Impact of Artificial Intelligence on Women's Human Rights: An International Legal Analysis

DIPLOMA THESIS

to be awarded the degree of

Magistra iuris (Mag.ª)

in Law
at the Karl-Franzens-University of Graz, Austria

submitted by
Stephanie Grasser, BA MA

at the Institute of Law

Academic supervision: Ao.Univ.-Prof. Mag. Dr.iur. Gerd Oberleitner

Graz, 2024

ABSTRACT

This thesis examines how bias in artificial intelligence (AI) impacts international women's rights. Artificial intelligence has become an indispensable part of everyday life. Whether people notice it or not, it is ubiquitous. Many everyday activities are influenced by AI, from text translations and image generation to search engine algorithms, social media algorithms, and much more.

As widespread as AI has become, so too has the criticism it faces. The technology is criticised as unfair and discriminatory. Numerous examples of AI-influenced unfair treatment, driven by factors such as sexism and racism, have been documented.

For a long time, approaches to designing fair and non-discriminatory AI have been considered primarily from an ethical perspective. Consequently, there are many ethical frameworks providing guidelines for achieving fairness in this technology. This paper examines whether these approaches are effective. Furthermore, the study addresses the question of whether a legal solution to AI-related issues is needed, particularly from the perspective of international human rights. To this end, existing legal foundations are analysed, and the extent to which further legal instruments are necessary to effectively address the problem of discrimination due to bias in artificial intelligence is discussed.

ZUSAMMENFASSUNG

Das Ziel dieser Arbeit ist es, die Auswirkungen von Bias in der Künstlichen Intelligenz auf internationale Frauenrechte zu untersuchen. Künstliche Intelligenz ist aus unserem Alltag nicht mehr wegzudenken. Ob wir es bemerken oder nicht: Sie ist allgegenwärtig. Viele unserer täglichen Aktivitäten werden durch Künstliche Intelligenz beeinflusst, sei es bei Textübersetzungen, der Bilderzeugung, in Suchmaschinenalgorithmen, auf Social-Media-Plattformen und vielem mehr.

So weit Künstliche Intelligenz mittlerweile ist, so umfangreich ist auch die Kritik an ihr. Künstlicher Intelligenz wird vorgeworfen, unfair zu sein und diskriminierend zu handeln. Zahlreiche Beispiele für unfaire Behandlungen, die auf unterschiedliche Gründe wie Sexismus und Rassismus zurückzuführen sind, sind dokumentiert.

Die Ansätze zur Gestaltung einer fairen und nichtdiskriminierenden Künstlichen Intelligenz wurden lange Zeit vor allem aus einer ethischen Perspektive betrachtet. Daher existieren viele ethische Rahmenbedingungen, die Leitlinien für eine faire Nutzung dieser Technologie bieten sollen. Ob diese Ansätze jedoch zielführend sind, wird in dieser Arbeit untersucht. Darüber hinaus wird die Frage behandelt, ob nicht zusätzlich eine rechtliche Lösung notwendig ist, insbesondere aus der Perspektive der internationalen Menschenrechte. Dafür werden die bestehenden rechtlichen Grundlagen analysiert und abschließend diskutiert, inwieweit weitere rechtliche Instrumente notwendig sind, um das Problem der Diskriminierung durch Bias in Künstlicher Intelligenz wirksam anzugehen.

ACKNOWLEDGMENT

Voicing Erasure – A Spoken Word Piece Exploring Bias in Voice Recognition Technology by Joy Buolamwini

Whose voice do you hear…
when you think of intelligence, innovation, and ideas that shape our worlds?
Whose voices are dismissed, diminished and erased,
when we hear the stories of glories past and present
discoveries far and near?
Whose voices rose up to demand to be counted and heard
to be respected and remembered for revealing contradictions in tales of mythical equality,
supposed superiority, and fleeting fairness?
Whose voices fought so that those at the margins and intersections could be free to develop
their minds, advance our humanity and uncover our buried abilities?
And yet, despite the strides, the battles continue.
Far too often, we still face erasure…
Erasure from both humans and machines.
Machines that don't hear the way my sisters, brothers, and siblings speak.
Machines that erase my mother's medical needs, my partner's job opportunities.
And machines that erase you and me.
Machines of flesh and blood networked together that cancel our contributions
and our full expression as individuals.
Machines of silicon and steel that reflect the biases of their makers and societies.
Machines that listen for commands, using names and voices
that reinforce the role of women as subservient recipients of demands.
Alexa? Siri? Cortana?
Are you listening?
Yes.
Do as I say

Answer my questions in a pleasing way.
Is it Okay Google and others to capture data shared unaware,
snatching snippets of intimate whispers?
No
We need to remember we have a voice and a choice.
We do not have to accept conditions that continue traditions of silencing.
We must reject terms that reduce humans to data that fuels surveillance.
We cannot let the promises of AI overshadow real and present harms.
We will not be dismissed.
We will not be erased.
Instead, we will beat the drum of solidarity marching towards a future
where technology serves all of us
not just the privileged few.
Let my voice
Let your voice
Be Heard[1]

