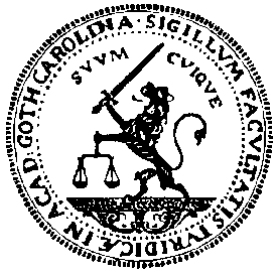


FACULTY OF LAW
Lund University

Elsa Alkan Olsson

Artificial Intelligence and Gender Equality

A Study on Legal Scholars' Understanding of Gender Discrimination and Suggested Solutions

LAGM01 Graduate Thesis

Graduate Thesis, Master of Laws Program
30 Higher Education Credits

Supervisor: Valentin Jeutner

Semester of Graduation: Spring Semester 2022

SUMMARY

The thesis aims to identify the dominant narrative within the legal research field of Artificial Intelligence (AI) and discuss the merit of the prevalent solutions offered by scholars in relation to substantive gender equality. The thesis finds the solutions from a systematic legal literature review covering AI and gender. It develops an analytical framework by applying the theory of social change to assess the recommended solutions' compatibility with the aim of the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW).

The literature review reveals that scholars focus on technological solutions to combat discrimination within AI. The thesis argues that these 'technocentric' solutions suffer from a significant methodological limitation: they define AI exclusively as a technological concept, detached from the social context. The findings suggest that the proposed solutions, similar to formal equality, may play a part in achieving social change. However, they fail to fulfil CEDAW's goal of substantive equality. Ultimately, eradicating discrimination solely through technological solutions is not adequate. Instead, legal scholars must broaden their scope of understanding; AI must not be viewed as an exclusively technological concept but as a system shaped by its social setting. Masquerading these so-called solutions as progress risks doing more harm than good, leaving substantive equality out of reach.

Keywords: Artificial Intelligence, Gender-based discrimination, Human Rights, CEDAW, Substantive Equality

SAMMANFATTNING

Syftet med denna uppsats är att identifiera den rådande diskursen inom juridisk forskning på området artificiell intelligens (AI) och genus, och att diskutera för- och nackdelar med de lösningar forskare presenterat i relation till substantiell jämställdhet. De rekommenderade lösningarna identifieras genom en systematisk juridisk litteraturgenomgång av AI och genus. Genom att tillämpa teorin om *social förändring* utvecklar uppsatsen ett analytiskt ramverk för att bedöma lösningarnas kompatibilitet med kvinnokonventionens (CEDAW) syfte.

Litteraturgenomgången visar att forskarna fokuserar på teknologiska lösningar i bekämpningen av könsdiskriminering inom AI. Dessa teknologiskt fokuserade lösningar lider av en metodologisk begränsning. AI definieras uteslutande som ett teknologiskt koncept, fristående från den sociala kontexten. I uppsatsen fastslås att de föreslagna lösningarna, likt formell jämställdhet, till viss del kan bidra till social förändring. Dock misslyckas de med att uppnå CEDAW:s långsiktiga mål om substantiell jämställdhet. Därmed är teknologiska lösningar otillräckliga medel i förhindrandet av diskriminering. Forskare måste därför vidga sin förståelse och inse att AI inte uteslutande kan förstås som ett teknologiskt koncept men som ett system format av sitt sociala sammanhang. Att maskera dessa lösningar som framgång riskerar att göra mer skada än nytta för substantiell jämställdhet.

Nyckelord: artificiell intelligens, genus, diskriminering, mänskliga rättigheter, CEDAW, substantiell jämställdhet

PREFACE

This thesis marks the end of my five years at Lund University. As such, I would like to take this opportunity to express my appreciation for the people who made this thesis possible.

On a personal note, the vast source of inspiration, motivation, and support throughout the years from my parents has guided me towards the human rights lawyer I aspire to be. Finishing my law degree would not have been possible without you.

I would also like to thank Donna, who has provided me with endless support and proved to be the best proofreader a friend could wish for.

Last but not least, I would like to give a special thanks to my supervisor Valentin Jeutner for providing me with helpful feedback on the thesis.

Stockholm, 20 May 2022

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CEDAW	Convention on the Elimination of All Forms of Discrimination against Women
ECHR	European Convention on Human Rights
EU	European Union
ICCPR	International Covenant on Civil and Political Rights
ICESCR	Covenant on Economic, Social and Cultural Rights
NHRS	National Human Rights Systems
STEM	Science, Technology, Engineering and Mathematics
UDHR	Universal Declaration of Human Rights
UN	United Nations

LIST OF TABLES AND FIGURES

1. INTRODUCTION

1.1 Background

In the last 50 years, global society has raced into the Information Age.[1] As wide-sweeping as the Industrial Revolution, computers have impacted every aspect of modern life in just one lifetime. Mass communication has connected the globe with devices which hold the summation of human knowledge. Countless jobs have been automated by computers that produce billions of calculations per second. As a growing sector, technology has joined the ranks of the revered STEM subjects (science, technology, engineering, mathematics). As such, technology has been framed as a foil to the humanities, lumped in with black-and-white calculations and amoral productivity. Technology is seen as a master key that uses science to unlock objective solutions.

Yet this perception crumbles into dust upon further investigation. Technology is not objective, but a subjective concept. As a product of human action, it represents the interconnection of assumptions, prejudice, social structure and human relationships. Technology's interaction with its social context is such that "technological developments frequently have social and human consequences that go far outside the immediate purposes of the technical devices and practices".[2] Thus, technology itself is not neutral but created by humans and intrinsically weighted by social connotations. Furthermore, the utility of technology is a subjective concept. How a society implements its

[1] The Information Age is characterised by the rapid shift from traditional industry to an economy where the consumption of information becomes central. See Noel Castree and others, *A Dictionary of Human Geography* (Oxford University Press 2013).
[2] Melvine Kranzberg, 'Technology and History: "Kranzberg's Laws"' (1986) 27 Technology and Culture, 545-546.

technology reflects its goals. As a society driven by efficiency, our technology is tailored for time-saving measures. Therefore, both technology and the situations in which it is used mirror subjective cultural values.

One particularly time-saving technology is Artificial Intelligence (AI). Permeating both private and public spheres of our lives, AI has changed society as we know it. The European Parliament recently described AI as the epicentre of the "fourth industrial revolution".[3] Many industries prioritise the implementation of AI for its massive utility within society. For instance, AI is used within social services, hiring selection and improving diagnostics in healthcare.[4] Pedro Domingo portrayed this conquest in his book, *The Master Algorithmic*, asserting that "if every algorithm suddenly stopped working, it would be the end of the world as we know it".[5]

However, despite the potential opportunities in technology and AI, scholars also have concerns for human rights.[6] As technology is such a wide-encompassing phenomenon, the realm of AI can serve as a microcosm of the range of issues that confront legal experts today. One problematic area is the technology's negative impact on gender equality. A prominent social scientist in AI and gender, Kate Crawford, believes people should not fear algorithms becoming too smart. Instead, she argues that the pressing issue is that algorithms risk hard-coding sexism into the digital web of infrastructure on which we now build our societies.[7] In 2018, Reuters reported that Amazon's AI hiring tool favoured male applicants over female applicants. For instance,

---

[3] Special Committee on Artificial Intelligence in a Digital Age, 'Draft report on artificial intelligence in a digital age' 2020/2266/(INI) para 7.
[4] McKenzie Raub, 'Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices' (2018) 71 Ark. L. Rev., 530.
[5] Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Out World* (Penguin Books Ltd 2015) 14.
[6] Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021) 8.
[7] Kate Crawford, 'Artificial Intelligence's White Guy Problem' *The New York Times* (June 25, 2016) < https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> accessed 1 February 2022.

the system deducted points for resumes that included '*women's* [emphasis added] chess club captain'.[8] Many such occasions illustrate where AI hems the promotion of gender equality, triggering regulatory human rights frameworks. One such regulatory protection is the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW).

As actors within the national human rights system (NHRS), scholars play an important role in promoting and respecting human rights.[9] Part of this role entails assessing current and future human rights issues.[10] Likewise, scholars use their publications to create narratives of how society should address human rights issues.[11] Due to the large-scale impact of AI, the errors within the algorithmic systems are disseminating swiftly, necessitating immediate action. Recognising this urgency, the UN High Commissioner implores human rights scholars to '[assess] and [address] the serious risk this technology poses to human rights'.[12]

1.2 Aim and Research Questions

Much work has emerged to tackle AI's discriminatory impacts on women in recent years. This thesis aims to identify the prevailing narrative built around AI and gender by mapping the legal literature. Furthermore, by positioning the narrative within the social change paradigm, the thesis discusses the proposed academic solutions' possible impacts on promoting substantive gender equality.

---

[8] Jeffry Dastin, 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women' *Reuters* (October 10, 2018) < https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> accessed 8 February 2022.
[9] The Danish Institute for Human Rights, 'HRS Concept Note: Academia' (May 2018) <https://www.humanrights.dk/sites/humanrights.dk/files/media/migrated/hrs_toolbox_concept_note_academia_may2018.pdf> accessed 9 February 2022, 2-3.
[10] ibid 3.
[11] ibid 4.
[12] United Nations, 'Urgent Action Needed over Artificial Intelligence Risks to Human Rights' *UN News* (15 September 2021) <https://news.un.org/en/story/2021/09/1099972> accessed 9 February 2022.

The specific research questions are:

- *What are the issues and prevailing perspectives discussed in the legal literature on artificial intelligence and gender?*
- *How do the solutions predominantly suggested by researchers endorse CEDAW's goal of greater substantive gender equality?*

1.3 Theoretical Approach

The thesis employs the *theory of social change* as an analytical framework to examine how the predominantly suggested academic solutions align with CEDAW, the international human rights framework for gender equality.

As a *concept*, social change is used within sociology to explain an alteration in established modes of behaviour in society.[13] Concurrently, social change as a *theory* explains the mechanisms behind the alteration of societal conduct. The thesis uses the theory's fundamental structure to explain that CEDAW's goal of achieving substantive gender equality demands a structural change in society. However, the thesis does not consider *how* this change should happen; therefore, it does not discuss how different approaches within the theory of social change best answer the needed social change for greater gender equality.

Overarching Components

The theory of social change has two dimensions: *descriptive* and *prescriptive*. The descriptive dimension uses data to describe the observable surface structure. On the other hand, the prescriptive dimension refers to the 'judgement of change' closely interlinked with the term development.[14] In

[13] John Lewis Gillin and John Philip Gillin, *Cultural Sociology* (Macmillan Company 1954) 561-562.
[14] Amitabha Das Gupta, 'Change, Development and a Theory of Social Science' (1989) 24 Economic and Political Weekly, 36.

other words, the descriptive facet explains how it *is*, while the prescriptive facet explains how it *should be*.

Building upon the two dimensions, the theory frames the thesis in two ways; the first research question connects to the descriptive dimension. The theory helps position the surveyed literature and suggested solutions within the social change paradigm through collected data. The second research question is part of the prescriptive dimension. It highlights the limitations of predominantly suggested solutions and the readjustments necessary to reach CEDAW's substantive gender equality objective.

Content

The theory expands upon the idea that 'society has a physiological structure containing different components, each of which has different functions but performing as a single whole'.[15] When these units are altered, either by external or internal factors, the dominant societal structure is affected.[16] Consequently, the theory creates a paradigm that encapsulates the mechanisms necessary for societal change.[17]

The content and weight given to each component differ in the social change literature. In his article, Burke Hendrix argues that it is possible to identify three core barriers to societal change: social, economic and cultural.[18] Each type is supported by the works of Michel Foucault, Karl Marx and John Stuart Mill, respectively.

Actual transitional change requires changes within social institutions.[19] Foucault postulates that the intellectual frameworks of our society form our

[15] ibid 40.
[16] Neil Flingstein and Doug McAdam, *A Theory of Fields* (Oxford Press 2012) 8.
[17] Gupta (n 14) 35.
[18] Burke A Hendrix, 'Where Should We Expect Social Change in Non-Ideal Theory?' (2013) 41 Political Theory, 116.
[19] Hendrix (n 18) 129.

behaviour and "construct us as persons" through multiple organisational methods.[20] Law is one of the institutional avenues of society that can create societal change. However, as law is only one fragment of the "complex organizational system" upon which society is built, law and regulations can "achieve certain things, but not others".[21] For effective change, law cannot succeed alone but requires the cooperation of other societal institutions.[22] One violin section cannot perform a symphony without the woodwinds or the brass accompaniment.

Marx also contributed to the ideas of social change. In one of his notebooks, he wrote, 'Philosophers have so far only changed their interpretation of the world; the point however is to change the world'.[23] While scholars debate Marx's theory of social change, it mainly focuses on the structural activities in the areas of economy and production.[24] Marx believed that the mode of production determines human behaviour. The mode of production is, in turn, influenced by technology, sometimes to the extent of technological determinism.[25] As Marx held in his book *Poverty of Philosophy*, 'the wind-mill gives you a society with the feudal lord; the steam-mill a society with the industrial capitalist'.[26] According to Marx, the changes in the mode of production and economic structure shape the organisation of society.

However, feminist scholars criticize Marx's view of technology as a driving force for social change. They assert that technology is far from an autonomous force that forms society—instead, it is a direct result of power structures and

[20] Michel Foucault, *Discipline and Punish: The Birth of the Prison* (Penguin Books 2019) Part III.
[21] Hendrix (n 18) 129.
[22] ibid.
[23] Frederick Engels, *Ludwig Feuerbach and the Outcome of Classical German Philosophy* (International Publishers 1941) 73.
[24] ibid 122.
[25] Bruce Bimber, 'Karl Marx and the Three Faces of Technological Determinism' (1990) 20 Social Studies of Science 333.
[26] Karl Marx, *The Poverty of Philosophy* (Harry Quelch tr, Cosimo Classics 2008) 119.

consequently far from a neutral concept. As such, feminist technology studies highlight that technology is 'socially constructed, or coproduced, alongside gender'.[27] Subsequently, uncritical endorsement of technology as the prominent actor for social change remains naïve to the social realities that shape it.

Unlike Marx, Mill believed that the key to societal change resides at the individual level. Through patterns of belief, ideas are enforced and passed on through generations, creating a culture within society.[28] Flawed cultures lead to flawed societies. Therefore, the intellectual inheritances from previous eras need to transform to achieve societal change.[29]

Despite the discrepancies the three approaches display, they all identify a single-factor explanation for social change. However, contemporary theories of social change move away from these single axes of determinism and view change instead as multi-dimensional and sectional.[30] Based on this approach, the identified three drivers are not *single* variables but a system of variables and, as such, must be viewed as interconnected.[31]

1.4 Method and Materials

Shaped by the descriptive component of the Theory of Change, this thesis conducts a *systematic legal literature review*. The descriptive dimension necessitates an 'observable structure of data', which in this case is attained through the mapping of relevant legal literature in the areas of AI and

[27] Wendy Faulkner, 'The Technology Question in Feminism: A View from Feminist Technology Studies' (2001) 24 Women's Studies International Forum, 79.
[28] Hendrix (n 18) 123.
[29] ibid 122.
[30] Wilbert E. Moore, 'A Reconsideration of Theories of Social Change' (1960) 25 American Sociological Review, 811.
[31] Raymond Boudon, *Theories of Social Change: A Critical Appraisal* (JC Whitehouse tr, Polity Press 1986) 15.

gender.[32] In turn, the data enables the dissemination of trends within the research community, later used as part of the prescriptive element (*chapter 5*) of the thesis.

The systematic legal literature review was conducted using two scientific literature databases: *LUBsearch* and *HeinOnline*. LUBsearch was chosen because it functions as a collective entry point for literature databases within many different scientific subjects. It provides a broad search coverage and proves an excellent tool for interdisciplinary research. However, whilst LUBsearch envelops many databases, including HeinOnline, it sometimes fails to encapsulate all the databases' unique materials. HeinOnline was specifically chosen because it targets legal journals and is perfectly tailored for legal research. Of course, other databases such as Scopus exist as well. However, preliminary searches showed that most articles therein leaned towards social sciences with a dominant focus on gender and lacked a legal emphasis and legal discussions.

The following section explains how the systematic legal literature was conducted. The inquiry comprised three steps: (1) gathering of literature, (2) exclusion of literature and (3) analysis.

1.4.1 Gathering of Legal Scientific Literature

To find relevant legal literature, two complementary search techniques were used to conduct a systematic search. The first technique consisted of identifying keywords related to AI and gender. The keywords were then compiled into a search string and inserted in the two scientific databases. The second complementary technique comprised of a so-called "snowball search".

[32] Marnix Snel and Janaina De Moraes (eds), *Doing a Systematic Literature Review in Legal Scholarship* (Eleven International Publishing 2018) 7.

As to the first technique, keywords representing the research questions were identified through an initial scoping of the most recent peer-reviewed articles in the databases. Then, building on these initial keywords, several searches were conducted to narrow down and develop the specific search terms. These initial searches produced a general overview of the keywords and synonyms used in the academic literature on AI and gender and thus aided in creating a suitable and specific "search string".[33] After the initial search, the process continued by testing different combinations of search terms, principally by including synonyms within each block and adding and removing blocks within the advanced search option (*see Appendix A*).

