assistants, humanoid robots, and hiring tools. The reports frame the diversity situation in AI as alarming, highlight that previous diversity initiatives have not worked, emphasize urgency, and call for a holistic approach that focuses not just on numbers but rather on culture, power, and opportunities to exert influence. While dedicated reports on intersectionality in AI provide a lot of depth, detail, and nuance on the topic, in the patriarchal system they are in danger of being pigeonholed as issues of relevance mainly for women and minorities rather than part of the core agenda.

Keywords

artificial intelligence; data; diversity; feminism; framing; gender; intersectionality; politics; power; technology

1. Introduction

Amid the hype and excitement surrounding the potential economic and social benefits of artificial intelligence (AI), uncomfortable questions regarding the wide-ranging problematic impacts of AI are multiplying (Radu, 2021; Schiff, 2023; Taeihagh, 2021; Ulnicane et al., 2021). While AI is often perceived as neutral and objective, growing evidence suggests that it reinforces and exacerbates human biases and stereotypes leading to disadvantages and discrimination based on gender, race, ethnicity, age, and other characteristics (Allhutter et al., 2020; Benjamin, 2019; Broussard, 2023; Browne et al., 2023; D'Ignazio & Klein, 2020; Noble, 2018; Søraa, 2023; Ulnicane & Aden, 2023). Numerous problematic examples include the hiring tool that discriminates against women, racial discrimination being built into sentencing algorithms, targeted ads, chatbots, voice assistants, and robots replicating gender and racial biases, or facial recognition working poorly on black female faces (Angwin et al., 2016; Buolamwini & Gebru, 2018; Collett & Dillon, 2019; Young et al., 2021).

These examples are closely related to the diversity crisis in AI, where founders as well as employees of major companies largely come from a homogenous group of white men from well-off socio-economic backgrounds and build AI systems according to their views and stereotypes (Little & Winch, 2021; S. M. West et al., 2019; Young et al., 2023). Moreover, this is happening at a time when illiberal pushback against gender equality in technology is encountered at the highest levels of political decision-making as well as through widely publicized initiatives from individuals at big tech companies (Schopmans & Cupac, 2021; S. M. West et al., 2019). Concerns have been raised that discriminatory AI systems might offset earlier advances made towards gender equality (UNESCO, 2020). These examples and concerns highlight problematic political and social impacts of AI, its reinforcement of existing power relationships, structural inequalities, and the highly unequal distribution of rewards and costs of new technologies across various social groups.

Against this background, this article examines an emerging agenda to address inequality issues in AI. To do that, it uses the concept of intersectionality, which deals with multiple inequalities arising from the interaction of various social identities, such as gender, race, ethnicity, and socio-economic background, contributing to the marginalization and disadvantaging of less powerful groups (Crenshaw, 1991; Verloo, 2006). In particular, this study asks how the problems of intersectionality in AI are framed and what recommendations have been outlined to address them. To get a good overview of the problems and solutions, four dedicated reports on intersectionality in AI are analyzed: Collett and Dillon (2019), UNESCO (2020), S. M. West et al. (2019), and Young et al. (2021). This article on intersectionality and AI contributes to other studies on gender and bias in AI policy (Guevara-Gómez et al., 2021; Rönnblom et al., 2023; Ulnicane & Aden, 2023), unpacking how inequalities and discrimination are understood and tackled in documents prepared by various experts and stakeholders aiming to influence and set AI policy.

This article proceeds as follows. First, it outlines the conceptual and methodological framework introducing the key concepts of AI and intersectionality (Section 2.1), as well as explaining the choice of empirical material (Section 2.2). Second, it examines the framing of concerns and recommendations in the four selected reports (Section 3). Finally, the findings from the analysis are discussed (Section 4).

2. Conceptual and Methodological Framework

2.1. Key Concepts: AI and Intersectionality

AI is a highly contested and political concept (Whittaker, 2021). This study follows actors' definitions of AI, considering how the authors of the documents understand and present AI through the definitions and examples they use. According to Collett and Dillon (2019, p. 7), AI refers to "a heterogeneous network of technologies—including machine learning, natural language processing, expert systems, deep learning, computer vision, robotics—which share in common the automation of functions of the human brain." Similarly, Young et al. (2021) use a definition that refers to AI when a machine or system performs tasks that would ordinarily require human or other biological brainpower to be accomplished. The UNESCO (2020, p. 4) report explains:

> Simply put, artificial intelligence (AI) involves using computers to classify, analyze, and draw predictions from data sets using a set of rules called algorithms. AI algorithms are trained using large datasets so that they can identify patterns, make predictions, recommend actions, and figure out what to do in unfamiliar situations, learning from new data and thus improving over time.

The reports mention numerous examples of AI applications. For example, the UNESCO (2020, p. 4) report says that "we interact with AI on a daily basis in our professional and personal lives, in areas such as job recruitment and being approved for a bank loan, in medical diagnoses, and much more." Through various examples of AI applications reinforcing sexist and racist stereotypes—sentencing algorithms, hiring tools, voice assistants, humanoid robots, chatbots, linguistic biases—a more contested and problematic nature of AI is revealed. These examples highlight that AI is not just a neutral tool but is co-created with society, and as such has major political and social implications in reinforcing existing power relationships, discrimination, and structural inequalities (Benjamin, 2019; Noble, 2018; Ulnicane & Aden, 2023).

To examine how AI impacts diverse social groups, intersectionality with its focus on multiple and overlapping identities provides a useful perspective. While gender has long dominated (and still often dominates) efforts toward equality, intersectionality helps to emphasize and remind us of the importance of addressing multiple inequalities and their reinforcement (Crenshaw, 1991; Fothergill et al., 2019; Verloo, 2006). These multiple identities relate to social categories like gender, race/ethnicity, sexual orientation, and class. Such categories are far from straightforward. For example, the traditional understanding of gender as binary (male/female) is being challenged today (D'Ignazio & Klein, 2020). Examining the convergence of various social identity categories is crucial for understanding processes of oppression, dominance, and power. Regarding the socially constructed categories of identity, Crenshaw (1991, pp. 1296–1297), in her groundbreaking work on intersectionality, highlights that:

> A large and continuing project for subordinated people…is thinking about the way power has clustered around certain categories and is exercised against others. This project attempts to unveil the processes of subordination and the various ways those processes are experienced by people who are subordinated and people who are privileged by them. It is, then, a project that presumes that categories have meaning and consequences. And this project's most pressing problem, in many if not most cases, is not the existence of the categories, but rather the particular values attached to them and the way those values foster and create social hierarchies.

Crenshaw's focus on intersecting identity categories and their relation to power and oppression is highly relevant for AI, where similar patterns of subordination and marginalization have been observed (Allhutter et al., 2020; Benjamin, 2019; Broussard, 2023; Browne et al., 2023; D'Ignazio & Klein, 2020; Little & Winch, 2021; Noble, 2018; Søraa, 2023; Young et al., 2023). Feminist, intersectional, and decolonial approaches to AI present resistance to discriminatory AI, draw attention to the wealth of experience that has been erased from the history of AI (women, trans people, people of color, and the disability community), and call for new forms of solidarity across diverse experiences and identities (Bentley et al., 2023; Ciston, 2019; Png, 2022; Toupin, 2023; S. M. West, 2020).

2.2. Methods and Data

To examine the emerging agenda on intersectionality in AI, this study analyzes four high-profile reports dedicated to intersectionality and AI (Table 1) published between 2019 and 2021. All selected reports explicitly deal with AI rather than digital technology more broadly (e.g., M. West et al., 2019). While some of the selected documents primarily focus on gender, they all explicitly recognize and discuss the importance of an intersectional approach. All four reports, ranging from 33–60 pages, provide an in-depth analysis of the topic. By undertaking a close analysis of a small number of dedicated in-depth reports, this article complements the studies that have interrogated gender and bias issues in AI policy documents at the national and international levels (Guevara-Gómez et al., 2021; Rönnblom et al., 2023; Ulnicane & Aden, 2023).

Three reports are published by well-known AI research centers in the UK and the US: the AI Now Institute, the Leverhulme Centre for the Future of Intelligence, and the Alan Turing Institute. The remaining document (UNESCO, 2020) has been prepared by an international organization. While all four are well-known organizations, they operate quite differently.

The AI Now Institute was founded in 2017 at New York University by internal tech industry critics and was initially funded by tech companies (Sadowski & Phan, 2022). It has regularly prepared influential critical reports on AI, which have been used by government agencies, experts, professional associations, and

Table 1. Overview of the reports analyzed.

Authors	Title	Organization	Date of publication	Number of pages
S. M. West, M. Whittaker, and K. Crawford	*Discriminating Systems: Gender, Race, and Power in AI*	AI Now Institute (US) https://ainowinstitute.org	April 2019	33
C. Collet and D. Dillon	*AI and Gender: Four Proposals for Future Research*	Leverhulme Centre for the Future of Intelligence (UK) http://lcfi.ac.uk	June 2019	43
UNESCO	*Artificial Intelligence and Gender Equality: Key Findings of UNESCO's Global Dialogue*	UNESCO https://www.unesco.org/en	August 2020	49
E. Young, J. Wajcman, and L. Sprejer	*Where Are the Women? Mapping the Gender Job Gap in AI*	The Alan Turing Institute (UK) https://www.turing.ac.uk	2021	60

international organizations (Ulnicane & Aden, 2023). As of 2022, AI Now Institute is an independent organization and currently does not take funding from corporate donors. The report analyzed (S. M. West et al., 2019) was supported by Pivotal Ventures, founded by Melinda Gates. The Leverhulme Centre for the Future of Intelligence at the University of Cambridge is funded by the UK research funding charity the Leverhulme Trust. It is an interdisciplinary AI research center working on a wide variety of research programs on topics such as AI governance, trust, and narratives. The report studied (Collett & Dillon, 2019) was sponsored by the Ada Lovelace Institute—an independent research body—and supported by the consulting company PwC. The Alan Turing Institute is the UK's national institute for data science and AI, launched by leading UK universities and a research council. The report examined (Young et al., 2021) has been prepared by the Women in Data Science and AI project of the Public Policy Program of the institute. The final report (UNESCO, 2020) comes from the United Nations Educational, Scientific and Cultural Organization, based in Paris. It was prepared by the UNESCO Division for Gender Equality and funded by the German Government.

While the organizational and funding contexts from which the four reports originate are diverse, all reports present a critical analysis of gender and diversity in AI. Importantly, all reports are authored by women at various career stages (an exception is the UNESCO report that does not name individual authors), including leading experts in AI like feminist scholar Judy Wajcman, or researchers of social and political aspects of AI, such as Meredith Whittaker and Kate Crawford. From an intersectional perspective, it is important to mention that all four well-known and influential reports originate from a small number of elite institutions in the Global North.