The main difficulty in producing a suitable search string was achieving an appropriate number of relevant articles. As the thesis aimed to map the legal scholarship on AI and gender, the search needed to contain a sufficient number of articles to give an accurate picture of the academic landscape.

After conducting a preliminary search, titles and abstracts of articles from the first page of the database were examined to determine if the search string resulted in relevant articles. The search was also narrowed down using the advanced search syntax options, such as Boolean Logic and Proximity Indicators.[34]

The final search string is shown in *Table 1*. Block 1-2 delineates the central area, artificial intelligence and gender. The term "equality" was an important keyword to incorporate into the search string for two reasons; firstly, most relevant literature included the term, and therefore relevant articles fell away if not included in the search string. Secondly, merely using the term gender was not specific enough to draw out articles concerning gender issues and technology. However, when the terms gender "OR" equality were used in

[33] ibid 50.
[34] ibid 51.

block two, it gave rise to too many hits. The idea with block three was thus to narrow down the search articles. Finally, block four intended to cover legal articles on the topic in order to map out the legal literature on AI and gender.

The search was limited to peer-reviewed articles in academic journals. Delimitation by English was also necessary as the final search string result contained ten articles in other languages. No temporal parameter was chosen as most relevant articles were written from 2018 to 2021. The final search string resulted in 183 hits in LUBSearch and 221 hits in HeinOnline.

***Table 1**: Final search string applied in both LUBsearch and HeinOnline.*

Keywords				**Delimitation**	**Database**	**Search date**	**Hits**
Block 1	**Block 2**	**Block 3**	**Block 4**				
"Artificial Intelligence" AND AI (All text)	Gender (All text)	Equality (All text)	Law AND "Human Rights" (SO Journal Title/Source)	"Peer Review" and "Journals" and "English"	LUBsearch	**2022-01-22**	183
					HeinOnline	**2022-01-27**	221

The second technique used to identify pertinent articles is referred to as 'snowballing', assuring that the thesis encapsulated the majority of relevant literature on the topic. The "snowballing technique" entails scrutinising the reference list of already-unearthed key publications.[35] The key publications were chosen in part depending on the number of citations. However, citations are not always a good proxy for a paper's quality. Therefore, key publications were also picked through abstracts from the already-chosen articles.[36] 'Bias

[35] ibid 52.
[36] Haochuan Cui and others, 'Identifying the Key Reference of a Scientific Publication' (2020) 29 Journal of Systems Science and Systems Engineering, 429.

Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law' is an example of one of the chosen key references.[37] When five consecutive key references did not produce new relevant articles, the search ceased. In total, six additional articles were found.[38]

1.4.2 Application of Exclusion Criteria

Not all identified articles were relevant for the research aim. Therefore, it was necessary to identify appropriate exclusion criteria. If the articles did not contain all three themes of law, artificial intelligence and gender, they were eliminated. Furthermore, duplicates and articles that only mentioned gender or AI in the footnotes were ruled out.

A manual selection was conducted using these exclusion criteria by reading titles and abstracts of the 183 articles found in LUBsearch. If uncertain of the relevance, a closer examination was conducted by studying the outline and skimming through the article. The same procedure was applied to the 221 hits found in HeinOnline. However, because the articles overlapped with the LUBsearch hits, duplicates already found in the previous step were screened out directly. The final number of identified articles from the first part of the literature search was 72 *(see table 2)*.

[37] Sandra Wachter and others, 'Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law' (2021) 123 West Virginia L Rev.
[38] See the *Saturation Principle* in Van Dijck Gijs, 'Legal Research When Relying on Open Access: A primer' (2016) 6 Law and Method, 10.

Table 2: *Number of included articles after applying the selection criteria.*

Database 1 – Lubsearch		
Date	**Hits**	**Selected**
22 January 2022	183	55
Database 2 – HeinOnline		
27 January 2022	210	11
Snowballing		6
Total		**72**

1.4.3 Analysis

All 72 articles were used to answer the first research question. They provided an overview of the specific research topics that scholars explore in the area of AI and gender. To answer the second research question, articles that focused on solutions to identified problems within AI and gender were selected, resulting in 24 articles. The full list of articles is catalogued in *Appendix B* and cited in the bibliography under "Academic Journals."

1.5 Delimitations

The AI and human rights field opens up numerous stimulating discussion points and issues. However, due to the scope of this project, certain delimitations have been applied to the thesis. Firstly, the thesis does not cover other grounds for discrimination than gender, such as ethnicity, age or disability.

Secondly, whilst the thesis touches on intersectional discrimination, it does not go fully in-depth, as the examined literature lacks discussions on intersectionality. To do so would inaccurately claim that intersectionality is currently involved in the discussions on gender and AI. This can, of course, be criticised as a flaw within the current literature but cannot be fixed in this paper.

Thirdly, the potential human rights impacts of AI concerning gender go far beyond merely the issue of discrimination. Other human rights such as the freedom of expression and the right to health are also impacted through AI, disproportionately affecting women. However, as the current literature mainly discusses discrimination, this thesis does not consider other human rights.

Finally, only technological solutions were examined, as the most prevalent, to keep the thesis focused. The examined literature also mentioned legal solutions, but these were not further analysed. Furthermore, most legal solutions discussed data privacy, the General Data Protection Regulation and patent law questions. These topics fall outside of this thesis' scope.

1.6 Outline

In her book, *Artificial Knowing*, Alison Adam claims that tackling a transdisciplinary issue through academia is like providing the reader with a 'Chinese banquet, made up of lots of little courses with different flavours'.[39] Drawing upon this analogy, I hope that the reader may leave the table feeling full, the palate satisfied by the combination of tastes, and hopefully, not discomforted.

The thesis consists of six chapters. Following the introductory chapter, C*hapter 2* defines AI, underlining that the technology is not limited to mathematics and engineering but is intrinsically shaped by social factors. After laying this foundation, the technology itself is examined, helping the reader understand the different types of AI biases and how human intervention plays a part in their creation.

[39] Alison Adam, *Artificial Knowing: Gender and the Thinking Machine* (Routledge 2006) 3.

Chapter 3 focuses on the legal elements of the thesis. Firstly, it describes the international human rights sources prohibiting discrimination. This illustration establishes a connection between AI biases and the prohibited discriminatory results. After that, the chapter examines the term "equality," linking the discussions to the core legal framework used in the thesis, CEDAW. It is this legal framework upon which later discussions build.

Chapter 4 disseminates the systematic legal literature review findings, providing the reader with a mapping of the issues and suggested solutions as articulated by the scholars.

Chapter 5 discusses the findings by positioning the prevalent techno-narrative within the social change paradigm, highlighting the solutions' effect on promoting substantive equality.

Finally, *Chapter 6* summarises the key takeaways, circling back to the thesis' research questions and aim.

2. ARTIFICIAL INTELLIGENCE AND REPRODUCED BIASES

2.1 Introduction

The chapter targets individuals with a non-technical background, aiming to provide readers with an overall comprehension of AI systems. Although this may seem daunting for lawyers, it is necessary. We can only adequately solve the technology's gender equality issues by understanding *how* biases are introduced and amplified through these systems.

The chapter is structured into three parts. The first part contextualizes the debate around the definition of AI. Part two centres around machine learning, a subfield of AI, and how it works. In part three, the focus shifts to the different biases that the technology produces.

2.2 Defining Artificial Intelligence

According to the computer scientist Jerry Kaplan, trying to define the term Artificial Intelligence is an 'easy question to ask and a hard one to answer'.[40] The difficulty in answering the question lies in the fact that the term does not have a general agreed-upon definition. However, the EU attempted to define it in the new AI Act:

[40] Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford University Press 2016) 1.

> "Artificial intelligence system" (AI system) refers to software developed with one or more of the techniques and approaches listed in Annex I. From a given set of human-defined objectives, it can generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.[41]

The approaches listed in Annex I of the act are machine learning, symbolic approaches and statistics. This definition has already created upheavals of dismay amongst lawyers and scientists. While some statisticians claim the definition to be overbroad, lawyers argue that it will soon be an outdated description of the technology, not a definition.[42] The futility in pinning down a precise definition stems from three factors.

Firstly, as soon as one closes in, the definition retreats into the distance, like trying to reach the end of a rainbow. AI is not static, instead, it is an umbrella term containing various constantly-evolving techniques, such a Neural Networks.[43] Subsequently, it is impossible to identify one form of a AI.[44]

The second issue in defining AI is the notion of *intelligence*. Although studied in depth by scientists, philosophers and psychologists, it remains an ambiguous concept. [45] Not only is 'intelligence' difficult to define in humans, the concept of AI 'intelligence' is also debated. And it gets more complicated; even though intelligence is a benchmark of determining whether or not a technology is "worthy" of AI status, it is often not enough on its own. In 1997,

---

[41] European Commission, 'Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (21 April 2021) COM/2021/206 final, art 3.
[42] See for example, Laurie Clarke, 'The EU's leaked AI regulation is ambitious but vague' *Tech Monitor* (15 April, 2021) <https://techmonitor.ai/policy/eu-ai-regulation-machine-learning-european-union> accessed 11 February 2022.
[43] Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer Nature 2020) 6.
[44] ibid 7.
[45] European Commission, 'A Definition of AI: Main Capabilities and Scientific Disciplines' (8 April 2019) <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> accessed 11 February 2022.

the Deep Blue program beat grandmaster Garry Kasparov at a game of chess. Even though the program clearly was more "intelligent" than Kasparov in this one area of expertise, the AI system was still not considered an overall "intelligent" creature in succeeding publications.[46] The rainbow had moved once again.

Thirdly, scholars have attempted to define AI through a purely technical lens. The predicament is that AI is a social construct in addition to a technical concept, inspiring the entire field of constructivism. Social constructivist scholars argue that technology is linked to social conditions and human action.[47] They claim that society shapes technology instead of the other way around. Therefore, to fully grasp and define AI, it is necessary to view the technology through the political, economic and historical forces that shape it.[48] Because AI is 'technical and social practices, institutions and infrastructures, politics and culture', finding a shared definition encapsulating these simultaneous concepts is challenging.[49]

Despite the difficulty in defining AI, it is possible to separate between *general* and *narrow* AI. General intelligence refers to systems that show human traits. However, this type of technological sophistication is yet to be achieved.[50] By contrast, narrow AI can resemble human competencies in a limited area.[51] For instance, Deep Blue, the chess-playing AI could beat one of the world's best chess champions because it was programmed for that specific task. However, presented with a mission outside its operation field, such as facial recognition, Deep Blue would fail. Consequently, narrow AI appears intelligent but does

[46] See for example, Murray Campbell and others, 'Deep Blue' (2002) 135 Artificial Intelligence, 57-83.
[47] See for example, Hans K Klein and Daniel Lee Kleinman, 'The Social Construction of Technology: Structural Considerations' (2002) 27 Science, Technology, & Human Values.
[48] Crawford (n 6) 8.
[49] ibid.
[50] ARTICLE 19, 'Privacy and Freedom of Expression in the Age of Artificial Intelligence' (April 2018) 6.
[51] ibid.

not truly understand reality.[52] In contrast, general AI is a mind that comprehends and can experience other cognitive states.[53]

In conclusion, AI remains a tenuous term. The difficulty in finding a standard definition also makes describing AI arduous. However, what remains clear is that AI is more than merely a technical term. Accordingly, the thesis understands AI as an umbrella term, encapsulating both the technical aspects of the machinery and the social structures which shape it. The next part of the chapter will focus on the core techniques of machine learning, which is a subfield of AI.

2.3 The Fundamentals of Machine Learning

Machine learning refers to a subfield of AI; one could even call it the "backbone" of AI. The technique allows computer programmes to learn from data inputs to independently improve its algorithms.[54] Of course, the computer does not "learn" in the same manner as human beings. Instead, the algorithms learn in an operative sense.[55] Through data input, the algorithms detect patterns and convert them into knowledge, allowing the computer to make predictions.[56] The more data the algorithm is exposed to, the more the AI improves its "intelligence". [57] Consequently, a core component of machine learning is data, the fuel that keeps AI running and the source of its "knowledge".

To help illustrate some basic machine learning features, let us use the email spam filter as an example. Email software programmes commonly use AI to

[52] Richard E Neapolitan and Xia Jiang, *Contemporary Artificial Intelligence* (Taylor & Francis Group 2013) 4.
[53] John R Searle, 'Minds, Brains and Programs' (1980) 3 Behavioural and Brain Sciences, 417- 424.
[54] Harry Surden, 'Machine Learning and Law' (2014) 89 Washington Law Review, 89.
[55] ibid.
[56] ibid.
[57] Harvard Business Review Press and others, *Artificial Intelligence: The Insights You Need from Harvard Business Review* (Harvard Business Review Press 2019) 14.

help filter out unwanted emails.[58] The algorithms are trained through data inputs containing different examples of emails predetermined by a human as either non-spam or spam.[59] This teaching method is called *supervised learning* and builds upon "pre-labelled training data".[60] From the inputs, the algorithm then identifies patterns using the data provided.[61] For instance, the algorithm may conclude that most spam emails use capital letters in the title or use the expression "Earn Money". The algorithm will then collect similar patterns and use them as a heuristic, creating a rule of thumb.[62] As more data is inserted into the input layer, the "rules" accumulate, increasing the accuracy in the output layer. Ultimately, the input data leads to an output of determining whether the email is spam or not.[63]

Whilst the algorithms used to detect spam emails are relatively simple, more intricate algorithms, such as Neural Networks, are employed for harder tasks. For example, the healthcare and judiciary systems require more complexity.[64] Today, Neural Networks are one of the most commonly used algorithms within AI, creating its own subsection, *deep learning*.[65]

Deep learning is an advanced form of machine learning. However, compared to machine learning, which uses conventional algorithms, Neural Networks are sophisticated algorithms that evolved by mimicking the human brain.[66] Neural Networks consist of interrelated components called neurons. There are

[58] Surden (n 54) 90.
[59] ibid 93.
[60] Apart from 'Supervised Learning', there are two other training methods: 'Unsupervised Learning' and 'Reinforcement Learning'. However, supervised learning is the most commonly used method within AI systems today. To read more about the other methods, see chapters 1.4-1.5 in Zoltán Somogyi, *The Application of Artificial Intelligence: Step-by-Step Guide from Beginner to Expert* (Springer Nature 2022).
[61] ibid 8.
[62] ibid 91.
[63] ibid 93.
[64] Francesco Contini, 'Artificial Intelligence: A New Trojan Horse for Undue Influence on Judiciaries?' (*UNODC,* 2019).
[65] Ian Goodfellow and others, *Deep Learning* (MIT Press 2016) 13.
[66] Charu C Aggarwal, *Neural Networks and Deep Learning: A textbook* (Springer International Publishing 2018) 1.

three different types of neurons: input, hidden and output neurons. These neurons are organised into different *layers*.[67] Typically, there exists an input layer, a hidden layer and an output layer. The figure beneath illustrates how Neural Networks can look (*see figure 1*).[68]

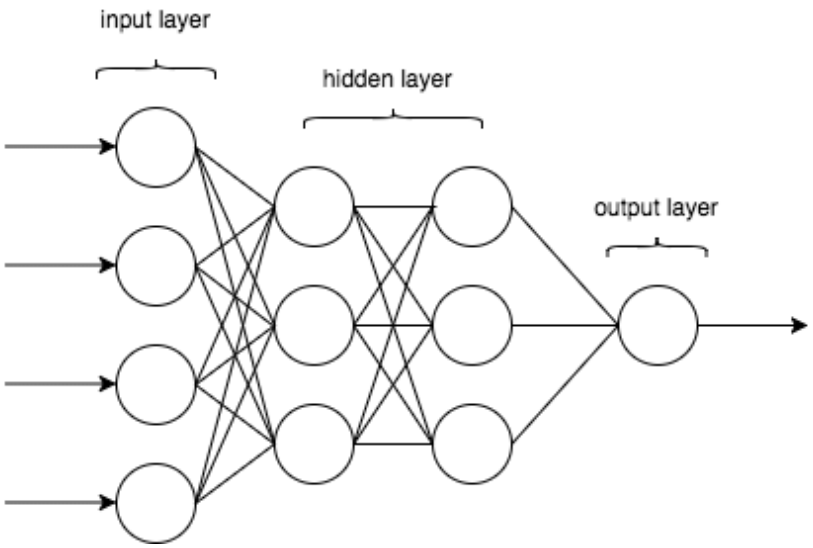


Figure 1: *An illustration of a Neural Network*

The data is introduced into the Neural Network through the input layer. The data is then directed towards the hidden layer. The neurons there process the received data by finding patterns from which it can draw conclusions.[69] These conclusions then cumulate into a result situated within the output layer.