To analyze how these reports define the problems of intersectionality in AI and the ways to address them, this study uses a framing approach (Bacchi, 2000; Rein & Schon, 1996; van Hulst & Yanow, 2016). Framing offers a politically nuanced and power-sensitive way to analyze narratives about the problems and possible ways to act upon them. Its focus is on sense-making, selecting, naming, categorizing, and storytelling in complex situations. This approach has been used to analyze policies relevant to this study such as gender equality (Verloo, 2005) and AI (Ulnicane, 2022). Thus, framing is a highly appropriate approach for this examination of how the selected documents make sense and tell stories about the issues of intersectionality in AI and formulate recommendations for tackling them.

3. Analysis of the Reports on Intersectionality and AI

This section examines the selected documents by focusing on how they frame concerns (Section 3.1), what methods and data they use (Section 3.2), how they approach the various social categories they analyze (Section 3.3), and what recommendations they suggest (3.4). For an overview see Table 2.

3.1. Concerns

The reports raise several interconnected concerns, such as the lack of diversity in the AI sector, reproduction and reinforcement of stereotypes and bias in AI systems, the disproportionate effect of job replacement for women, as well as broader issues of social and economic justice and concentration of power. While previously the diversity problem of the AI industry and issues of bias in AI systems have been examined separately, these reports show that "issues of discrimination in the workforce and in system building are deeply intertwined" (S. M. West et al., 2019, p. 6).

Table 2. Intersectionality in the reports analyzed.

Report	Concerns	Methods and data	Social categories	Examples	Recommendations
S. M. West et al. (2019)	Diversity crisis in the AI sector intertwined with bias and discrimination in AI systems Pushback against diversity	Literature review Data on diversity, pay, discrimination, and harassment	Gender (non-binary, fluid), race, and other identities in the context of existing power structures	Sentencing algorithms Amazon hiring tool Targeted ads Facial recognition	Improving workplace diversity and addressing bias and discrimination in AI systems Worker-led initiatives
Collett and Dillon (2019)	Biased datasets Lack of diversity in the AI workforce	Workshop Literature review	Gender (non-binary) in relation to power dynamics, and functioning intersectionally with race, ethnicity, and sexuality	Humanoid robotics Virtual personal assistants Facial recognition Crime and policing technologies Linguistic biases Health technologies	Four research proposals: bridging gender theory and AI practice; law and policy; biased datasets; diversity in the AI workforce
UNESCO (2020)	AI reinforces gender stereotypes Women at higher risk of being replaced by automation	Dialogue with experts Additional research (including on AI principles)	Gender (non-binary) Intersectionality	Digital voice assistants Amazon hiring tool Targeted ads	Gender equality in AI principles Increasing awareness, education, and skills Industry action Coalition building
Young et al. (2021)	The absence of women in AI and data science leads to gender bias being built into machine learning systems Social and economic justice	Dataset on AI and data science workforce Review of existing datasets Case study of AI platforms Literature review	Gender Binary gender data and lack of intersectional data as a problem	Amazon hiring tool Racist and sexist chatbot Targeted ads Facial recognition Linguistic biases Voice assistants	Reporting standards on gender and intersectional data Government programs Targets and quotas for recruiting and promoting women

Lack of diversity in the AI workforce is seen as the key problem in all four reports. It is called the "diversity crisis" and "diversity disaster" (S. M. West et al., 2019) and illustrated with data on the lack of women, black people, and minorities in the AI sector. Women comprise 15% of AI research staff at Facebook and just 10% at Google, while only 2.5% of the workforce at Google and 4% of the workforce at Facebook and Microsoft is black (S. M. West et al., 2019, p. 5). Furthermore, Facebook reported just 5% of Hispanic workers, and only 3.6% of Google's full-time workers are Latinx (S. M. West et al., 2019, p. 11). Merely 18% of authors at leading AI conferences are women and more than 80% of AI professors are male (S. M. West et al., 2019). Only 7% of students studying computer science and 17% of those working in technology in the UK are women (Collett & Dillon, 2019, p. 25). Worldwide, 78% of AI and data science professionals are male (Young et al., 2021, p. 2). In the broader field of computer science, women make up 24.4% of the workforce and receive median salaries that are only 66% of the salaries of their male counterparts (S. M. West et al., 2019, p. 11). When discussing the diversity crisis, the reports mention issues of harassment, discrimination, unfair compensation, microaggressions, unwelcoming environments, stereotypes, a culture of inequity, and a lack of promotion for women at tech companies (S. M. West et al., 2019; Young et al., 2021).

Young et al. (2021) find diverging career trajectories, where women are more likely to occupy jobs associated with less status and pay, usually within analytics, data preparation, and exploration, rather than the more prestigious jobs in engineering and machine learning. Moreover, there are fewer women in industries that traditionally entail more technical skills and, thus, are seen as "masculine" like information technology, while more women are in industries that involve fewer technical skills and are perceived as "feminine," such as healthcare. This pattern reflects the historical association of technology with men, while perceiving femininity as incompatible with technology (Wajcman, 2010). Young et al. (2021) find that there are also fewer women in leadership positions in AI, even though they are better qualified, i.e., have higher education levels than men. Still, women self-report fewer skills on LinkedIn than men and are less active on online data science platforms. Furthermore, women working in AI and data science have higher turnover and are more likely to leave the sector than men.

Considering that these issues have been known and investments have been made to address them for decades, the current situation is described as "alarming" (S. M. West et al., 2019). Despite diversity initiatives, the broader field of computer science has experienced a sharp decline of women in its ranks from 37% of computer science majors in the US in 1984 to 18% in 2015 (S. M. West et al., 2019, p. 11). This is reflected in broader concerns discussed in the literature that diversity initiatives have changed little in the AI sector, are still often poorly understood among tech workers, and are sometimes even seen as a threat to scientific excellence (Browne et al., 2024; Stinson & Vlaad, 2024). Furthermore, experts also highlight the pushback to diversity by those who question or even reject that racism, misogyny, and harassment are problems in the AI field (Collett & Dillon, 2019; S. M. West et al., 2019). S. M. West et al. (2019) are critical of so-called "pipeline studies" that focus on the absence of diverse candidates in the hiring pool, which is often used by the industry to justify the lack of diversity and place the problem outside their remit. They refer to the well-known work of historian Mar Hicks who has demonstrated structural discrimination in the computing sector that led to the exclusion and marginalization of women who initially dominated computer operation and programming (Hicks, 2018).

Importantly, the reports emphasize that these are systemic issues reflecting existing power relationships in the AI sector, which are shaped by a feedback loop of discriminatory workplace practices leading to discriminatory tools (S. M. West et al., 2019). The major gender disparity in the AI workforce hinders the development of equitable AI (Collett & Dillon, 2019). "A troubling and persistent absence of women" (Young et al., 2021, p. 2) is seen as having wider consequences when it results in gender bias being built into machine learning systems. It is also seen as fundamentally an ethical issue of social and economic justice. The reports mention numerous cases where AI applications reinforce gender and racial stereotypes and discrimination, including targeted ads shown according to gender and racial stereotypes, or Amazon hiring tool that downgraded women applicants because it was built using historical—predominately male—employment data (S. M. West et al., 2019). S. M. West et al. (2019, p.5) highlight the systemic nature of the problems:

> The diversity problem is not just about women. It is about gender, race, and most fundamentally about power. It affects how AI companies work, what products get built, who they are designed to serve, and who benefits from their development.

Experts stress that AI tools spreading and reinforcing gender stereotypes would further stigmatize and marginalize women in economic, political, and social life as well as offset an earlier progress made toward gender equality (UNESCO, 2020). They state that the design and implementation of AI perpetuates a vicious cycle, where technology captures and amplifies controlling and restrictive conceptions of gender and race, such as gender binary and racial hierarchies, which are then repetitively reinforced (Collett & Dillon, 2019). This can be seen, for example, in digital voice assistants. Almost all voice assistants, such as Amazon's Alexa or Apple's Siri, "are given female names and voices, and express a 'personality' that is engineered to be uniformly subservient" (UNESCO, 2020, p. 2).

Not only women are largely underrepresented in the AI sector, but they are also at a higher risk of being displaced in the workforce because they are disproportionately represented in sectors that are undergoing automation, like clerical, administrative, bookkeeping, and cashier jobs (UNESCO, 2020).

The reports point out the limitations of emerging AI policies and guidelines. Collett and Dillon (2019, p. 13) emphasize the risk "that economic prosperity and political power will play an underlying role in shaping laws and policies on AI at the expense of more socially equalizing motivations." Analysis of gender equality issues in AI principles reveals that "direct references to gender equality and women's empowerment in existing AI and ethics principles are scarce" (UNESCO, 2020, p. 9). Nevertheless, the report finds implicit alignment with gender equality in discussions of issues such as justice and solidarity.

To sum up, the reports diagnose the diversity situation in the AI field as alarming, troubling, and being in crisis. Problems are seen as systemic, leading to a negative feedback loop and a vicious cycle where discrimination in the workforce leads to the building of discriminating tools. Importantly, experts highlight that previous diversity initiatives have not led to positive changes but rather to a decline in diversity. Despite the acceptance of the diversity rhetoric by tech companies, it is often poorly understood and has experienced some pushback. Reports highlight that AI reinforces gender and racial stereotypes and can offset earlier progress made toward equality.

3.2. Methods and Data

To identify problems and come up with recommendations, the reports draw on a range of research methods and data sources, including a literature review, data on diversity, pay, discrimination, and harassment, workshop and dialogue with experts, and analysis of AI principles. Reports build on feminist and occasionally also intersectional scholarship and review existing literature in this area, including many well-known studies (e.g., Angwin et al., 2016; Benjamin, 2019; Broussard, 2018; Buolamwini & Gebru, 2018; Caliskan et al., 2017; Criado Perez, 2019; D'Ignazio & Klein, 2020; Eubanks, 2019; Hicks, 2018; Noble, 2018; O'Neil, 2016; Zou & Schiebinger, 2018). A considerable part of this literature is authored by women, including black women. Furthermore, documents refer to each other, with the UNESCO (2020) report citing earlier documents on AI, gender, and race (Collett & Dillon, 2019; S. M. West et al., 2019) and Young et al. (2021) reviewing the reports by UNESCO (2020) and S. M. West et al. (2019). They also review other related reports and guidelines; for example, the UNESCO (2020) document draws on UNESCO's work on digital skills and gender divide (M. West et al., 2019) and AI principles.

Additionally, the reports build on an expert workshop and consultation. One of the documents (Collett & Dillon, 2019) results from the trans-disciplinary and trans-sectoral Gender and AI workshop held at the University of Cambridge. The workshop identified four proposals for future research on bridging gender theory and AI practice, law and policy, biased datasets, and diversity in the AI workforce. It is primarily situated in the UK context but argues that research should be as international as possible (Collett & Dillon, 2019). The workshop had 47 attendees, including well-known experts in the field of AI from industry, government, non-governmental sector, and leading universities like Cambridge, Oxford, and the London School of Economics. Many of them come from the UK but some also from other countries like the US and Ireland.