Although both the input and output layers are possible to interpret and comprehend, one of the central conundrums with Neural Networks is that it has no observable hidden layer. Thus, the conclusions the algorithm draws in the hidden layer are inscrutable. Within scholarly literature, this issue is called

[67] ibid 18.
[68] The figure is taken from Tanvi Bhandarkar and others, 'Earthquake Trend Prediction Using Long Short-term Memory Neural Networks' (2019) 9 International Journal of Electrical and Computer Engineering, 1305.
[69] Fahmi Nurfikri, 'An Illustrated Guide to Artificial Neural Networks' (Towards Data Science, July 20 2020) < https://towardsdatascience.com/an-illustrated-guide-to-artificial-neural-networks-f149a549ba74> accessed 20 February 2022.

the "black-box".[70] Like in mathematics, if one cannot show the process, it can be difficult to trust the end result.

2.4 Mapping the Emergence of Gender Biases in Artificial Intelligence

This segment guides readers to understand the multifaceted issue of AI biases and, later, apply this framework to the proposed solutions identified in the academic literature. This chapter will thus help structure the discussions of AI biases by enabling the reader to systematise how gender biases emerge in AI.

Biases can enter AI systems in three different phases: (1) through the dataset, (2) through the algorithm and (3) through the application of the AI system. This section breaks down *how* and *why* the biases are infused through the system. Each phase is dissected into smaller parts *(see Figure 2)* and accompanied by examples.

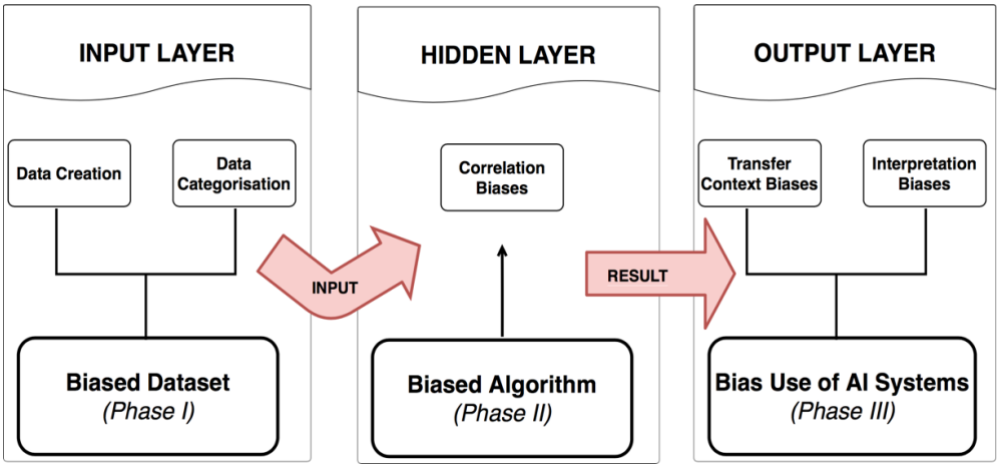


Figure 2: *Mapping of AI biases.*

[70] Jianlong Zhou and Fand Chen, *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent* (Springer 2018) 120.

2.4.1 Biased Dataset

As mentioned in *section 3.2*, machine learning systems are trained on data. The datasets form the core of the machine's "intelligence". The systems also use the data to make predictions.[71] In the first phase, there are two main ways biases materialise and seep into the computer system. It happens either through the creation of the data itself or once handled by the computer scientist through the categorisation of data.

Data Creation

Data is neither a neutral concept nor objective.[72] Numbers collected in statistical graphs are not raw material. Instead, like any other product, it is produced, collected and passed through a series of political and institutional actors with specific interests.[73] Thus, in the case of Artificial Intelligence, the saying 'numbers speak for themselves' does not ring true.[74] Furthermore, because data is a product of historical and current inequalities, certain data may be missing, or data may exist but is coloured by biases.[75] Thus, the popular STEM saying, 'garbage in, garbage out', helps summarise the issue of incorrect or poor quality data.[76] Incomplete data cannot be magically transformed into comprehensive results, but retains its inherent quality. Consequently, data must be understood together with the context in which it is produced.

---

[71] Crawford (n 6) 97.
[72] Catherine D'lgnazio and Lauren F Klien, *Data Feminism* (MIT Press 2020) 149.
[73]Zoe Corbyn, 'Catherine D'lgnazio: "Data is never a raw, truthful input- and it is never neutral"' *The Guardian* (21 March, 2020) <https://www.theguardian.com/technology/2020/mar/21/catherine-dignazio-data-is-never-a-raw-truthful-input-and-it-is-never-neutral> accessed 17 February 2022.
[74] ibid.
[75] Genevieve Smith and Ishita Rustagi, 'Mitigating Bias in Artificial Intelligence: An equity Fluent Leadership Playbook' (Berkeley Hass, July 2020) 24.
[76] Catherine D'lgnazio, '5 Questions on Data & Justice with Cathy O'Neil' *Medium* (26 November, 2017) <https://medium.com/data-feminism/5-questions-on-data-justice-with-cathy-oneil-87f42355ce55> accessed 4 April 2022.

As mentioned above, one issue is the lack of relevant data, so-called *data gaps*. The collection of data or "non-collection" closely intertwines with power and who decides what becomes information and what is silenced. However, the gender data gap is not just about silence; 'these silences have consequences'.[77] In the UK, the government-approved AI healthcare app *Babylon* misdiagnosed a woman for having a panic attack instead of a heart attack.[78] The app underdiagnosed her because her symptoms were "atypical". The app only recognised male heart attack symptoms, which differ compared to female symptoms. This is a problem in many medical textbooks as well; historically, most medical studies have been conducted on male subjects. The problem stemmed from a fallacy in the algorithm's source material, and thus the app lacked data for a correct diagnosis.[79] Similarly, sexual harassment and rape statistics are usually insufficient in many countries. Subsequently, Siri can locate the closest prostitutes but cannot understand the term "rape".[80]

Even when the data *is* collected, it can still be biased—echoing historical and present societal disparities. One study examined work done by Google Translate from gender-neutral languages to English. The researchers concluded that the AI translation reflected gender stereotypes.[81] For example, the app changed the gender-neutral Turkish pronoun "*O* bir doctor" and "*O* bir hemsire" to *he* is a doctor, and *she* is a nurse.[82] As the example shows, the algorithms using biased data can create a vicious circle of gender stereotypes.

---

[77] Caroline Criado Perez, *Invisible Women: Data Bias in a World Designed for Men* (Abrams Press 2019) 3.
[78] Shanti Das, 'It's Hysteria not a Heart Attack, GP App Tells Women' *The Sunday Times* (13 October, 2019) <*https://www.thetimes.co.uk/article/its-hysteria-not-a-heart-attack-gp-app-tells-women-gm2vxbrqk*> accessed 16 February 2022.
[79] Perez (n 77) Chapter 11.
[80] Pam Belluck, 'Hey Siri, Can I Rely on You in a Crisis? Not Always, a Study Finds' *New York Times* (14 March, 2016) <https://well.blogs.nytimes.com/2016/03/14/hey-siri-can-i-rely-on-you-in-a-crisis-not-always-a-study-finds/?mtrref=undefined&gwh=AC21A99572031D9A121000842D645CDC&gwt=pay&assetType=PAYWALL> accessed 16 February 2022.
[81] Aylin Caliskan and others, 'Semantics Derived Automatically from Language Corpora Contain Human-like Biases' (2017) 356 Science, Annex 3.
[82] ibid Annex 4.

Data Categorisation

As mentioned in *section 3.2*, the datasets on which the algorithm is trained are pre-labelled. Data labelling and categorisation, familiar concepts within science, involve human discretion.[83] However, the act of classification is not only scientific but inherently political, with widespread consequences. For example, the Uppsala University library still has the remains of the Institute of Racial Biology's methods for categorising the indigenous Sami people hung up on the walls. From skull measurements to photographs of bodies, this data was taken to prove a pseudoscientific theory based on racial prejudice.[84] Unfortunately, similar practices that are found all over the world have weaselled into our technology.

In one of the most influential studies on classifications, Geoffrey Bowker and Susan Leigh Star wrote that 'classifications are powerful technologies. Embedded in working infrastructures, they become relatively invisible without losing any of their power'.[85] Consequently, classification is a powerful instrument, and as Bowker and Star observe, they become part of everyday life, something we rarely notice. However, when these seemingly pivotal categorisations are built into AI systems, they play a forceful role in moulding society.

The database ImageNet reinforces the gender binary through categorisation of facial images. Including over 14 million categorised images, the database is a primary data source used to develop research for facial recognition software.[86] Regarding gender, the *Female Body* and *Male Body* images within

---

[83] Smith and others (n 75) 29.
[84] Åsa Malmberg, 'Så drabbades samerna av den rasbiologiska forskningen' (*Uppsala Universitet,* 8 December 2021*)* <*https://www.uu.se/nyheter/artikel/?id=17896&typ=artikel&lang=sv*> accessed 18 February 2022.
[85] Geoffrey C Bowker and Susan Leigh Star, *Sorting Things Out: Classification and its Consequences* (MIT Press 2000) 319.
[86] ImageNet (March 11 2021)

ImageNet fall under the categorisation: Natural Object – Body– Human Body– Female and Male Body.[87] Similarly, ImageNet also has a category for *Transgender Body* situated within the subgroup: Person–Sensualist–Transgender Body. The example shows how compartmentalisation reinforces harmful ideas. Firstly, the categorisation of the male and female body under "Natural Object" enforces the false idea that gender is a biological binary concept. Simultaneously, the exclusion of transgenderism from the predominant dichotomy further "others" transgender people.[88]

In practice, this categorisation has real consequences, as seen in Uber's built-in facial recognition program. Intended as a safety feature, a facial recognition system identifies the driver before it unlocks the car. However, because the technology is built upon data that only categorises individuals as male or female, the technology fails to recognise transgender drivers. Consequently, the drivers were kicked off the app because they did not "fit" into the binary norm, costing them income and time.[89] Uber tried to solve the data gap by collecting additional images. The videos of transgender people that Uber sourced from YouTube, however, were taken without the individuals' consent.[90] Subsequently, the simplistic societal understanding of gender, which inadequately grasps gender fluidity, is built into the system by categorising training data—resulting in the eradication of identities in society, affecting the already-marginalised.

---

[87] Crawford (n 6) 138.
[88] For an in-depth reading on how facial recognition, building upon pre-categorised data, reinforces gender norms see Os Keyes, 'Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition' (2018) 2 Association for Computing and Machinery 1.
[89] Sigal Samuel, 'Some AI Just Shouldn't Exist' (*Vox*, 19 April 2019) <https://www.vox.com/future-perfect/2019/4/19/18412674/ai-bias-facial-recognition-black-gay-transgender> accessed 20 February 2022.
[90] James Vincent, 'Transgender Youtubers had their Videos Grabbed to Train Facial Recognition Software' *The Verge* (22 August, 2017) <https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset> accessed 20 February 2022.

2.4.2 Biased Algorithm

Apart from the biased dataset, the algorithm can contribute to biased outcomes. As mentioned in *section 3.2*, algorithms draw conclusions by depicting correlational relationships from the input data. However, correlation does not necessarily equal causation. As such, 'algorithms risk seeing patterns where none exist, simply because massive quantities of data can offer connections that radiate in all different directions'.[91] Thus, the correlations which the algorithms make are not always necessarily correct. These types of biases are called *correlation biases*.[92]

Correlation biases are a problem in job recruitment and hiring settings that use AI systems. These systems are designed to cut administrative costs by eliminating weak candidates automatically.[93] Therefore, the AI system aims to depict how a particular person will fit the job. Usually, this is done using proxy attributes, such as title at previous work, university grades and length of tenure at a past job.[94] However, one study found that AI systems searching for "creative" applicants drew a correlation between creativity and the length of employment at the person's previous job.[95] For women, this faulty correlation created a discriminatory outcome. Typically, gaps in employment or shorter position lengths are directly linked to longer parental leaves taken by women. As a result, the women applying for the job were overlooked, reinforcing societal biases within the corporate culture.[96]

Additionally, the issue of the "black box" can hinder attempts to solve algorithmic biases. Because of the obscurity of the calculation process, even

[91] Danah Boyd and Kate Crawford, 'Critical Questions for Big Data' (2012) 15 Information, Communication & Society, 668.
[92] ibid.
[93] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books 2017) 118.
[94] ibid.
[95] ibid.
[96] ibid.

the visible portions of code are difficult to interpret due to the incomplete data. Subsequently, the inability to test independent variables within the algorithms makes it impossible to test if and how biases affect the outcome.

2.4.3 Biased Application

The final phase in which biases may enter an AI system is through its application. There are two main biases related to application: *transfer context biases* and *interpretation biases*.[97]

Biases can occur even after the algorithm has produced an output; transfer context biases occur when the output is utilised separately from its anticipated context, causing discriminatory results. As noted in *chapter 3.1*, narrow AI systems are developed for particular purposes and specific settings.[98] Outside these operational settings, they will not necessarily perform as envisioned. For example, transfer context biases could arise if a healthcare prediction algorithm created for people in Sweden was transferred to a rural hospital in Lebanon. The algorithm would certainly have noteworthy biases because the transfer context would be entirely different. Thus, AI systems cannot be used as off-the-shelf tools but must be adapted to a specific environment.

Interpretation bias appears in situations where AI is used as decision support.[99] When an output is generated, the result needs to be interpreted by a human being. However, the act of interpretation is highly subjective, affected by individuals' lived experiences. Thus, interpretation bias is a discrepancy between the result that the AI system produces and the information requisite of the user.[100] For example, the judiciary uses AI systems to calculate recidivism. Suppose that the algorithm produces a score

[97] Xavier Ferrer, Tom van Nuenen and others, 'Bias and Discrimination in AI: A Cross-Disciplinary Perspective' (2021) 40 IEEE Technology and Society Magazine, 72.
[98] ibid 73-74.
[99] Smith (n 75) 36.
[100] David Danks and Alex John London, 'Algorithmic Bias in Autonomous Systems' (Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017) 4.

of seven in the case of one defendant. It is ultimately upon the judge to deduce the score and decide the defendant's sentence. However, depending on the judge and their preconceived notions, some may view a score of seven as low, whereas others may interpret it as high.[101]

In conclusion, AI remains a tenuous term for three reasons. Firstly, the field keeps evolving. Secondly, there exists no common understanding of how to define *intelligence*. Thirdly, scholars predominantly view AI purely through a technical lens. However, as the subsequent section illustrates, it is necessary to move away from this one-dimensional understanding and look at the political, economic, social and gender context in which the technology is situated and created. This adjustment enables us to view AI biases as both a technical issue and a product of societal structures and human acts.

[101] Selena Silva and Marin Kenney, 'Algorithms, Platforms, and Ethnic Bias: An Integrative Essay' (2018) 55 The Clark Atlanta University Review of Race and Culture, 22.

3. A LEGAL FRAMEWORK FOR EQUALITY

3.1 Introduction

As established in chapter 2, AI systems can pass along and even create biases. These biases lead to discriminatory results.[102] Considering the relationship between AI biases and discrimination, the purpose of this chapter is to (1) provide an overview of the international norms which prohibit gender discrimination, (2) describe the concepts of direct, indirect and intersectional discrimination as related to AI biases, (3) consider the meaning of *equality* in general terms, and (4) introduce CEDAW's strategies for achieving equality.

3.2 International Legal Sources Prohibiting Gender Discrimination

Equality and non-discrimination loom large in the realm of human rights objectives. Equality is more than an ideal, though; it is an actionable goal. Almost all core human rights instruments refer to the right against discrimination and equal treatment, and manifold efforts work on apprehending it within the human rights legal framework. The Vienna Declaration and Programme of Action depict this right as 'a fundamental rule of international human rights law'.[103]

Discrimination based on gender is prohibited in most human rights treaties. For example, Article 1 of the Universal Declaration of Human Rights

[102] See for example, Trine Rogg Korsvik and others, 'Artificial Intelligence and Gender Equality: A review of Norwegian Research' (Kilden, December 2020).
[103] Vienna Declaration and Programme of Action, adopted by the World Conference on Human Rights on 25 June 1993 (A/CONF.157/24) para 15.

prohibits discrimination based on *sex*.[104] In addition, the UN Charter recognises that human rights should be enjoyed 'without distinction as to (…) sex', as does the European Convention on Human Rights (ECHR).[105]

The International Covenant on Civil and Political Rights (ICCPR) echo similar assertions.[106] Article 2(1) of the ICCPR provides a dependent provision of protection that only concerns civil and political rights within the Covenant. On the other hand, Article 26 of the ICCPR is an autonomous right, prohibiting 'discrimination in law or in fact in any field regulated and protected by public authorities'.[107] Additionally, gender equality is emphasised in Article 3 of the ICCPR, stressing 'the equal right of men and women to the enjoyment of all civil and political rights set forth in the present Covenant'.[108] The Covenant on Economic, Social and Cultural Rights (ICESCR) also prohibits discrimination concerning sex.[109]

Furthermore, CEDAW draws specific attention to gender-related issues, focusing on affirmative measures to address gender inequality.[110] Given the importance of the Convention for gender equality, particular attention is given to CEDAW in *section 3.4*.