For the UNESCO report, with a particular focus on the private sector, 17 experts in AI, digital technology, and gender equality from academia, civil society, and the private sector were consulted (UNESCO, 2020). Many of these experts come from the UK and the US. Two out of 17 experts have also attended the workshop that led to the report by Collett and Dillon (2019). The purpose of the dialogue was to identify issues, challenges, and good practices to help overcome the built-in gender biases in AI devices, data sets and algorithms, improve the global representation of women in technical roles and boardrooms in the technology sector, and create robust and gender-inclusive AI principles, guidelines, and codes of ethics within the industry.

Overall, for these two reports (Collett & Dillon, 2019; UNESCO, 2020) a relatively small number of well-known experts have been involved, with many of them coming from only two countries—the UK and the US. From an intersectional perspective, this raises important questions about who has an opportunity to be part of identifying problems and developing recommendations and who is left out.

To study the gender job gap in AI, Young et al. (2021) reviewed existing statistics and datasets in this area. They conclude that existing data is sparse, fragmented, incomplete, and inadequate to analyze the careers of women and men in AI. Shortcomings of available data include binary gender data and the lack of intersectional data on age, race, geography, disability, sexual orientation, and socioeconomic status. The limited data availability hinders opportunities for intersectional analysis (see Section 3.3). In partnership with an executive search and consulting firm specializing in data science, advanced analytics, and AI, they develop a novel data science and AI career dataset. It consists of 19,535 profiles out of which 11.3% are women, belonging mostly to the US, France, Germany, and the UK (Young et al., 2021, 2023). Their sample is very senior with an average of almost 20 years of work experience and with over 55% holding a graduate or postgraduate degree. The authors admit that their sample is not comprehensive, intersectional, and representative of the entire global data and AI population. Thus, although a novel dataset contributes to better information about the little-studied gender issues in AI and data science, it still is limited to senior professionals in a small number of most developed countries.

To summarize, while all four reports use methods and data typical for preparing such documents, like literature review, workshop, dialogue, and a new data set, they have some limitations, especially concerning intersectional analysis. The sample in the dataset and participants in a workshop and dialogue come predominantly from elite ranks in a tiny number of Global North countries (US, UK, France, and Germany). In a certain way, this selection reproduces existing power imbalances, where those based in the periphery or Global South, and doing less prestigious work in AI and data science, like data labeling or content moderation, remain voiceless and marginalized. The reports admit that their analysis is not comprehensive. The UNESCO (2020, p. 9) report states upfront that "this is not a comprehensive exploration of the complexities of the AI ecosystem in all its manifestations and all its intersections with gender equality. Rather, this is a starting point for conversation and action."

3.3. Social Categories

While one of the reports already in its title explicitly refers to gender and race (S. M. West et al., 2019), the other three documents (Collett & Dillon, 2019; UNESCO, 2020; Young et al., 2021) primarily focus on gender but recognize the importance of intersectional approach. Gender is typically understood as non-binary and like other social categories, it is analyzed in the context of existing power structures. Collett and Dillon (2019), who mainly discuss gender, recognize that gender is inseparable from race, ethnicity, and sexuality. In their report, "gender is understood to have an inextricable relationship with unequal power dynamics, and to function intersectionally with other protected characteristics such as race, ethnicity, and sexuality" (Collett & Dillon, 2019, p. 7). Their focus on equality includes trans, queer, and non-binary equality. Similarly, the UNESCO report, which predominantly explores gender, recognizes the importance of intersectionality:

> Women are a multifaceted and heterogenous group and have different experiences based on realities or characteristics which include: women living in rural and remote areas; indigenous women; racial, ethnic or religious minority women; women living with disabilities; women living with HIV/AIDS; women with diverse sexual orientations and gender identities; younger or older women; migrant, refugee or internally displaced women or women in humanitarian settings. (UNESCO, 2020, pp. 7–8)

Documents discuss a well-known problem (see D'Ignazio & Klein, 2020), namely, the limited availability of intersectional data, such as the lack of data on trans persons and other gender minorities. They find that "the literature almost solely looked at gender, and represented gender as binary. It much less frequently examined race, or other identities, and even more rarely examined the intersection of such identities" (S. M. West et al., 2019, p. 20). S. M. West et al. (2019) emphasize that, instead of the overwhelming and too narrow focus on "women in tech" likely to privilege white women, it is important to acknowledge intersections of gender, race, and other identities as well as existing power structures.

3.4. Recommendations

In response to the concerns discussed, the reports outline several recommendations for tackling the diversity problem in the AI sector as well as improving AI policy and research in this area. Interestingly, one of the reports highlights that its proposals are not intended to be prescriptive but rather should be seen as a provocative mechanism to raise awareness, summarize the current challenges, and prompt practical action (Collett & Dillon, 2019).

Importantly, the UNESCO report emphasizes that to address diversity issues an adequate framing of the overarching AI landscape is needed. It suggests to "shift the narrative of AI as something 'external' or technologically deterministic, to something 'human' that is not happening to us but is created, directed, controlled by human beings and reflective of society" (UNESCO, 2020, p. 33).

The reports emphasize a critical moment and urgency to address the diversity crisis, where "the AI sector needs a profound shift in how it addresses the current diversity crisis. The AI industry needs to acknowledge the gravity of its diversity problem, and admit that existing methods have failed" (S. M. West et al., 2019, p. 3). Recommendations for increasing the diversity of the AI workforce emphasize the need to go beyond just hiring more women and minorities. Collett and Dillon (2019) suggest looking not only at balancing the numbers of the AI workforce but also at creating a culture of diversity in educational institutions and the workplace. When discussing diverse development teams, the UNESCO report argues for a broad approach emphasizing that "this is not a matter of numbers, but also a matter of culture and power, with women actually having the ability to exert influence" (UNESCO, 2020, p. 23). Additionally, it calls for a robust approach to raise awareness and literacy, technical and ethical education, skills development, and capacity building.

Similarly, S. M. West et al. (2019) underline the systemic nature of problems of diversity and bias in AI, which cannot just be addressed by technical approaches but rather require an integration of social and technical approaches. Recommendations to address diversity problems include ending pay and opportunity inequality, developing and implementing reporting standards on gender and other characteristics in AI and data science companies, publishing transparency reports on harassment and discrimination, and broadening recruitment beyond elite universities. Young et al. (2021) call for intersectional gender mainstreaming in human resources policy so that women and men are given equal access to well-paid jobs and careers as well as for actionable targets, incentives, and quotas for recruiting and promoting women. S. M. West et al. (2019) argue that a broader approach is needed that considers existing racism and misogyny and focuses on changing workplace cultures. They highlight the role of worker-led initiatives to address the issues of harassment, discrimination, diversity, and equal pay.

To tackle the development of biased and discriminatory AI tools, experts call for going beyond technical debiasing to include a wider social analysis (S. M. West et al., 2019). Reports stress the importance of diverse perspectives and participation of a wide range of stakeholders from academia, civil society, government, and the private sectors. Such multi-stakeholder forums would include users as well as diverse disciplines from computing to social sciences and humanities. The UNESCO report suggests "establish[ing] a multi-disciplinary and inter-generational coalition that builds partnerships across sectors and groups for a holistic society/AI ecosystem approach. Create avenues for dialogue and learning by creating a common understanding and language and through collaborations and collective impact models" (UNESCO, 2020, p. 34). However, Young et al. (2021) emphasize that the AI industry should avoid "participation washing," when the mere fact that somebody has participated in a project (rather than having had a chance to meaningfully shape it) is supposed to lend it legitimacy.

The UNESCO (2020) report mentions that AI has the potential to contribute to gender equality. To illustrate this point, the report provides an example that AI can help employers use gender-sensitive language to write inclusive job postings to increase the diversity of their workforce. In the reports studied, this is a rare example of the positive impact of AI on equality.

Importantly, the reports also suggest exploring where AI should not be deployed (Collett & Dillon, 2019), if it results in inequality and discrimination. Experts recommend undertaking risk assessments of whether certain systems should be designed at all (S. M. West et al., 2019). Tools that claim to detect sexuality from headshots, predict criminality based on facial features, or assess worker competence via micro-expressions are seen as particularly problematic and in need of urgent reconsideration (S. M. West et al., 2019). To avoid harm, the UNESCO report suggests accepting "that some things may not be able to be fixed and therefore should not be done at all, or should ultimately be abandoned (e.g., the example of Amazon's hiring algorithm which remained biased after multiple attempts to fix it)" (UNESCO, 2020, p. 17).

To address diversity issues in AI, reports emphasize the need for effective and gender-inclusive policies (Young et al., 2021). The UNESCO (2020) report calls for governments to commit to policies, regulations, and mechanisms that proactively promote gender equality in and through AI. Experts suggest that governments together with national and international organizations must initiate research and advocacy programmes to tackle gender gaps. To ensure justice and fairness, they should also take proactive steps to include women and marginalized groups in the design and development of AI (Young et al., 2021). Moreover, the reports focus on guidelines and principles that can support diversity. The UNESCO report suggests strengthening and operationalizing gender equality in AI principles, while Collett and Dillon (2019) recommend context and gender-specific guidelines for data collection and handling, in particular in three contexts that significantly impact gender equality: crime and policing, health, and the financial sector.

Experts call for research in this area to be collaborative, intersectional, pluralistic, interdisciplinary, and trans-sectoral (Collett & Dillon, 2019). It needs to provide a deeper analysis of the power and structural challenges AI systems pose to communities examining the relationship between technology development and the lived experiences of individuals of different racialized, gendered, and classed identities (S. M. West et al., 2019). Collett and Dillon (2019) suggest utilizing gender theory, including trans, non-binary, queer, feminist, as well as postcolonial theory. Considering the crucial role that law and policy play in shaping AI, they recommend analyzing their impact on gender, as well as race, ethnicity, sexuality, disability, etc., so that "the intersectional nature of the research would enable it to consider different standpoints, working for widespread social justice and redistribution of power" (Collett & Dillon, 2019, p. 18). The UNESCO (2020) report highlights the importance of learning from past experiences: successes, gaps, and failures of diversity initiatives.

To summarize, the reports emphasize the urgency and critical moment to acknowledge the gravity of the diversity problem in AI. They call for a holistic and broad approach that goes beyond just increasing the numbers of women in AI and focuses on culture, power, and opportunities to exert influence. Rather than just focusing on technical fixes of bias, a socio-technical approach that involves perspectives from multiple disciplines and sectors is suggested (see Ulnicane & Aden, 2023). Importantly, the reports argue that some AI systems, that result in discrimination and inequality, should not be built at all.