---

[104] Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR) art 1.
[105] See, Charter of the United Nations (adopted 26 June 1945) 1 UNTS XVI, art 1(4) and the European Convention for the Protection of Human Rights and Fundamental Freedoms (adopted 4 November 1950, entered into force 3 September 1953) ETS 5; 213 UNTS 222 (ECHR) art 14. Important to note is that art 14 is a dependent provision and exists solely in relation to the other convention articles.
[106] See, International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR) art 2(1) and art 26.
[107] See, UN Human Rights Committee (HRC), 'General Comment No. 18' (10 November 1989) UN Doc HRI/GEN/1/Rev.5, para 12.
[108] See, The Office of the High Commissioner for Human Rights (CCPR), 'General Comment No. 28 on Article 3' (29 March 2000) CCPR/C/21/Rev. 1/Add.10.
[109] See, International Covenant on Economic, Social and Cultural Rights (adopted 16 December 1966, entered into force 3 January 1976) 993 UNTS (ICESCR) art 2(2), art 3 and art 7(a)(i). Observe that there exists no equivalent autonomous article in ICESCR compared to the ICCPR.
[110] Daniel Moeckli and others (eds), *International Human Rights Law* (3rd edn, Oxford University Press 2018) 310.

3.3 Forms of Discrimination

After establishing the prohibition against discrimination of gender, we now move on to grasp the nexus between AI biases and discrimination. Therefore, the following section will briefly consider three different forms of discrimination related to AI: direct, indirect and intersectional discrimination.

Direct Discrimination

Direct discrimination is described in General Comment 28 of CEDAW as 'different treatment explicitly based on grounds of sex and gender differences'.[111] Thus, direct discrimination emerges when a person is treated less favourably than a similar individual solely because of their gender.[112] One example of direct discrimination can be when AI systems are used as a hiring tool. If they have been trained on datasets which contain previously-hired applicants who are majority male, that results in a *biased dataset*. Due to the biased dataset, the AI system, filters out CVs that contain the word "women" during the decision-making process. Consequently, as the AI system is taught to prefer men, fewer women are hired with a direct causal link to their gender.

Indirect Discrimination

Indirect discrimination against women arises when a 'practice appears to be neutral, but has a discriminatory effect in practice on women, because pre-existing inequalities are not addressed by the apparently neutral measure'.[113] Hence, the definition encapsulates practices that outwardly seem "neutral" but result in unequal outcomes because of structural inequalities. Furthermore, indirect discrimination aggravates prevailing disparities

[111] UN Committee for the Elimination of All Forms of Discrimination against Women (CEDAW), 'General Recommendation No 28' (16 December 2010) CEDAW/C/GC/28, para 16.
[112] *Biao v. Denmark* App no 38590/10 (ECtHR, 24 May 2016) para 89.
[113] General Recommendation No 28, para 16.

because of a failure to acknowledge historical discrimination schemes and existing power dynamics between women and men.[114] Building on the previous example of AI and recruitment, one study showed that the frequently-used Human Resources tool 'Xerox' had *correlation biases* related to geography. [115] As a consequence job applicants with long commutes were eliminated because the algorithm identified a correlation between long commutes and decreased efficiency. However, the study discovered another effect of including distance to work: people with longer commutes tended to travel from poor neighbourhoods.[116] Consequently, the AI system implicitly eliminated people based on economic background, causing indirect discrimination.

Intersectional Discrimination

Intersectional discrimination occurs when multiple grounds for discrimination interact simultaneously.[117] The term highlights how categories such as gender and class mutually reinforce each other, perpetuating patterns of injustice. Although CEDAW does not explicitly refer to the term, throughout the Convention it acknowledges that certain groups of women may be discriminated against on multiple grounds.[118] The Committee, on the other hand, explicitly refers to the term "intersectionality," stating that discrimination against gender is inextricably interlinked with other grounds.[119] For example, take the two aforementioned illustrations concerning AI hiring tools. A black woman living in a poorer neighbourhood would experience intersecting forms of discrimination based on gender, race and class.

---

[114] UN Committee for the Elimination of All Forms of Discrimination against Women (CEDAW), 'General Recommendation No 25' (2004) para 7.
[115] O'Neil (n 93) 119.
[116] ibid.
[117] Marsha A Freeman, Christine Chinkin and Beata Rudolf (eds), *The UN Convention on the Elimination of all Forms of Discrimination Against Women: A Commentary* (Oxford University Press 2012) 68-69.
[118] ibid.
[119] General Recommendation No 28, para 18.

3.4 Legal Standards of Equality

As established in *section 3.1*, international human rights instruments aspire to the notion of equality. Many societies claim it as a foundational organising tenet of democracy. On the surface, it may seem obvious what equality entails. Much like the observer effect in physics, however, the more we try to capture the term, the more it eludes us. The core value of human rights may also be the most contested. Accordingly, human rights scholars must accept that views differ, especially regarding the scope of *equality*. Nevertheless, how we interpret equality is not a matter of logic but is intrinsically connected to politics.[120]

Regardless of the difficulty in defining the term, clarifying the conceptual foundation of equality is crucial since our standpoint on equality deeply colours the solutions we seek. This section aspires to consider the two main approaches to equality, *formal* and *substantive equality*.[121]

3.4.1 Formal Equality

Formal equality is attained when two individuals in similar circumstances are treated equally, irrespective of the result.[122] The Aristotelian dictum, 'treating likes alike', is conceivably the most prevalent interpretation of the right to equality, and closely connects to prohibitions against direct discrimination.[123] Furthermore, the concept focuses on individual justice and merit, irrespective of group identity. Formal equality supports the position that a person's characteristics, including gender, ethnicity and age, should be irrelevant when

[120] Moeckli and others (n 110) 149.
[121] Some legal literature views 'transformative equality' as a *third* distinct form of equality. However, in this thesis, as the Committee of CEDAW, it is viewed as part of substantive equality. See CEDAW, 'General Recommendation No 25', para 8.
[122] Moeckli and others (n 110) 150.
[123] Aristotle, *Ethica Nicomachea* (WD Ross trans., Oxford University Press 1925) book v, 1131 a-b.

determining merit. Thus, formal equality may combat negative stereotyping and, to a certain degree, help develop equal enjoyment of rights.[124]

However, some scholars highlight the inadequacy of formal equality in achieving *true* equality. One criticism is that the concept assumes "merit" to be an abstract entity, free from concepts of religion, gender and other characteristics.[125] Instead, critics voice that merit should be viewed as a product of society rather than an intangible entity.[126] Sandra Fredman explains that 'formal equality fails to recognise that such discrimination cannot be attributed to individual acts by specific perpetrators but flows instead from institution and structures of society'.[127] Therefore, the requirement to treat everyone the same, despite differing backgrounds, may entrench pre-existing patterns of inequality.[128]

3.4.2 Substantive Equality

The idea of substantive equality has grown from the criticism directed towards formal equality. Although substantive equality is also a contested concept, there are two main variants: equality of opportunity and equality of results.[129] Building on CEDAW's interpretation of the term, the thesis understands substantive equality as equality of results.[130]

CEDAW recognises the necessity for a contextual approach to equality by eradicating practices and policies that increase disadvantages within

[124] Freeman, Chinkin and Rudolf (n 117) 54. See also, Committee on the Rights of Persons with Disabilities, 'General Recommendation No. 6' (2018) CRPD/C/GC/6, para 10.
[125] Sandra Fredman, 'Substantive Equality Revisited' (2016)14 ICON, 719.
[126] ibid.
[127] ibid.
[128] Moeckli and others (n 110) 150.
[129] Fredman (n 125) 720.
[130] General Recommendation No 25, para 8-9.

society.[131] For example, one method that helps confront structural disadvantages is the prohibition against indirect discrimination.

Unlike formal equality, substantive equality requires going behind the façade of similar treatment. Lifting the smokescreen enables one to view the actual impact of practices and policies. It acknowledges and unravels historical inequalities that form society—pushing for the removal of structural barriers.[132] Thus, substantive equality requires actors to take proactive, targeted steps necessary to eliminate 'asymmetrical structures of power, dominance and disadvantage at work in society'.[133] More generally, it discards the emphasis on the individual. As developed by Ian Hacking, substantive equality highlights that '… the existence or character of X is not determined by the nature of things. X is not inevitable. X was brought into existence or shaped by social events, forces, and history, which could have been different'.[134]

3.5 The Strategies of CEDAW in Achieving Equality

With the incentive to strengthen the position of women, CEDAW is the primary international human rights treaty concerning the protection and promotion of women's rights. Article 1 of CEDAW defines the term "discrimination against women" as:

> Any exclusion or restriction made based on sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, based on equality of men and women,

[131] OM Arnadóttir, *Equality and Non-Discrimination under the European Convention on Human Rights* (Kluwer Law International 2003) 27.
[132] Fredman (n 125) 732.
[133] Arnadóttir (n 131) 27.
[134] Ian Hacking, *The Social Construction of What?* (Harvard University Press 1999) 6-7.

> human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field.[135]

The definition in Article 1 and Article 4(1) highlights CEDAW's purpose to eliminate discrimination against women, achieving both *de jure* and *de facto* equality.[136] Consequently, the strategy to eliminate all forms of discrimination against women focuses on achieving gender equality.[137]

The Convention does not directly define the term *equality.* Instead, it is upon the Committee to articulate the substance. As evidenced most clearly by *General Recommendation No 28*, the Committee has adopted a flexible understanding of gender equality.[138] According to the Committee, the strive for women's equality is achieved through both formal and substantive equality. As Rikki Holtmaat has pointed out, these objectives should not be separated, but used in combination to eliminate discrimination.[139]

The Committee has underlined the significance of formal equality as a building block for change on several occasions.[140] However, aware of formal equality's limitations, the Committee has emphasised that a purely formal approach is 'not sufficient to achieve women's de facto equality with men, which the Committee interprets as substantive equality'.[141] As long as the underlying reasons for discrimination against women are not tackled, the position of women will remain unchanged – unimpaired by a wholly formal equality model.[142]

---

[135] Convention on the Elimination of All Forms of Discrimination against Women (adopted 18 December 1979, entered into force 3 September 1981) 1240 UNTS 13 (CEDAW) art 1.
[136] General Recommendation No. 28, para 4.
[137] Simon Cusack and Lisa Pusey, 'CEDAW and the Rights to Non-discrimination and Equality' (2013) 14 Melbourne Journal of International Law, 10.
[138] Freeman, Chinkin and Rudolf (n 114) 62.
[139] Rikki Holtmaat, 'European Women and the CEDAW-Convention; the way Forward' (EWLA Conference, 2002) 3.
[140] Freeman, Chinkin and Rudolf (n 117) 64.
[141] General Recommendation No. 25, para 8.
[142] ibid. para 10.

For example, in their concluding observations on Serbia, the Committee welcomed the recent legislative changes concerning early marriage and practices against polygamy.[143] However, the Committee disapproved the ineffectiveness of the laws as not 'conducive to the achievement of substantive gender equality in practice'.[144] Simultaneously, Holtmaat asserts that it is, in fact, impossible to eradicate discrimination without addressing its root cause.[145] Subsequently, the Committee asserts that formal equality plays a part in achieving CEDAW's overarching goal, but not enough to reach true gender equality.

In addition to formal equality, CEDAW necessitates measures which ensure substantive equality. The Committee understands substantive equality as *equality of results* because equality of opportunities does not constitute equality 'to the fullest sense'.[146] Consequently, according to the Committee, substantive equality requires strategies for conquering the power imbalance between genders and the underrepresentation of women in society.[147]

CEDAW's approach to equality cumulate into three strategies which promote substantive equality and consequently eradicate discrimination against women; (1) the broad definition of women's discrimination, (2) the promotion of certain permanent and "special measures" and (3) the obligation for state parties to tackle the key sources of women's inequality.[148]

Firstly, the comprehensive understanding of women's discrimination covers a range of conduct. CEDAW prohibits both direct and indirect discrimination

[143] The Committee on the Elimination of Discrimination against Women (CEDAW), 'Concluding Observations: Serbia' (14 May-1 June 2007) CEDAW/C/SCG/CO/1.
[144] ibid. para 15.
[145] Holtmaat (n 139) 4.
[146] *Timor-Leste* CEDAW/C/TLS/CO/1 (CEDAW, 7 August 2009) para 17.
[147] General Recommendation No. 25, para 8.
[148] Moeckli and others (n 110) 315.

and a span of discrimination grounds from "sex" to "pregnancy".[149] The definition also obliges states to ensure non-discrimination in both public and private spheres.[150]

Secondly, CEDAW acknowledges that non-identical treatment may be necessary for certain circumstances. Article 4(1) endorses *special temporary measures* aimed at 'remedying the effects of past or present discrimination against women and promoting the structural, social, and cultural changes necessary'.[151] For example, the adoption of quotas in education can increase the number of female computing students.

The third and final strategy, related to the "transformative requirement" of substantive equality, compels states parties to take positive action in resolving the underlying causes of women's systematic discrimination.[152] The transformative strategy can be separated into two related units. The first unit concerns the 'transformation of institutions, systems and structures that cause or perpetuate discrimination and inequality',[153] while the second unit covers the adjustment of stereotypes. For example, Article 5(a) requires states parties to work towards eliminating harmful gender stereotypes.

In conclusion, the chapter establishes the international human rights sources prohibiting discrimination which set the ground for equality. Furthermore, the chapter demonstrates the nexus between AI biases and discrimination, asserting that AI biases cause direct, indirect and intersectional discrimination. Finally, the chapter examines CEDAW's purpose: to achieve substantive equality. Only when substantive equality is attained can discrimination against women be eliminated. Nevertheless, the Committee

[149] ibid.
[150] ibid.
[151] ibid. 316.
[152] Cusack and Pusey (n 137) 11-12.
[153] ibid. 11.

recognises formal equality as a stepping stone for achieving substantive gender equality. Thus, formal and substantive equality are complementary strategies for achieving social change.

4. FINDINGS

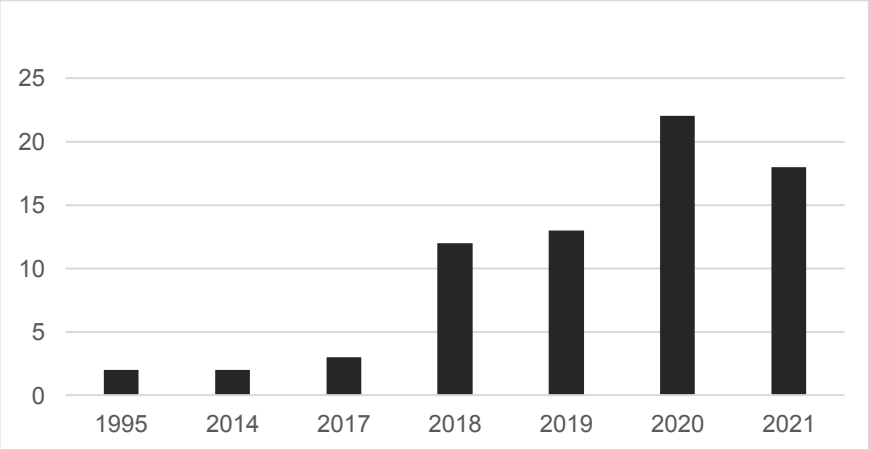
4.1 Introduction

This chapter disseminates the results from the systematic legal literature review. Building on the thesis' two research questions, the chapter is divided into two main sections. The first part presents the reader with a general overview of the legal literature written on AI and gender. The second section provides an account of the solutions against gender discrimination suggested in the literature.

4.2 Mapping of the Literature on Discrimination

Scholars' interest in AI and gender has grown over the past few decades (*see table 3).* Because of this increased attention, it is possible to see a recent but swift development of research on the topic. Yearly output of articles has skyrocketed. Almost no literature was written before 2017, yet 18 legal articles published in 2021 alone.

***Table 3**: Publications over time.*



As established in *section 1.2,* 72 articles were included in the review. Of the 72 articles, 68 of them concerned the issue of discrimination. The remaining four articles focused on granting AI legal personality and AI's impact on the freedom of expression.[154] These numbers indicate that among legal articles on AI and gender, the prevailing focus lies on discrimination. Therefore, this section focuses on the 68 articles related to *discrimination*.

Based on an identification of the principal subject of investigation, the 68 articles are divided into six groups: (i) articles mapping human rights issues, (ii) articles on responsibility, (iii) articles on regulatory protection, (iv) articles targeting specific fields, (v) articles situating AI within a discourse analysis and (vi) articles on LGBTQ+ communities.

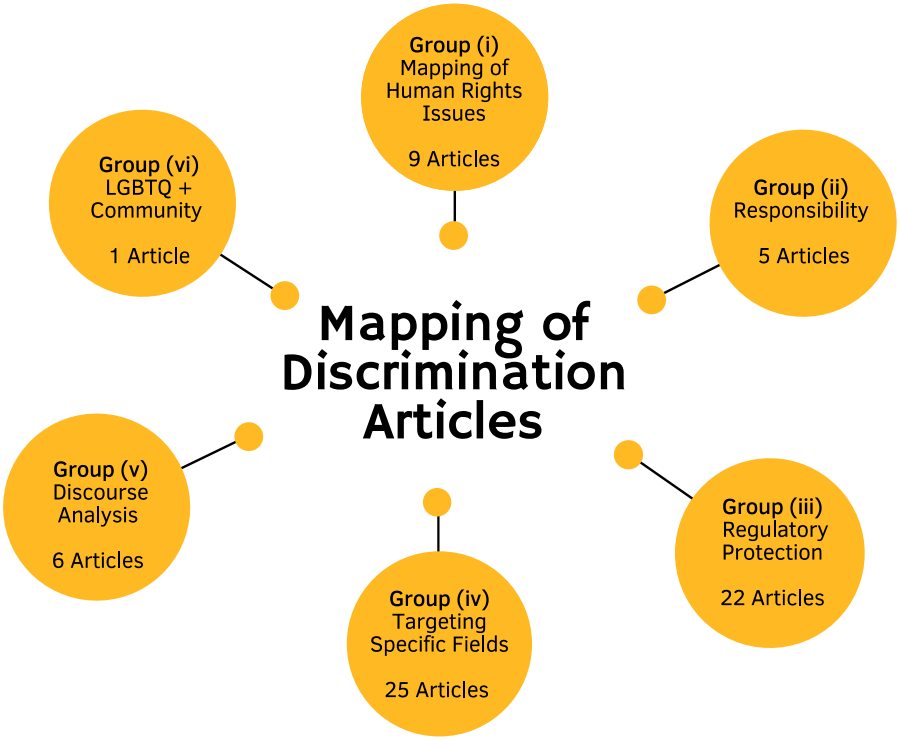


***Figure 3**: Mapping of discrimination articles.*

[154] See for example, Arushi Gupta, 'Assessing the Legal Personality of Sexbots in the Light of Human Rights' (2020) 8 Kathmandu School of Law Review, 90 and Maggie Redden, 'Sophia: The Intersection of Artificial Intelligence and Human Rights' (2020) 10 Journal Global Rights & Org, 155.

Based on this categorisation, it is possible to identify a development in the legal literature. Whilst earlier articles establish the correlation between AI biases and gender discrimination, recent literature from 2021 considers the relationship a fact, focusing instead on critical approaches and solutions. For example, most articles in *group (i)* are published between 2017 and 2018, whereas articles in *group (v)* are written between 2020 and 2021. Overall, articles transition from descriptive to prescriptive. Also, many articles focus on specific fields where AI produces discriminatory effects *(group iv)*. Within this group, it is possible to see specific focal points. Articles on employment practices are most common, followed by articles on the judiciary, advertisement, healthcare, and social services.

4.3 Mapping of Solutions

Out of the 72 articles, 24 discuss solutions; see *figure 4*. Of the 24 articles, the majority are written in 2019 and onwards. The solutions discussed can, in turn, be categorised into four main categories; (i) technical, (ii) regulatory, (iii) team diversification, and (iv) corporate governance and utilisation.

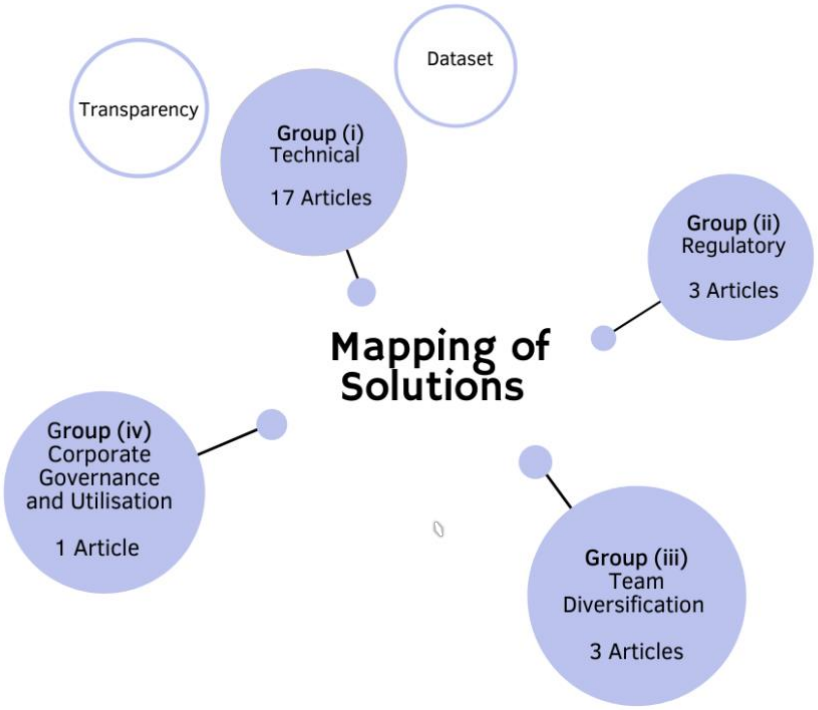


***Figure 4**: Mapping of solutions.*

Technological Solutions

From the articles, it is evident that the dominant focus lies on technological solutions. More than half of the 24 articles recommend tackling the issue of biases with technological fixes. One article states that 'computer science techniques can be used to avoid outcomes that could be considered discriminatory'.[155] In the same spirit, another article emphasises that 'to eliminate biases in the future of AI, the safest way to begin is to incorporate non-discrimination in the initial design of algorithms'.[156] When examined more closely, it is possible to further separate these technical recommendations into two categories: improvement of datasets (9 articles) and increased transparency (8 articles).

Improving datasets is linked to eradicating biases within the dataset *(see figure 2).* According to these articles, the problem 'lies in the data'.[157] Consequently, their solutions focus on correcting the input data. For example, one article discusses algorithms that enable one to "block" specific data. In doing so, the AI can disregard certain features, such as gender.[158]

Similarly, other articles discuss using so-called "corrective training data", which means training algorithms on more inclusive data. Thus, if the algorithm cannot recognise the faces of black women, then "corrective data", would feed it new images, minimising the risk for misidentification.[159]

---

[155] Joshua A Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review, 678.
[156] Haley Moss, 'Screened out Onscreen: Disability Discrimination, Hiring Bias, and Artificial Intelligence' (2021) 98 Denver Law Review, 800.
[157] Ignacio N Cofone, 'Algorithmic Discrimination Is an Information Problem' (2019) 70 Hastings Law Journal, 1410.
[158] ibid 1411.
[159] Valerie Schneider, 'Locked out by Big Data: How Big Data Algorithms and Machine Learning May Undermine Housing Justice' (2020) 52 Columbia Human Rights Law Review, 294.

The remaining articles on technical solutions emphasise *increased transparency* by solving the "black-box issue" *(see chapter 2.4)*. These articles target biases within the algorithm. According to one of these articles, 'the only way to avoid unfair or discriminatory algorithms is to demand greater disclosure of how they operate'. [160] In other words, these articles argue that identifying the technology's discriminatory features requires transparent AI systems. The articles pushing for so-called *Accountable Algorithms* view technology as a tool that 'can help assure accountability' and remove the veil from the hidden layer.[161]

Regulatory Solutions

The three articles discussing protection options through regulatory solutions approach the issue of discrimination from divergent positions. One article argues that the discrimination risk with AI 'can be mitigated by the inclusion of disadvantaged groups (…) in the regulation-making stage'.[162] Others emphasise the gaps between current legal frameworks and reality. For example, several articles discuss how anti-discrimination laws and treaties, such as CEDAW, can be developed to tackle digital discrimination, including those created by AI.[163] At the same time, others highlight the possibility of mitigating risks through data protection and copyright laws.[164]

Team Diversification

Three other articles concentrate on solutions that facilitate the diversification of teams working in STEM. These solutions target biases within the dataset,

[160] Jon Kleinberg and others, 'Discrimination in the Age of Algorithms' (2018) 10 Journal of Legal Analysis, 189.
[161] Kroll (n 155) 657.
[162] Ifeoma Elizabeth Nwafor, 'AI ethical bias: a case for AI vigilantism (AIlantism) in shaping the regulation of AI' (2021) 29 International Journal of Law & Information Technology, 226.
[163] See Tetyana Krupiy, 'Meeting the Chimera: How the CEDAW Can Address Digital Discrimination' (2021) 10 International Human Rights Law Review, 3.
[164] See for example, Laura Stanila, 'Artificial Intelligence and Human Rights: A Challenging Approach on the Issue of Equality' (2018) 2 Journal of Eastern European Criminal Law, 19.

typically focusing on *data labelling*.[165] In 2017, only 19.4 % of the people working in software development in the United States were women.[166] Articles, therefore, call to change the homogenous makeup in STEM through 'policies and practices to facilitate more significant inclusion in the new technological environment'.[167]

Corporate Governance and Utilisation

Lastly, two articles raise the necessity of reconsidering the role of private sectors. The authors believe that the private sectors should ensure that AI systems are utilised in a manner that is 'responsive to women's needs' — necessitating a 'wholesale reform within the industry'.[168] This solution is connected to the biased usage of AI as it reconsiders *how* the finished product should be utilised in a non-discriminatory manner. According to Sonia Katyal, reform must come not from the state but through the increased involvement of the industry in creating the systems.[169]

[165] See, for example, Kimberly Houser, 'Artificial Intelligence and the Struggle between Good and Evil' (2021) 60 Washburn Law Journal, 485.
[166] Moss (n 156) 802.
[167] Peter K Yu, 'The Algorithmic Divide and Equality in the Age of Artificial Intelligence' (2020) 72 Florida Law Review, 368.
[168] ibid 257. See also Rachel Adams and Nóra Ní Loideáin, 'Addressing indirect discrimination and gender stereotypes in AI virtual personal assistants: the role of international human rights law' (2019) 8 Cambridge International Law Journal, 252.
[169] Sonia Katyal, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66 UCLA Law Review, 54.

5. DISCUSSIONS

The thesis now moves to its prescriptive segment, which builds upon the data collected in the descriptive section *(see chapter 4)*. The findings show an increased interest amongst scholars towards AI and gender, especially towards biases. The interest is justified as these biases lead to discrimination, a central human rights issue. However, when the focus area is examined more closely, a narrative begins to form. As scholars begin to realise the scale and human rights effects of AI biases, they are procuring hasty solutions. For example, one article argues that "data can be neutralised" and fosters a sense of overall optimism that once these "issues" are dealt with, the technology can help further equality.[170]

The remaining part of the chapter lifts four interrelated discussion points: (1) identification of the prevalent narrative built around AI and gender, (2) establishment of the link between technocentric solutions and formal equality, (3) positioning the identified narrative within the social change paradigm and (4) discussing the effects of the recommended solutions on the promotion of substantive equality.

The Dominant Perception of AI Frames the Causes of Discrimination and its Solutions as Technical

How we interpret and contextualise a problem shapes its solutions. In this case, our views on discrimination's causes affect how we try to solve it. Consequently, it is crucial to understand how scholars define AI, as their definition forms a narrative concerning the causes of AI discrimination.

[170] See for example, Schneider (n 159) 291-295.

As discussed in *section 2.1*, scholars commonly interpret AI as a technical concept, distinct from the social realities which shape it. However, technology does not materialise out of thin air; on the contrary, it is a direct consequence of human action.[171] Technology is intimately attached to society and is created, adapted and used to carry out societal objectives of any moral standing.

Technology, like maps, are powerful symbols that both echo and manufacture power dynamics and understandings of society.[172] On the surface, maps are considered value-free depictions of the world. However, maps "seek to school the eye", teaching the observer to focus on particular details and features concurrent with politics.[173] Maps provide the viewer with a particular outlook and orientation of the world. Thus, they are a subjective endeavour instead of merely longitudes and latitudes.[174]

By exploring the analogy of a map, we understand the importance of grasping AI in a wider context. As Kate Crawford asserts 'once we connect AI within these broader structures of social systems, we escape the notion of AI as a purely technical domain'.[175] Unfortunately, scholars' limited definition of AI as a solely technical notion has consequences; it frames discrimination within AI as something unexpected and, as a result, masks the root causes for the discrimination.[176]

Although the technical aspects of AI may seem intimidating, AI builds upon a straightforward premise: use data from the past, find patterns and calculate

[171] Smith (n 75) 23.
[172] Crawford (n 6) 9-10.
[173] Lorraine Daston, 'Cloud Physiognomy' (2016) 135 University of California Press, 59.
[174] Crawford (n 6) 11.
[175] ibid. 8.
[176] Julia Powles, 'The Seductive Diversion of "Solving" Bias in Artificial Intelligence' (*OneZero*, 7 December 2018) < https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53> accessed 15 May 2022.

the future.[177] Subsequently, discussing this issue as a system malfunction, like a bug in an app, is incorrect. Instead, we must acknowledge that the discriminatory AI systems are working as envisioned and that the issues are systematic. Thus, the discriminatory outputs are not necessarily a result of insufficient data but accurately reflect social realities. Ultimately, ignoring technology's interconnection with social structures leads to the skewed view of AI biases as technical errors.[178] Instead, we must view this problem both as technical and social, or "technosocial".[179] Without upgrading to a new lens, the critical human rights issue cannot be solved.

Secondly, under the current technical paradigm, AI issues are understood as technical problems, which require technical solutions. Alvin Weinberg coined the term "technological fix" to encapsulate situations where technology is used 'to solve social, political and cultural problems'.[180] Technological fixes prevail in modern society, such as when food famines are combatted with genetically modified organisms and climate change with electric cars. Accordingly, the result from the literature review is not surprising. The dominant narrative frames AI biases as a technical issue, focusing on technical solutions *(see chapter 4.2)*. For instance, one article describes the necessity for lawyers to consider using engineering in their recommended solutions.[181]

[177] Anthony Elliott (ed), *The Routledge Social Science Handbook of AI* (Routledge 2021) 218.
[178] Wachter (n 37) 743.
[179] Shakir Mohamed and others, 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence' (2020) 405 Philosophy and Technology, 4.
[180] Sean F Johnston, 'The Technological Fix as Social Cure-All: Origins and Implication' (2018) 37 IEEE Technology and Society Magazine, 48.
[181] Shlomit Yanisky-Ravid and Sean K Hallisey, '"Equality and Privacy by Design": A New Model of Artificial Intelligence Data Transparency via Auditing, Certification, and Safe Harbor Regimes' (2019) 46 Fordham Urban Law Journal, 428-429.

The technical fixation also explains why most articles fail to look past the 'AI phases' and focus on the processes surrounding the system.[182] Accordingly, no questions are asked on how data is created, collected and used in practice. Instead, discussions on data revolve mostly around quantity and modification. For example, one article stipulates that 'we should use existing pre-processing techniques to alter the data that is fed to the algorithms to prevent disparate impact outcomes'.[183] Similarly, absent from the discussions are debates concerning the context in which the results from the AI systems are used.

Technical Solutions Echo Similarities with Formal Equality

The focus on "fixing" AI biases through technical solutions finds legal parallel in international human rights law, aligning closely with one of the two normative concepts in CEDAW: formal equality.[184] Like formal equality, the framing of AI biases as a technical issue focuses on discrimination attributable to 'individual acts by a specific perpetrator', i.e., AI.[185]

Additionally, the technical solutions seek to reassert the baseline status quo without pursuing the causes of inequality. It confers the discussions to a specific situation and point in time.[186] For example, as mentioned in *chapter 4.2*, one of the articles discusses the technology's ability to block specific data, enabling the AI to disregard certain features, such as gender. Whilst this technical fix can help solve cases of direct discrimination, it is a blunt weapon against indirect discrimination. Moreover, the elimination of details in such a manner is troublesome because it builds upon the assumption that 'we can somehow identify "social factors" which can be cut out, leaving a realm of the purely technical underneath'.[187] Consequently, technical fixations suffer

[182] Sarah Myers West and others, 'Discriminating Systems: Gender, Race, and Power in AI' (AI Now Institute, April 2019) 18.
[183] Moss (n 156) 1389.
[184] Wachter (n 37) 744.
[185] Fredman (n 125) 719.
[186] Wachter (n 37) 744.
[187] Adam (n 39) 2.

from many of the same limitations and criticisms directed towards formal equality. Rather than assuring an actual outcome, it remains essentially passive by overlooking social realities.

Positioning the Technocentric Narrative Within the Social Change Paradigm

Chapter 3.4 of the thesis highlights the Committee's understanding of the right to non-discrimination and equality, interpreting 'equality' in the broadest sense. This ensures that the treaty remains a responsive instrument toward protecting women's rights.[188] However, as established, CEDAW's absolute objective is substantive equality. Thus, CEDAW not only works towards formal equality, but also understands it as complementary to CEDAW's overarching goal. Consequently, legal scholars recommending actions to prevent gender discrimination should align with CEDAW's ultimate goal.