4. Conclusions

While the political agenda on AI tends to be dominated by questions of potential economic and social benefits (Schiff, 2023; Ulnicane, 2022), the four high-profile reports on intersectionality in AI analyzed in this study play an important role in bringing the problematic impacts of AI on gender, race, socio-economic background, and other intersecting identities to broader attention.

This study demonstrates that there is a lot of convergence among the four reports in the way they frame problems, provide evidence through data and examples, and outline recommendations for action (Table 2). Experts demonstrate how the diversity crisis in the AI workforce is closely intertwined with biased AI tools. They emphasize the systemic nature of problems when discrimination in the workforce leads to the development of discriminatory systems. This diagnosis of the problem resonates with feminist scholarship of technology that focuses on the mutual shaping of gender and technology and highlights how processes of technical change can influence gender power relations (Wajcman, 2010). A major concern discussed in the reports is the ineffectiveness of earlier diversity initiatives in tech. The reports highlight the urgency and critical moment to reconsider previous approaches and address problems in a holistic way that focuses not only on increasing the number of women and minorities but also on considering culture, power, and opportunities to exert influence.

Remarkably, almost all examples of the impact of AI on gender, race, and other identities in these reports as well as in other AI documents (Ulnicane & Aden, 2023) highlight problematic consequences, while it is difficult to find examples that would confirm optimistic views that AI could help to eliminate human bias. In the four reports analyzed, the only positive example suggests that AI could help to use gender-sensitive language to write inclusive job postings to help increase the diversity of their workforce (UNESCO, 2020).

These reports, written by some of the leading scholars on feminism and the social implications of AI, are important in highlighting the AI diversity problems. However, from an intersectional perspective, these influential reports originating from a small number of elite institutions in the Global North are somewhat limited because their data and methods mainly focus on senior professionals in a few most developed countries.

Nevertheless, the problems and recommendations outlined in these reports are important initial steps in shaping the agenda and calling for action on intersectionality in AI. These reports, and most of the literature they cite, come from a community of women dedicated to the issues of intersectionality and AI. However, in the patriarchal system, these issues can be at risk of being pigeonholed as issues relevant mostly to women, black people, and minorities rather than part of the core AI agenda (D'Ignazio & Klein, 2020; Ulnicane & Aden, 2023). Thus, one of the questions for future research is to explore how intersectional and feminist approaches can be seen not just as an add-on but rather as a challenge and alternative to the existing system.

This study contributes to the growing intersectionality research across disciplines and sectors, including AI and data science (Bentley et al., 2023). While further work examining and tackling intersectionality issues in AI is urgently needed, it has to be recognized that intersectionality is not a straightforward concept. Rather:

> [It is] a multifaceted area of theory and praxis that is often contradictory. Intersectionality can be an expression of one's identity, which can be singular, multiple, and/or intersectional, rooted and stable, or changing constantly. It can constitute belonging and/or unpack marginalization and disadvantage. It can unite people in their endeavours and/or detonate struggles against systems of oppression, discrimination or persecution. It can be an abstract ideological project and/or rich and detailed experiences of existence. Crucially, it is a significant forum for investigating and transforming relationships between people, places and institutions, towards human rights, reduced inequality and

social justice. Likewise, it can be and do none of those things—serving merely as a buzzword. (Bentley et al., 2023, p. 13)

Acknowledgments

Helpful comments and suggestions from two anonymous reviewers are gratefully acknowledged.

Funding

The research reported in this article has received funding from the EU's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 945539 (HBP SGA3).

Conflict of Interests

The author declares no conflict of interests.

References

Allhutter, D., Cech, F., Fischer, F., Grill, G., & Mager, A. (2020). Algorithmic profiling of job seekers in Austria: How austerity politics are made effective. *Frontiers in Big Data*, *3*, Article 5. https://doi.org/10.3389/fdata.2020.00005

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bacchi, C. (2000). Policy as discourse: What does it mean? Where does it get us? *Discourse: Studies in the Cultural Politics of Education*, *21*(1), 45–57.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.

Bentley, C., Muyoya, C., Vannini, S., Oman, S., & Jimenez, A. (2023). Intersectional approaches to data: The importance of an articulation mindset for intersectional data science. *Big Data & Society*, *10*(2). https://doi.org/10.1177/20539517231203667

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. MIT Press.

Browne, J., Cave, S., Drage, E., & McInerney, K. (Eds.). (2023). *Feminist AI: Critical perspectives on algorithms, data, and intelligent machines*. Oxford University Press.

Browne, J., Drage, E., & McInerney, K. (2024). Tech workers' perspectives on ethical issues in AI development: Foregrounding feminist approaches. *Big Data & Society*, *11*(1). https://doi.org/10.1177/20539517231221780

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, *81*, 77–91.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Ciston, S. (2019). Intersectional AI is essential: Polyvocal, multimodal, experimental methods to save artificial intelligence. *Journal of Science and Technology of the Arts*, *11*(2), 3–8. https://doi.org/10.7559/citarj.v11i2.665

Collett, C., & Dillon, D. (2019). *AI and gender: Four proposals for future research*. The Leverhulme Centre for the Future of Intelligence. https://doi.org/10.17863/CAM.41459

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*(6), 1241–1299.