Substantive equality goes beyond equal procedural treatment and focuses on outcomes and structural inequalities.[189] Addressing these inequalities has a 'transformative' dimension, requiring societal change to amend the underlying causes of the disparities.[190] The theory of social change explains the necessary components for the societal transformation required to reach substantive equality. Furthermore, it helps explain why the infatuation of "fixing" AI biases through technical solutions falls short of reaching CEDAW's objectives.

The theory of social change establishes mechanisms necessary for transformation: social, economic, and cultural. Although the Committee does

[188] Cusack and Pusey (n 137) 38.
[189] ibid. 11.
[190] Sandra Fredman, 'Beyond the Dichotomy of Formal and Substantive Equality: Towards a New Definition of Equal Rights' in Ineke Boerefin et al (eds), *Temporary Special Measures: Accelerating De Facto Equality of Women under Article 4(1) UN Convention on the Elimination of All Forms of Discrimination against Women* (Intersentia 2003) 115.

not use these exact terms, they mention the necessity to transform 'institutions, systems and structures'.[191] 'Institutions' refer to the social element, 'systems' relate to the economic dimension and 'structures' link to the cultural and ideological dimensions. All are related to the power relations that social structures produce in any given society.

The technical solutions and, to that extent, formal equality, are essential mechanisms for creating the social change necessary to achieve substantive equality. However, as Michel Foucault claims, social institutional change is only one fragment of society's multifaceted organisational structure.[192] Therefore, although formal equality is part of social change and complementary to the achievement of substantive equality, other targeted solutions are necessary. For instance, a few articles suggest diversifying people working within STEM. One way of achieving this is by improving diversity amongst students entering engineering and computing programmes.[193] Moreover, like ripples from a stone cast into a lake, alterations in one social change component can build momentum of change in others. For example, increased institutional diversity in STEM can tackle stereotyping, targeting the cultural element of social change.

In conclusion, the overarching emphasis on solutions is undoubtedly technical. The focus on technological solutions is expected considering that 'the starting point for most equality regimes (...) is a reactive one, namely the eradication of direct discrimination'.[194] Once the reactive regime matures, they tend to push for substantive equality rather than eradicate the discrimination causing the inequality. However, the theory of change,

[191] ibid. 11.
[192] Foucault (n 20) Part III.
[193] Adam (n 39) 19. Alison Adam identifies these solutions as a liberal feminist position which tends to view technology as a neutral enterprise.
[194] B Fitzpatrick and others, 'Comparative Review of the Law on Equality of Opportunity', in Magill D and Rose S (eds), *Fair Employment Law in Northern Ireland: Debates and Issues* (Belfast, SACHR 1996) 152.

highlights a deficiency in reaching CEDAW's ultimate goal. The solutions targeting discrimination lack variation, and inhibit an atmosphere of "technical fixes". The following section will explore the effects of creating a technical narrative for achieving substantive equality.

The Exclusive Focus on Technical Solutions Risk Reinforcing and Perpetuating Inequality

The following section examines the impact of the technocentric narrative, on substantive equality. Specifically, two harmful effects are lifted concerning the recommended technical solutions: (1) they help reinforce the status quo and (2) they risk perpetuating inequality.

The focus on technical solutions suffers from a significant methodical limitation, mirroring some of the anti-discrimination discourses' most problematic features. Namely, it confines analysis into isolated parts, sustaining inequality. Moreover, this limited understanding of discrimination is poorly equipped to enhance substantive equality, as it fails to consider social dynamics. Like a Band-Aid, these solutions may stop the bleeding, but the harm will continue if the injury is left untreated. From this perspective, identifying AI biases with an attempt to fine-tune them 'becomes an exercise in futility'.[195]

Bias is a social problem, and resolving the issue requires social transformation, according to CEDAW and the theory of social change. Therefore, seeking to solve it within the algorithm is always inadequate. Instead, solutions should address the entire algorithmic life cycle, from the "creation of data" to AI's deployment in society.[196] Only when the causes of discrimination are examined in their proper social context is it possible to understand the interrelations with power structures. Consequently,

[195] West (n 182) 10.
[196] ibid 4.

focusing on technical fixes obscures the root causes for inequality. This maintains inequality, although it does not worsen it. However, using technology to preserve the status quo cannot be considered a neutral choice, but a legally significant one.[197] Scholars need to carefully consider how to better nurture the proactive fulfilment of substantive equality.

Apart from reinforcing inequality, a sole focus on technical solutions can actively increase inequalities. For example, facial recognition is known to misidentify women of colour as they are underrepresented in both the data and amongst the developers of the algorithms.[198] Adding additional data will paint a more accurate picture from which the algorithm can learn. However, the blind conquest to eliminate bias and subsequently "balance" representation may have serious consequences.[199]

Firstly, surveillance tool perpetuate discrimination against 'people of colour and other minorities, women or persons with disabilities'.[200] If not purposeful, the refinement of these instruments can contribute to systematic harm against women. Secondly, the "improvement" of AI entrenches the categorisation of individuals through increased labelling of data, separating people into binary groups of black and white, women and men. This "making" of gender and race within technological systems entrenches harmful norms which CEDAW actively disavows.[201] Accordingly, producing more gender-cognizant AI systems does not necessarily equate to an embrace of diversity. Such an

[197] Wachter (n 37) 774.
[198] See, Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (2018) 81 Proceedings of Machine Learning Research.
[199] Powles (n 176).
[200] The UN Office of the High Commissioner for Human Rights (OHCHR), 'Impact of New Technologies on the Promotion and Protection of Human Rights in the Context of Assemblies, Including Peaceful Protests' (25 June 2020) UN Doc A/HRC/44/24, para 32.
[201] CEDAW, art 5.

adoption of categorisation also enables a monetization of identity 'as market segments for corporate profit'.[202]

Focusing on technical biases asserts considerable pressure on reforming technical tools and diversifying data. However, similar to the limitations of formal equality, the narrow approach redirects focus from other critical questions: Who is producing these systems and why? How are they being used? Moreover, what is at stake when we systematically classify individuals?[203] Asking these questions requires human rights lawyers to step outside of the incomplete technological narrative. Scholars must use human rights law to help ensure that the solutions they recommend are efficient but not harmful. International human rights frameworks, such as CEDAW, provide a robust guide. The convention emphasises the prohibitions, but most importantly, what we should aim towards. Of course, this is no easy task; it becomes particularly difficult when the subject area is transdisciplinary, demanding lawyers to cover both technological and legal facets. However, the gap between law and practice becomes apparent only when these spheres are truly merged. Accordingly, when aligned with CEDAW's goal for substantive equality, it stands clear that the recommended technological solutions fail to fully merge, like oil and water.

In conclusion, CEDAW aims to achieve substantive equality. However, substantive equality necessitates social change, aligning with CEDAW's transformative provisions. Like formal equality, technocentric solutions are one dimension of social change as they help prevent direct and ongoing instances of discrimination. However, an exclusive focus on formal equality is not enough to reach substantive equality and, at worst, risks both reinforcing and perpetuating inequality. Consequently, the legal scholarly focus on technical solutions, locating the discrimination solely within the AI

[202] West (n 182) 19.
[203] Crawford (n 6) 149.

system itself, is not only inadequate in its intervention but can also be actively harmful.[204] Instead, legal scholars must develop the established techno-narrative around AI. They must go beyond the technocentric solutions and include a broader analysis of the technology's interaction with society.

[204] West (n 182) 18.

6. CONCLUSION

The AI frontier presents exciting opportunities to enhance society; however, as the technology encompasses the social fabric, it impacts human rights and gender equality. As required by the descriptive dimension of the Theory of Social Change, the thesis examined the human rights issues by conducting a mapping of the legal literature on AI and gender. The data showed that the literature predominantly focuses on discrimination. Subsequently, scholarly work has started to emerge over the last few years to address the technology's disparate impact on gender equality.

Additionally, the thesis established that scholars perceive AI discrimination as purely technical. They describe AI using the language of mathematics, concealing AI with a sterilised screen of objectivity. This 'epistemological flattening of complexity' masks the core issue by restricting our understanding of AI biases to the machine learning system.[205] Conversely, the thesis has shown that AI is far from a neutral system detached from human subjectivity. On the contrary, cultural power structures shape AI to fit social contexts.

We can only fully grasp AI's impacts when it is understood as a technosocial entity. Armed with this nuance, we can work towards effective solutions that align with CEDAW's primary aim. The thesis' first research question explored the issues and prevailing perspectives discussed within legal literature on AI and gender. Resoundingly, it is evident that the dominant issue is *discrimination*. Unfortunately, the prevailing technological narrative pushes scholars to view discrimination as a purely technical issue rather than incorporate the social context.

[205] Crawford (n 6) 213.

There is much at stake in how legal scholars delineate an issue, as this nuance shapes our perspective, which in turn influences the chosen solutions. The data from the legal literature review establishes that scholars perceive AI as a technical concept. As technical problems require technical solutions, most of the scholars' suggestions are therefore technical. When these technocentric solutions are positioned against CEDAW, the solutions show parallels to formal equality. However, similar to formal equality, technical tools cannot avert discriminatory results by themselves. The biases lie not in the machine learning systems but are created through social processes.

Therefore, positioned together with the theory of social change, the suggested technocentric solutions fail to achieve the underlying social change and fulfil CEDAW's ultimate goal by themselves. Consequently, scholars must move away from the one-dimensional pathway and explore other solutions to tackle AI biases. These new solutions require looking past the machine learning systems. Regrettably, failure to do so may directly impact substantive equality.

The thesis has shown that the recommended technical solutions conceal and perpetuate inequality. Firstly, they conceal it because the solutions fail to acknowledge the true causes of discrimination. Secondly, they perpetuate it as the solutions feed into a narrative that builds upon the categorisation of people—blinding us from asking other important questions.

The thesis' second research question examined how effectively the solutions endorsed CEDAW's goal of greater substantive gender equality. Technology-focused fixes may be part of the solution. However, altering other societal components is also necessary to achieve social change and substantive equality. Moreover, far from aligning with CEDAW's ultimate purpose, failure to grasp the deep-rooted causes of biases only exacerbates the discriminatory outcomes.

In conclusion, as the scholarly focus on AI biases and discrimination increases, the research scope must also expand. Furthermore, AI systems must not be viewed as exclusively technical concepts but as systems shaped by their social context and those who construct them. Merging both AI's technical and social aspects provides a more realistic picture of the issue, helping develop solutions that efficiently mitigate discrimination and align with CEDAW.

Solving the intricacies of biases and discrimination requires more than the current limited technocentric solutions dominating the debates on AI and gender. While it is only human nature to hope for a painless solution, I fear technology is not a silver bullet. The recommended solutions may be one step towards achieving necessary social change, but remaining only within the technological narrative threatens substantive equality.

APPENDIX A- DEVELOPMENT OF SEARCH STRING

Date	Keywords[206]					
	Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**	**Hits**
#1 18/1	Artificial Intelligence (All text)	Gender (All text)	Law (All text)			43,290
#2 18/1	"Artificial Intelligence" OR AI (Abstract)	Gender (Abstract)	Law (Abstract)			37
#3 18/1	"Artificial Intelligence" OR AI (All text)	Gender OR sex (All text)	Law (SO Journal Title/Source)			4,458
#4 19/1	"Artificial Intelligence" OR AI (All text)	Gender OR equality (All text)	Law (SO Journal Title/Source)			7,204
#5 19/1	"Artificial Intelligence" OR AI (All text)	Gender (All text)	Equality (All text)	Law (SO Journal Title/Source)		2,072
#6 20/1	Artificial Intelligence (All text)	AI (All text)	Gender (All text)	Equality (All text)	Law (SO Journal Title/Source)	307
#7 20/1	Artificial Intelligence OR AI (All text)	Gender (All text)	Equality (All text)	Law (SO Journal Title/Source)		399
#8 20/1	Artificial Intelligence OR AI (Selected field optional)	Gender OR Equality OR women's rights (Selected field optional)	Law (SO Journal Title/Source)			49
#9 22/1	Artificial Intelligence OR AI (All text)	Gender (All text)	Equality (All text)	Law OR "Human Rights" (SO Journal Title/Source)		2,735
#10 22/1	"Artificial Intelligence" AND AI (All text)	Gender (All text)	Equality (All text)	Law AND "human rights" (SO Journal Title/Source)		183

[206] All search strings had the delimitation criteria: 'Peer reviewed', 'Academic Journals' and 'English'.

APPENDIX B- DATA COLLECTION

ID	Title	Year	Authors	Journal	Topic (general)	Topic (specific)	Field Focus	Recommended Solutions	Type of Recommended Technical Solution
1	A legal framework for AI training data—from first principles to the Artificial Intelligence Act	2021	Hacker, P.	Law, Innovation & Technology	Discrimination	Regulatory Protection			
2	Adapting Our Anti-Discrimination Laws to Protect Workers' Rights in the Age of Algorithmic Employment Assessments and Evolving Workplace Technology	2021	Yang, J. R.	ABA Journal of Labor & Employment Law	Discrimination	Targeting Specific Fields	Employment	Technical	Improve Dataset
3	Addressing indirect discrimination and gender stereotypes in AI virtual personal assistants: the role of international human rights law	2019	Adams, R. and Nóra Ní Loideáin	Cambridge International Law Journal	Discrimination	Discourse Analysis		Team Diversification	
4	Affinity Profiling and Discrimination by Association in Online Behavioral Advertising	2020	Wachter, S.	Berkeley Technology Law Journal	Discrimination	Targeting Specific Fields	Advertisment	Technical	Algorithmic Transparency
5	AI ethical bias: a case for AI vigilantism (Allantism) in shaping the regulation of AI	2021	Nwafor, I. E.	International Journal of Law & Information Technology	Discrimination	Regulatory Protection		Regulatory	
6	Algorithmic Personalized Pricing	2020	Chapdelaine, P.	New York University Journal of Law and Business	Discrimination	Targeting Specific Fields	Advertisment		
7	Algorithms and Fairness	2021	Foggo, V., Villasenor, J. and Garg, P.	Ohio State Technology Law Journal	Discrimination	Regulatory Protection		Technical	Algorithmic Transparency
8	Algorithms as Allies: Regulating New Technologies in the Fight for Workplace Equality	2019	Heasier, J.	Temple International & Comparative Law Journal	Discrimination	Targeting Specific Fields	Employment		
9	Algorithms as Legal Decisions: Gender Gaps and Canadian Employment Law in the 21st Century	2020	Niblett, A.	University of New Brunswick Law Journal	Discrimination	Targeting Specific Fields	Judical System		
10	Artificial Intelligence and Human Rights. A Challenging Approach on the Issue of Equality	2018	Stănilă, L.	Journal of Eastern European Criminal Law	Discrimination	Targeting Specific Fields	Employment	Technical	Algorithmic Transparency
11	Artificial Intelligence and Human Rights: A Business Ethical Assessment	2020	KRIEBITZ, A. and LÜTGE, C.	Business & Human Rights Journal	Discrimination	Responsability			
12	Artificial Intelligence and the Struggle between Good and Evil	2021	Houser, K. A.	Washburn Law Journal	Discrimination	Mapping of Human Rights Issues		Team Diversification	
13	Artificial Intelligence and the Threat to Human Rights	2020	Humble, K. P. and Altun, D.	Journal of Internet Law	Discrimination	Mapping of Human Rights Issues		Technical	Algorithmic Transparency
14	Artificial Intelligence, Rights and the Virtues	2021	Opderbeck, D. W.	Washburn Law Journal	Legal Personhood				
15	Better decision support through exploratory discrimination-aware data mining: Foundations and empirical evidence	2014	Berendt, B. and Preibusch, S.	Artificial Intelligence and Law	Discrimination	Regulatory Protection		Technical	Improve Dataset
16	Beyond Intent: Establishing Discriminatory Purpose in Algorithmic Risk Assessment	2021	Harvard Law Review	Harvard Law Review	Discrimination	Responsability			
17	Beyond state v loomis: Artificial intelligence, government algorithmization and accountability	2019	Liu, H.-W. (1), Lin, C.-F. (2) and Chen, Y.-J. (3,4)	International Journal of Law and Information Technology	Discrimination	Targeting Specific Fields	Judical System	Technical	Algorithmic Transparency
18	Big Data and Artificial Intelligence: New Challenges for Workplace Equality	2019	Kim, P. T.	University of Louisville Law Review	Discrimination	Targeting Specific Fields	Employment		
19	Will Technological Skill Bias Exacerbate Residual Market Inequalities? Lessons from EU Non-Discrimination Law	2020	Grozdanovski, L.	Labor Law Journal	Discrimination	Mapping of Human Rights Issues			
20	Building Global Algorithmic Accountability Regimes: A Future-focused Human Rights Agenda Beyond Measurement	2021	Gottardo, R.	Peace Human Rights Governance	Discrimination	Regulatory Protection		Technical	Algorithmic Transparency
21	Combating discrimination using Bayesian networks	2014	Mancuhan, K. and Clifton, C.	Artificial Intelligence and Law	Discrimination	Regulatory Protection		Technical	Improve Dataset
22	Constitutional Rights in the Machine-Learning State	2020	Huq, A. Z.	Cornell Law Review	Discrimination	Regulatory Protection			
23	Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression	2020	Oliva, T. D.	Human Rights Law Review	Freedom of Expression				
24	Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness	2020	Abu-Elyounes, D.	Journal of Law, Technology & Policy	Discrimination	Regulatory Protection			
25	Designing intersections—designing subjectivity: Feminist theory and praxis in a sex discrimination legislation system	1995	Adam, A. and Furnival, C.	Information & Communications Technology Law	Discrimination	Discourse Analysis			

26	Discrimination by Design	2019	Cahn, N., Carbone, J. and Levit, N.	Arizona State Law Journal	Discrimination	Mapping of Human Rights Issues			
27	Discrimination through Optimization : How Facebook's Ad Delivery Can Lead to Biased Outcomes	2019	Ali, M. et al.	Proceedings of the ACM on Human-Computer Interaction	Discrimination	Targeting Specific Fields	Advertisment	Technical	Improve Dataset
28	"Equality and Privacy by Design": A New Model of Artificial Intelligence Data Transparency via Auditing, Certification, and Safe Harbor Regimes	2019	Yanisky-Ravid, S. and Hallisey, S. K.	Fordham Urban Law Journal	Discrimination	Regulatory Protection			
29	How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem	2018	Levendowski, A.	Washington Law Review	Discrimination	Regulatory Protection			
30	Introduction: Gender, War, and Technology: Peace and Armed Conflict in the Twenty-First Century	2018	Jones, E., Kendall, S. and Yoriko Otomo	Australian Feminist Law Journal	Discrimination	Targeting Specific Fields	Warfare		
31	Locked out by Big Data: How Big Data Algorithms and Machine Learning May Undermine Housing Justice	2020	Schneider, V.	Columbia Human Rights Law Review	Discrimination	Targeting Specific Fields	Housing and Social Services	Technical	Improve Dataset
32	Meeting the Chimera: How the CEDAW Can Address Digital Discrimination	2021	Krupiy, T.	International Human Rights Law Review	Discrimination	Regulatory Protection		Regulatory	
33	Modelling law using a feminist theoretical perspective	1995	Edwards, L.	Information & Communications Technology Law	Discrimination	Targeting Specific Fields	Judical System		
34	Price Discrimination-Driven Algorithmic Collusion: Platforms for Durable Cartels	2021	Mehra, S. K.	Stanford Journal of Law, Business & Finance	Discrimination	Targeting Specific Fields	Advertisment		
35	Proving Algorithmic Discrimination in Government Decision-Making	2020	Maxwell, J. and Tomlinson, J.	Oxford University Commonwealth Law Journal	Discrimination	Targeting Specific Fields	Judical System		
36	Proxy Discrimination in the Age of Artificial Intelligence and Big Data	2020	Prince, A. E. R. and Schwarcz, D.	Iowa Law Review	Discrimination	Regulatory Protection			
37	Putting Human Values into the Machine	2018	Santow, E.	Human Rights Defender	Discrimination	Mapping of Human Rights Issues			
38	Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age	2018	Schider, C. A.	Denver Law Review	Discrimination	Regulatory Protection			
39	Screened out Onscreen: Disability Discrimination, Hiring Bias, and Artificial Intelligence	2021	Moss, H.	Denver Law Review	Discrimination	Targeting Specific Fields	Employment	Technical	Improve Dataset
40	Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare	2020	Davide Cirillo et al.	npj Digital Medicine	Discrimination	Targeting Specific Fields	Healthcare	Technical	Algorithmic Transparency
41	Sex Causation, and Algorithms: How Equal Protection Prohibits Compounding Prior Injustice	2020	Hellman, D.	Washington University Law Review	Discrimination	Regulatory Protection			
42	Show Us the Data: Privacy, Explainability, and Why the Law Can't Have Both	2020	Grant, T. D. and Wischik, D. J.	George Washington Law Review	Discrimination	Regulatory Protection			
43	Strengthening legal protection against discrimination by algorithms and artificial intelligence	2020	Zuiderveen Borgesius, F. J.	International Journal of Human Rights	Discrimination	Regulatory Protection		Regulatory	
44	Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under Eu Law	2018	Hacker, P.	Common Market Law Review	Discrimination	Regulatory Protection			
45	The Algorithmic Divide and Equality in the Age of Artificial Intelligence	2020	Yu, P. K.	Florida Law Review	Discrimination	Mapping of Human Rights Issues		Technical	Improve Dataset
46	The Gender Panopticon: AI, Gender, and Design Justice	2021	Katyal, S. K. and Jung, J. Y.	UCLA Law Review	Discrimination	LGBTQ+ community			
47	The Paradox of Automation as Anti-Bias Intervention	2020	Ajunwa, I.	Cardozo Law Review	Discrimination	Targeting Specific Fields	Employment		
48	Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity	2021	Chesterman, S.	American Journal of Comparative Law	Discrimination	Targeting Specific Fields	Judical System		
49	Unexpected Inequality: Disparate-Impact from Artificial Intelligence in Healthcare Decisions	2021	Takshi, S.	Journal of Law and Health	Discrimination	Targeting Specific Fields	Healthcare		
50	Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy	2018	Mateen, H.	Berkeley Journal of Employment and Labor Law	Discrimination	Targeting Specific Fields	Employment		
51	Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience	2020	Xenidis, R.	Maastricht Journal of European and Comparative Law	Discrimination	Regulatory Protection			
52	Private Accountability in the Age of Artificial Intelligence	2019	Katyal, S. K.	UCLA Law Review	Discrimination	Responsability		Corporate Governance and Utilisation	
53	Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices	2018	McKenzie Raub	Arkansas Law Review	Discrimination	Targeting Specific Fields	Employment	Team Diversification	
54	Targeting, gender, and international posthumanitarian law and practice: Framing the question of the human in international humanitarian law	2018	Arvidsson, M	Australian Feminist Law Journal	Discrimination	Targeting Specific Fields	Warfare		
55	Employing AI	2018	Sullivan, C. A	Villanova Law Review,	Discrimination	Regulatory Protection			

56	AI's Transformational Role in Making HR More Objective while Overcoming the Challenge of Illegal Algorithm Biases	2018	Garry Mathiason	The Journal of Robotics, Artificial Intelligence & Law	Discrimination	Targeting Specific Fields	Employment		
57	Algorithmic Discrimination Is an Information Problem	2019	Ignacio N. Cofone	Hastings Law Journal	Discrimination	Regulatory Protection		Technical	Improve Dataset
58	Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non-Discrimination Law	2021	Sandra Wachter, Brent Mittelstadt & Chris Russell	West Virginia Law Review	Discrimination	Discourse Analysis			
59	"Let the Algorithm Decide": Is Human Dignity at Stake?	2021	Marcela Mattiuzzo	Brazilian Journal of law Public Policy	Discrimination	Responsability			
60	Assessing the Legal Personality of Sexbots in the Light of Human Rights	2020	Arushi Gupta	Kathmandu School of Law Review	Legal Personhood				
61	Silencing women in the digital age	2019	Arimatsu, L.	Cambridge International Law Journal	Discrimination	Discourse Analysis			
62	Human Rights and Digital Health Technologies	2020	Nina Sun, Kenechukwu Esom, Mandeep Dhaliwal & Joseph J. Amon,	Health & Human Rights Journal	Discrimination	Targeting Specific Fields	Healthcare		
63	Artificial Intelligence as an Instrument of Discrimination in Workforce Recruitment	2019	Alessandro Miasato & Fabiana Reis Silva	Acta Univ Sapientiae: Legal Studies	Discrimination	Targeting Specific Fields	Employment		
64	Sophia: The Intersection of Artificial Intelligence and Human Rights	2020	Maggie Redden	Journal Global Rights & Org	Legal Personhood				
65	Connectivity: The Global Gender Digital Divide and Its Implications for Women's Human Rights and Equality	2019	Mary Pat Treuthart	Gonzaga Jorunal of International Law	Discrimination	Mapping of Human Rights Issues			
66	Feminist perspectives to artificial intelligence: Comparing the policy frames of the European Union and Spain	2021	Guevara-Gómez, A.	The International Journal of Government & Democracy in the Information Age	Discrimination	Discourse Analysis			
67	Gender as Emotive AI and the Case of 'Nadia': Regulatory and Ethical Implications	2021	Ni Loideain Nora and Adams, Rachel and Clifford, Damian	Computer Law & Security Review	Discrimination	Discourse Analysis			
68	Auditing Algorithms for Discrimination	2017	Pauline T. Kim	Univeristy of Penssylvania Law Review	Discrimination			Technical	Algorithmic Transparency
69	Discrimination in the Age of Algorithms	2018	Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, Cass R Sunstein	Journal of Legal Analysis	Discrimination	Mapping of Human Rights Issues			
70	Trust but Verify: A Guide to Algorithms and the Law	2017	Desai, D. R. and Kroll, J. A.	Harvard Journal of Law & Technology	Discrimination	Regulatory Protection			
71	The fundamental rights challenges of algorithms	2019	Gerards, J.	Netherlands Quarterly of Human Rights	Discrimination	Mapping of Human Rights Issues			
72	Accountable Algorithms	2017	Joshua A. Kroll , Joanna Huey , Solon Barocas , Edward W. Felten , Joel R. Reidenberg , David G. Robinson & Harlan Yu	Univeristy of Penssylvania Law Review	Discrimination	Responsability		Technical	Improve Dataset

BIBLIOGRAPHY

BOOKS

Adam A, *Artificial Knowing: Gender and the Thinking Machine* (Routledge 2006)

Aggarwal CC, *Neural Networks and Deep Learning: A textbook* (Springer International Publishing 2018)

Aristotle, *Ethica Nicomachea* (WD Ross trans., Oxford University Press 1925)

Arnadóttir OM, *Equality and Non-Discrimination under the European Convention on Human Rights* (Kluwer Law International 2003)

Boudon R, *Theories of Social Change: A Critical Appraisal* (JC Whitehouse tr, Polity Press 1986)

Bowker GC and Star SL, *Sorting Things out: Classification and Its Consequences* (MIT Press 2000)

Castree N and other, *A Dictionary of Human Geography* (Oxford University Press 2013)

Crawford K, *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021)

D'lgnazio C and Klien LF, *Data Feminism* (MIT Press 2020)

Domingos P, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Out World* (Penguin Books Ltd 2015)

Elliott A (ed.), *The Routledge Social Science Handbook of AI* (Routledge 2021)

Engels F, *Ludwig Feuerbach and the Outcome of Classical German Philosophy* (International Publishers 1941)

Fitzpatrick B and others, 'Comparative Review of the Law on Equality of Opportunity', in Magill D and Rose S (eds.), *Fair Employment Law in Northern Ireland: Debates and Issues* (Belfast, SACHR 1996)

Flingstein N and McAdam D, *A Theory of Fields* (Oxford University Press 2012)

Foucault M, *Discipline and Punish: The Birth of the Prison* (Penguin Books 2019)

Fredman S, 'Beyond the Dichotomy of Formal and Substantive Equality: Towards a New Definition of Equal Rights' in Ineke Boerefin et al (eds), *Temporary Special Measures: Accelerating De Facto Equality of Women under Article 4(1) UN Convention on the Elimination of All Forms of Discrimination against Women* (Intersentia 2003)

Freeman MA, Chinkin C and Rudolf B (eds), *The UN Convention on the Elimination of all Forms of Discrimination Against Women: A Commentary* (Oxford University Press 2012)

Gillin J.L. and Gillin J.P, *Cultural Sociology* (Macmillan Company 1954)

Goodfellow I, Bengio Y and Courville A, *Deep Learning* (MIT Press 2016)

Hacking I, *The Social Construction of What?* (Harvard University Press 1999)

Harvard Business Review Press and others*, Artificial Intelligence: The Insights You Need from Harvard Business Review* (Harvard Business Review Press 2019)

Kaplan J, *Artificial Intelligence: What Everyone Needs to Know* (Oxford University Press 2016)

Marx K, *The Poverty of Philosophy* (Harry Quelch tr, Cosimo Classics 2008)

Moeckli D and others (eds.), *International Human Rights Law* (3rd edn, Oxford University Press 2018)

Neapolitan RE and Jiang X, *Contemporary Artificial Intelligence* (Taylor & Francis Group 2013)

O'Neil C, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books 2017)

Perez CC, *Invisible Women: Data Bias in a World Designed for Men* (Abrams Press 2019)

Snel M and Des Moraes J (eds.), *Doing a Systematic Literature Review in Legal Scholarship* (Eleven International Publishing 2018)

Somogyi Z, *The Application of Artificial Intelligence: Step-by-Step Guide from Beginner to Expert* (Springer Nature 2022)

Wishmeyer T and Rademacher T (eds.), *Regulating Artificial Intelligence* (Springer Nature 2020)

Zhou J and Chen F, *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent* (Springer 2018)

ACADEMIC JOURNALS

Abu-Elyounes D, 'Contextual Fairness: A Legal and Policy Analysis of Algorithmic Fairness' (2020) 20 Journal of Law, Technology & Policy 1

Adam A and Furnival C, 'Designing intersections—designing subjectivity: Feminist theory and praxis in a sex discrimination legislation system' (1995) 4 Information & Communications Technology Law 161

Adams R and Loideáin NN, 'Addressing indirect discrimination and gender stereotypes in AI virtual personal assistants: the role of international human rights law' (2019) 8 Cambridge International Law Journal 241

Ajunwa I, ‘The Paradox of Automation as Anti-Bias Intervention’ (2020) 41Cardozo Law Review 1671

Ali M and others, ‘Discrimination through Optimization: How Facebook’s Ad Delivery Can Lead to Biased Outcomes’ (2019) 3 CSCW 1

“Anonymous”, ‘Beyond Intent: Establishing Discriminatory Purpose in Algorithmic Risk Assessment’ (2021) 134 Harvard Law Review 1760

Arimatsu L, ‘Silencing women in the digital age’ (2019) 8 Cambridge International Law Journal 187

Arvidsson M, ‘Targeting, Gender, and International Posthumanitarian Law and Practice: Framing the Question of the Human in International Humanitarian Law’ (2018) 44 Australian Feminist Law Journal 9

Berendt B and Preibusch S, ‘Better decision support through exploratory discrimination-aware data mining: Foundations and empirical evidence’ (2014) 22 Artificial Intelligence and Law 175

Bhandarkar T and others, ‘Earthquake Trend Prediction Using Long Short-term Memory Neural Networks’ (2019) 9 International Journal of Electrical and Computer Engineering 1304

Bimber B, ‘Karl Marx and the Three Faces of Technological Determinism’ (1990) 20 Social Studies of Science 333

Boyd D and Crawford K, ‘Critical Questions for Big Data’ (2012) 15 Information, Communication & Society 662

Buolamwini J and Gebru T, ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’ (2018) 81 Proceedings of Machine Learning Research 1

Cahn N and others, ‘Discrimination by Design’ (2019) 51Arizona State Law Journal 1

Caliskan A, Bryson JJ and Narayanan A, ‘Semantics Derived Automatically from Language Corpora Contain Human-like Biases’ (2017) 356 Science 183

Campbell M, Hoane AJ and Hsu FH, 'Deep Blue' (2002) 135 Artificial Intelligence 135

Chapdelaine P, 'Algorithmic Personalized Pricing' (2020) 17 New York University Journal of Law and Business 1

Chesterman S, 'Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity' (2021) 69 American Journal of Comparative Law 271

Cofone IN, 'Algorithmic Discrimination Is an Information Problem' (2019) 70 Hastings Law Journal 1389

Cui H and others, 'Identifying the Key Reference of a Scientific Publication' (2020) 29 Journal of Systems Science and Systems Engineering 429

Cusack S and Pusey L, 'CEDAW and the Rights to Non-discrimination and Equality' (2013) 14 Melbourne Journal of International Law 54

Daston L, 'Cloud Physiognomy' (2016) 135 University of California Press 45

Davide Cirillo and others, 'Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare' (2020) 3 HR Digital Medicine 1

Desai DR and Kroll JA, 'Trust but Verify: A Guide to Algorithms and the Law' (2017) 1 Harvard Journal of Law & Technology 1

Edwards L, 'Modelling law using a feminist theoretical perspective' (1995) 4 Information & Communications Technology Law 95

Faulkner W, 'The Technology Question in Feminism: A View from Feminist Technology Studies' (2001) 24 Women's Studies International Forum 79

Ferrer X and others, 'Bias and Discrimination in AI: A Cross-Disciplinary Perspective' (2021) 40 IEEE Technology and Society Magazine 1