Criado Perez, C. (2019). *Invisible women: Data bias in a world designed for men*. Random House.
D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.
Eubanks, V. (2019). *Automating inequality: How high-tech tools profile, police, and punish the poor*. Picador.
Fothergill, B. T., Knight, W., Stahl, B. C., & Ulnicane, I. (2019). Intersectional observations of the Human Brain Project's approach to sex and gender. *Journal of Information, Communication and Ethics in Society*, *17*(2), 128–144. https://doi.org/10.1108/JICES-11-2018-0091
Guevara-Gómez, A., de Zárate-Alcarazo, L. O., & Criado, J. I. (2021). Feminist perspectives to artificial intelligence: Comparing the policy frames of the European Union and Spain. *Information Polity*, *26*(2), 173–192.
Hicks, M. (2018). *Programmed inequality: How Britain discharged women technologists and lost its edge in computing*. MIT Press.
Little, B., & Winch, A. (2021). *The new patriarchs of digital capitalism: Celebrity tech founders and networks of power*. Routledge.
Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin Books.
Png, M.-T. (2022). At the tensions of South and North: Critical roles of Global South stakeholders in AI governance. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1434–1445). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533200
Radu, R. (2021). Steering the governance of artificial intelligence: National strategies in perspective. *Policy and Society*, *40*(2), 178–193. https://doi.org/10.1080/14494035.2021.1929728
Rein, M., & Schon, D. (1996). Frame-critical policy analysis and frame-reflective policy practice. *Knowledge and Policy*, *9*, 85–104. https://doi.org/10.1007/BF02832235
Rönnblom, M., Carlsson, V., & Öjehag-Pettersson, A. (2023). Gender equality in Swedish AI policies. What's the problem represented to be? *Review of Policy Research*, *40*(5), 688–704. https://doi.org/10.1111/ropr.12547
Sadowski, N., & Phan, T. (2022). "Open secrets": An interview with Meredith Whittaker. In T. Phan, J. Goldenfein, D. Kuch, & M. Mann (Eds.), *Economies of virtue: The circulation of "ethics" in AI* (pp. 140–152). Institute of Network Cultures.
Schiff, D. (2023). Looking through a policy window with tinted glasses: Setting the agenda for U.S. AI policy. *Review of Policy Research*, *40*(5), 729–756. https://doi.org/10.1111/ropr.12535
Schopmans, H., & Cupac, J. (2021). Engines of patriarchy: Ethical artificial intelligence in times of illiberal backlash politics. *Ethics & International Affairs*, *35*(3), 329–342.
Søraa, R. (2023). *AI for diversity*. CRC Press.
Stinson, C., & Vlaad, S. (2024). A feeling for the algorithm: Diversity, expertise, and artificial intelligence. *Big Data & Society*, *11*(1). https://doi.org/10.1177/20539517231224247
Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and Society*, *40*(2), 137–157. https://doi.org/10.1080/14494035.2021.1928377
Toupin, S. (2023). Shaping feminist artificial intelligence. *New Media & Society*, *26*(1), 580–595. https://doi.org/10.1177/14614448221150776
Ulnicane, I. (2022). Emerging technology for economic competitiveness or societal challenges? Framing purpose in artificial intelligence policy. *Global Public Policy and Governance*, *2*, 326–345. https://doi.org/10.1007/s43508-022-00049-8
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

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2021). Framing governance for a contested emerging technology: Insights from AI policy. *Policy and Society*, *40*(2), 158–177. https://doi.org/10.1080/14494035.2020.1855800

UNESCO. (2020). *Artificial intelligence and gender equality. Key findings of UNESCO's global dialogue*. https://unesdoc.unesco.org/ark:/48223/pf0000374174

van Hulst, M., & Yanow, D. (2016). From policy "frames" to "framing" theorizing a more dynamic, political approach. *The American Review of Public Administration*, *46*(1), 92–112.

Verloo, M. (2005). Mainstreaming gender equality in Europe. A critical frame analysis approach. *The Greek Review of Social Research*, *117*, 11–34. https://doi.org/10.12681/grsr.9555

Verloo, M. (2006). Multiple inequalities, intersectionality and the European Union. *European Journal of Women's Studies*, *13*(3), 211–228.

Wajcman, J. (2010). Feminist theories of technology. *Cambridge Journal of Economics*, *34*(1), 143–152.

West, M., Kraut, R., & Ei Chew, H. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

West, S. M. (2020). Redistribution and rekognition: A feminist critique of algorithmic fairness. *Catalyst: Feminism, Theory, Technoscience*, *6*(2), 1–24.

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating systems: Gender, race, and power in AI*. AI Now Institute. https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2

Whittaker, M. (2021). The steep cost of capture. *Interactions*, *28*(6), 50–55. https://doi.org/10.1145/3488666

Young, E., Wajcman, J., & Sprejer, L. (2021). *Where are the women? Mapping the gender job gap in AI*. The Alan Turing Institute. https://www.turing.ac.uk/news/publications/report-where-are-women-mapping-gender-job-gap-ai

Young, E., Wajcman, J., & Sprejer, L. (2023). Mind the gender gap: Inequalities in the emergent professions of artificial intelligence (AI) and data science. *New Technology, Work and Employment*, *38*(3), 391–414.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—It's time to make it fair. *Nature*, *559*, 324–326 https://doi.org/10.1038/d41586-018-05707-8

About the Author

Inga Ulnicane is a research fellow at the University of Birmingham (UK) and an honorary senior research fellow at De Montfort University (UK). She works at the intersection of political science and social studies of science and technology. She has published extensively on topics such as politics and policy of AI, governance of emerging technologies, responsible research and innovation, and grand challenges. In addition to her academic research, she has prepared commissioned reports for the European Parliament and the European Commission.