Foggo V and others, ‘Algorithms and Fairness’ (2021) 17 Ohio State Technology Law Journal 123

Gerards J, ‘The fundamental rights challenges of algorithms’ (2019) 37 Netherlands Quarterly of Human Rights 205

Gijs Van Dijck, ‘Legal Research When Relying on Open Access: A primer’ (2016) 6 Law and Method 1

Gottardo R, ‘Building Global Algorithmic Accountability Regimes: A Future-focused Human Rights Agenda Beyond Measurement’ (2021) 5 Peace Human Rights Governance 65

Grant TD and Wischik DJ, ‘Show Us the Data: Privacy, Explainability, and Why the Law Can’t Have Both’ (2020) 88 George Washington Law Review 1350

Grozdanovski L, ‘Will Technological Skill Bias Exacerbate Residual Market Inequalities? Lessons from EU Non-Discrimination Law’ (2020) 71 Labor Law Journal 58

Guevara-Gómez A and others, ‘Feminist perspectives to artificial intelligence: Comparing the policy frames of the European Union and Spain’ (2021) 26 International Journal of Government & Democracy in the Information Age 173

Gupta AD, ‘Assessing the Legal Personality of Sexbots in the Light of Human Rights’ (2020) 8 Kathmandu School of Law Review 90

——, ‘Change, Development and a Theory of Social Science’ (1989) 24 Economic and Political Weekly 35

Hacker P, ‘A Legal Framework for AI Training Data’ (2020) 13 Law, Innovation & Technology 257

——, ‘Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law’ (2018) 55 Common Market Law Review 1143

Hellman D, ‘Sex Causation, and Algorithms: How Equal Protection Prohibits Compounding Prior Injustice’ (2020) 98 Washington University Law Review 481

Hendrix BA, ‘Where Should We Expect Social Change in Non-Ideal Theory?’ (2013) 41 Political Theory 116

Hensler J, ‘Algorithms as Allies: Regulating New Technologies in the Fight for Workplace Equality’ (2019) 34 Temple International & Comparative Law Journal 31

Houser K, ‘Artificial Intelligence and the Struggle between Good and Evil’ (2021) 60 Washburn Law Journal 475

Humble KP and Altun D, ‘Artificial Intelligence and the Threat to Human Rights’ (2020) 24 Journal of Internet Law 1

Huq AZ, ‘Constitutional Rights in the Machine-Learning State’ (2020) 105 Cornell Law Review 1875

Ignacio NC, 'Algorithmic Discrimination Is an Information Problem' (2019) 70 Hastings Law Journal 1389

Johnston SF, ‘The Technological Fix as Social Cure-All: Origins and Implication’ (2018) 37 IEEE Technology and Society Magazine 47

Jones E and others, ‘Introduction: Gender, War, and Technology: Peace and Armed Conflict in the Twenty-First Century’ (2018) 44 Australian Feminist Law Journal 1

Katyal SL, ‘Private Accountability in the Age of Artificial Intelligence’ (2019) 66 UCLA Law Review 54

Katyal SK and Jung JY, ‘The Gender Panopticon: AI, Gender, and Design Justice’ (2021) 68 UCLA Law Review 692

Keyes OS, ‘The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition’ (2018) 2 Association for Computing Machinery 1

Kim PT, ‘Auditing Algorithms for Discrimination’ (2017) 166 University of Pennsylvania Law Review 189

——, ‘Big Data and Artificial Intelligence: New Challenges for Workplace Equality’ (2019) 57 University of Louisville Law Review 313

Klein HK and Kleinman DL, ‘The Social Construction of Technology: Structural Considerations’ (2002) 27 Science, Technology, & Human Values 28

Kleinberg J and others, ‘Discrimination in the Age of Algorithms’ (2018) 10 Journal of Legal Analysis 113

Kranzberg M, ‘Technology and History: “Kranzberg’s Laws”’ (1986) 27 Technology and Culture 544

Kriebitz A and Lütge C, ‘Artificial Intelligence and Human Rights: A Business Ethical Assessment’ (2020) 5 Business & Human Rights Journal 84

Kroll JA and others, ‘Accountable Algorithms’ (2017) 165 University of Pennsylvania Law Review 633

Krupiy T, ‘Meeting the Chimera: How the CEDAW Can Address Digital Discrimination’ (2021) 10 International Human Rights Law Review 1

——, ‘Meeting the Chimera: How the CEDAW Can Address Digital Discrimination’ (2021) 10 International Human Rights Law Review 1

Levendowski A, ‘How Copyright Law Can Fix Artificial Intelligence’s Implicit Bias Problem’ (2018) 93 Washington Law Review 579

Liu HW and others, ‘Beyond state v Loomis: Artificial intelligence, government algorithmization and accountability’ (2019) 27 International Journal of Law and Information Technology 122

Loideain N and others, ‘From Alexa to Siri and the GDPR: The Virtual Personal Assistants and the Role of Data Protection Impact Assessments’ (2020) 36 Computer Law & Security Review 1

Mancuhan K and Clifton C, ‘Combating discrimination using Bayesian networks’ (2014) 22 Artificial Intelligence and Law 211

Mateen H, ‘Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy’ (2018) 39 Berkeley Journal of Employment and Labor Law 285

Mathiason G, ‘AI's Transformational Role in Making HR More Objective while Overcoming the Challenge of Illegal Algorithm Biases’ (2018) 1 The Journal of Robotics, Artificial Intelligence & Law 1

Mattiuzzo M, ‘Let the Algorithm Decide": Is Human Dignity at Stake?’ (2021) 11 Brazilian Journal of law Public Policy 343

Maxwell J and Tomlinson J, ‘Proving Algorithmic Discrimination in Government Decision-Making’ (2020) 20 Oxford University Commonwealth Law Journal 352

McKenzie R, 'Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices' (2018) 71 Ark L Rev 529

Mehra SK, ‘Price Discrimination-Driven Algorithmic Collusion: Platforms for Durable Cartels’ (2021) 26 Stanford Journal of Law, Business & Finance 171

Miasato A and Silva FR, ‘Artificial Intelligence as an Instrument of Discrimination in Workforce Recruitment’ (2019) 8 Acta University Sapientiae Legal Studies 191

Mohamed S and others, ‘Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence’ (2020) 405 Philosophy and Technology 1

Moore WE, ‘A Reconsideration of Theories of Social Change’ (1960) 25 American Sociological Review 810

Moss H, ‘Screened out Onscreen: Disability Discrimination, Hiring Bias, and Artificial Intelligence’ (2021) 98 Denver Law Review 775

Niblett A, ‘Algorithms as Legal Decisions: Gender Gaps and Canadian Employment Law in the 21st Century’ (2020) 71 University of New Brunswick Law Journal 112

Nwafor IE, ‘AI ethical bias: a case for AI vigilantism (AIlantism) in shaping the regulation of AI’ (2021) 29 International Journal of Law & Information Technology 225

Oliva TD, ‘Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression’ (2020) 20 Human Rights Law Review 607

Opderbeck DW, ‘Artificial Intelligence, Rights and the Virtues’ (2021) 60 Washburn Law Journal 445

Prince AER and Schwarcz D, ‘Proxy Discrimination in the Age of Artificial Intelligence and Big Data’ (2020) 105 Iowa Law Review 1257

Raub M, ‘Bots, Bias and Big Data: Artificial Intelligence, Algorithmic Bias and Disparate Impact Liability in Hiring Practices’ (2018) 71 Ark. L. Rev. 529

Redden M, ‘Sophia: The Intersection of Artificial Intelligence and Human Rights’ (2020) 10 Journal Global Rights & Org 155

Sandra Fredman, ‘Substantive Equality Revisited’ (2016) 14 ICON 712

Santow E, ‘Putting Human Values into the Machine’ (2018) 27 Human Rights Defender 13

Schider CA, ‘Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age’ (2018) 96 Denver Law Review 87

Schneider V, ‘Locked out by Big Data: How Big Data Algorithms and Machine Learning May Undermine Housing Justice’ (2020) 52 Columbia Human Rights Law Review 251

Searle JR, ‘Minds, Brains and Programs’ (1980) 3 Behavioural and Brain Sciences 417

Selena S and Kenney M, ‘Algorithms, Platforms, and Ethnic Bias: An Integrative Essay’ (2018) 55 The Clark Atlanta University Review of Race and Culture 9

Stănilă L, ‘Artificial Intelligence and Human Rights. A Challenging Approach on the Issue of Equality’ (2018) 2 Journal of Eastern European Criminal Law 19

Sullivan CA, ‘Employing AI’ (2018) 63 Villanova Law Review 395

Sun N and others, ‘Human Rights and Digital Health Technologies’ (2020) 22 Health & Human Rights Journal 21

Surden H, ‘Machine Learning and Law’ (2014) 89 Washington Law Review 87

Takshi S, ‘Unexpected Inequality: Disparate-Impact from Artificial Intelligence in Healthcare Decisions’ (2021) 34 Journal of Law and Health 215

Treuthart MP, ‘Connectivity: The Global Gender Digital Divide and Its Implications for Women's Human Rights and Equality’ (2019) 23 Gonz J Int'l L 1

Wachter S and others, ‘Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non-Discrimination Law’ (2021) 123 West Virginia Law Review 735

——, ‘Affinity Profiling and Discrimination by Association in Online Behavioral Advertising’ (2020) 35 Berkeley Technology Law Journal 367

Xenidis R, ‘Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience’ (2020) 27 Maastricht Journal of European and Comparative Law 736

Yang JR, ‘“Adapting Our Anti-Discrimination Laws to Protect Workers” Rights in the Age of Algorithmic Employment Assessments and

Evolving Workplace Technology' (2021) 35 ABA Journal of Labor & Employment Law 207

Yanisky-Ravid S and Hallisey SK, '"Equality and Privacy by Design": A New Model of Artificial Intelligence Data Transparency via Auditing, Certification, and Safe Harbor Regimes' (2019) 46 Fordham Urban Law Journal 428

Yu PK, 'The Algorithmic Divide and Equality in the Age of Artificial Intelligence' (2020) 72 Florida Law Review 331

Zuiderveen Borgesius FJ, 'Strengthening legal protection against discrimination by algorithms and artificial intelligence' (2020) 24 International Journal of Human Rights 1527

CONFERENCE PAPERS AND REPORTS

ARTICLE 19, 'Privacy and Freedom of Expression in the Age of Artificial Intelligence' (April 2018)

Danks D and London AJ, 'Algorithmic Bias in Autonomous Systems' (Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017)

Holtmaat R, 'European Women and the CEDAW-Convention; the way Forward' (EWLA Conference, 2002)

Korsvik TR and others, 'Artificial Intelligence and Gender Equality: A review of Norwegian Research' (Kilden, December 2020)

Smith G and Rustagi I, 'Mitigating Bias in Artificial Intelligence: An equity Fluent Leadership Playbook' (Berkeley Hass, July 2020)

West SM and others, 'Discriminating Systems: Gender, Race, and Power in AI' (AI Now Institute, April 2019)

EU DOCUMENTS

European Commission, 'Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (21 April 2021) COM/2021/206 final

——, ‘A Definition of AI: Main Capabilities and Scientific Disciplines’ (8 April 2019) <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> accessed 11 February 2022

Special Committee on Artificial Intelligence in a Digital Age, ‘Draft report on artificial intelligence in a digital age’ (2 November 2021) 2020/2266/(INI)

UNITED NATIONS DOCUMENTS

Committee for the Elimination of All Forms of Discrimination against Women (CEDAW), ‘General Recommendation No 28’ (16 December 2010) CEDAW/C/GC/28

——, ‘Concluding Observations: Serbia’ (14 May-1 June 2007) CEDAW/C/SCG/CO/1

——, ‘General Recommendation No 25’ (2004)

Committee on the Rights of Persons with Disabilities, ‘General Recommendation No. 6’ (2018) CRPD/C/GC/6

Human Rights Committee (HRC), ‘General Comment No. 18’ (10 November 1989) UN Doc HRI/GEN/1/Rev.5

Office of the High Commissioner for Human Rights, ‘Impact of New Technologies on the Promotion and Protection of Human Rights in the Context of Assemblies, Including Peaceful Protests’ (25 June 2020) UN Doc A/HRC/44/24

——, ‘General Comment No. 28 on Article 3’ (29 March 2000) CCPR/C/21/Rev. 1/Add.10

NEWS AND MEDIA

Belluck P, ‘Hey Siri, Can I Rely on You in a Crises? Not Always, a Study Finds’ *New York Times* (14 March, 2016)
< https://well.blogs.nytimes.com/2016/03/14/hey-siri-can-i-rely-on-you-in-a-crisis-not-always-a-study-finds/?mtrref=undefined&gwh=> accessed 16 February 2022

Clarke L, 'The EU's leaked AI regulation is ambitious but vague' *Tech Monitor* (15 April, 2021) https://techmonitor.ai/policy/eu-ai-regulation-machine-learning-european-union accessed 11 February 2022

Corbyn Z, 'Catherine D'lgnazio: "Data is never a raw, truthful input-and it is never neutral"' *The Guardian* (March 21, 2020) <https://www.theguardian.com/technology/2020/mar/21/catherine-dignazio-data-is-never-a-raw-truthful-input-and-it-is-never-neutral> accessed 17 February 2022

Crawford K, 'Artificial Intelligence's White Guy Problem' *The New York Times* (June 25, 2016) <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> accessed 1 February 2022

D'lgnazio C, '5 Questions on Data & Justice with Cathy O'Neil' *Medium* (26 November, 2017) <https://medium.com/data-feminism/5-questions-on-data-justice-with-cathy-oneil-87f42355ce55> accessed 4 April 2022

Das S, 'It's Hysteria not a Heart Attack, GP App Tells Women' *The Sunday Times* (13 October, 2019) <*https://www.thetimes.co.uk/article/its-hysteria-not-a-heart-attack-gp-app-tells-women-gm2vxbrqk*> accessed 16 February 2022

Dastin J, 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women' *Reuters* (October 10, 2018) < https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> accessed 8 February 2022

Vincent J, 'Transgender Youtubers had their videos grabbed to train facial recognition software' *The Verge* (22 August, 2017) <https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset> accessed 20 February 2022

WEBSITES

Contini F, 'Artificial Intelligence: A New Trojan Horse for Undue Influence on Judiciaries?' (*UNODC*) <https://www.unodc.org/dohadeclaration/en/news/2019/06/artificial-intelligence_-a-new-trojan-horse-for-undue-influence-on-judiciaries.html> accessed 16 February 2022

Danish Institute for Human Rights, 'HRS Concept Note: Academia' (May 2018) <https://www.humanrights.dk/sites/humanrights.dk/files/media/migrated/hrs_toolbox_concept_note_academia_may2018.pdf> accessed 9 February 2022

ImageNet (March 11 2021) < https://www.image-net.org/> accessed 20 February 2022

Malmbers Å, 'Så drabbades samerna av den rasbiologiska forskningen' (*Uppsala Universitet,* 8 December 2021*)* <*https://www.uu.se/nyheter/artikel/?id=17896&typ=artikel&lang=sv* > accessed 18 February 2022

Nurfikri F, 'An Illustrated Guide to Artificial Neural Networks' *(Towards Data Science*, 20 July 2020*)* < https://towardsdatascience.com/an-illustrated-guide-to-artificial-neural-networks-f149a549ba74> accessed 20 February 2022

Powles J, 'The Seductive Diversion of "Solving" Bias in Artificial Intelligence' (*OneZero*, 7 December 2018) <https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53> accessed 15 May 2022

Samuel S, 'Some AI Just Shouldn't Exist' (*Vox*, 19 April 2019) <https://www.vox.com/future-perfect/2019/4/19/18412674/ai-bias-facial-recognition-black-gay-transgender> accessed 20 February 2022

United Nations, 'Urgent action needed over artificial intelligence risks to human rights' (*UN News,* 15 September 2021) <https://news.un.org/en/story/2021/09/1099972> accessed 9 February 2022

TREATIES

Charter of the United Nations (adopted 26 June 1945) 1 UNTS XVI

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights) (adopted 4 November 1950, entered into force 3 September 1953) ETS 5; 213 UNTS 222

Convention on the Elimination of All Forms of Discrimination against Women (adopted 18 December 1979, entered into force 3 September 1981) 1240 UNTS 13

International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171

International Covenant on Economic, Social and Cultural Rights (adopted 16 December 1966, entered into force 3 January 1976) 993 UNTS

Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A) (III)

Vienna Declaration and Programme of Action (adopted by the World Conference on Human Rights on 25 June 1993) (A/CONF.157/24)

TABLE OF CASES

EUROPEAN COURT OF HUMAN RIGHTS

Biao v. Denmark App no 38590/10 (ECtHR, 24 May 2016)

UN HUMAN RIGHTS TREATY BODIES

Timor-Leste CEDAW/C/TLS/CO/1 (CEDAW, 7 August 2009)