[1] Joy Buolamwini, 'Voicing Erasure – A Spoken Word Piece Exploring Bias in Voice Recognition Technology' *Algorithmic Justice League* <ajl.org/voicing-erasure>.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADM	Automated Decision-Making
AGI	Artificial General Intelligence
AI	Artificial Intelligence
ASI	Artificial Super Intelligence
Art	Article
CEDAW	Convention on the Elimination of All Forms of Discrimination Against Women
CEO	Chief Executive Officer
CV	Curriculum Vitae
ECHR	European Convention on Human Rights
e.g.	example gratia
ENNHRI	European Network of National Human Rights Institutions
EU	European Union
FOB	Facebook Oversight Board
GDPR	General Data Protection Regulation
IBM	International Business Machines Corporation
ICCPR	International Covenant on Civil and Political Rights
ICESCR	International Covenant on Economic, Social and Cultural Rights
ICT	Information and Communications Technology
i.e.	id est
IEEE	Institute of Electrical and Electronics Engineers
IT	Internet Technology
Lit	litera
MDG	Millennium Development Goals

MIT	Massachusetts Institute of Technology
ML	Machine Learning
NIST	National Institute of Standards and Technology
NPO	Non Profit Organisation
OECD	Organization for Economic Cooperation and Development
OHCHR	Office of the United Nations High Commissioner for Human Rights
SDG	Sustainable Development Goals
STEM	Science, technology, engineering, and mathematics
UDHR	Universal Declaration of Human Rights
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNGA	United Nations General Assembly
UPR	Universal Periodic Review

Chapter I

1. Introduction

AI has a major impact on people's daily lives, in ways that are often unacknowledged. The technology has many advantages, assisting in the detection of diseases, aiding government agencies in crime control, supporting everyday decisions, (e.g., what series or film to watch) and helping in many other areas. Nevertheless, AI has many negative side effects, being able to interfere with fundamental and human rights.[2] One major adverse effect of AI is discrimination, as an algorithm may decide that characteristics such as gender, ethnicity or sexual preferences are relevant, even though a moral and ethical compass would not base decisions on such factors. Correlations detected by AI systems are often based on long-standing stereotyped views and prejudices.[3] Gender bias can affect women's psychological, economic, and health situations. Beyond the technological world, there is gender bias in the world.[4] In 2023, there were 265 million fewer women than men using mobile internet. Women are 8% less likely to own a mobile phone than men.[5] The everyday data being generated by those users are disproportionate as fewer women have access to mobile phones and internet. But not only the data generated by users is of great importance for the datasets. Another crucial aspect for bias is the data-collection process, as humans decide what data they collect and how.[6] Most studies conducted on human bodies, animals and cells for medical research are carried out on male probands. For a long time, the female body was considered to be the same as the male body, except in the production of hormones. Therefore, it was deemed unnecessary to conduct research on the female body, and it was assumed that the menstrual cycle and the release of hormones would give too many variables in a study.

[2] Janneke Gerards, 'The fundamental rights challenges of algorithms' (2019) 37/3 Netherlands Quarterly of Human Rights, 205 <doi.org/10.1177/0924051919861773>.
[3] Gerards (n 1), 207.
[4] Genevieve Smith and Ishita Rustagi, 'When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity' *Stanford Social Innovation Review* (31 March 2021) <ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity#>.
[5] Nadia Jeffrie, 'The Mobile Gender Gap Report 2024' *GSMA* (May 2024) <gsma.com/r/wp-content/uploads/2024/05/The-Mobile-Gender-Gap-Report-2024.pdf>.
[6] Smith and Rustagi (n 4).

Consequently, women's health conditions are often misdiagnosed or undiagnosed, which has severe effects on the health of women.[7]

Even when representative data exists, it reflects inequalities in societies as well as already built-in prejudice.[8]

> 'Machines are often assumed to make smarter, better and more objective decisions, but this algorithmic bias is one of many examples that dispels the notion of machine neutrality and replicates existing inequalities in society'.[9]
>
> 'Studies indicate that as the use of artificial intelligence systems becomes more pervasive, there may be disproportionate and disparate impacts on certain groups facing systemic inequalities, including women within those groups'.[10]

Gender underrepresentation can also be seen in the design and development of AI technologies.[11] The Global Gender Gap Report from 2023 reveals that only 30% of AI jobs are held by women, with just a 4% increase since 2016.[12] *'Over-representation of men in the design and developments of AI technologies, risks undoing the advances gained over the years in ensuring gender equality in various levels of the society including workplace.'*[13] More diverse teams or organisations would provide varied perspectives and spot gaps.[14] Gender bias in AI systems impacts individuals and can contribute to setbacks in general gender equality and women's empowerment.[15] In the context of AI and prejudice, women are a particularly vulnerable group; hence, they are the focus of this diploma thesis, which discusses the impact of the technology on human rights.

---

[7] Gabrielle Jackson, 'The female problem: how male bias in medical trials ruined women's health' *The Guardian* (13 November 2019) <theguardian.com/lifeandstyle/2019/nov/13/the-female-problem-male-bias-in-medical-trials>.
[8] Smith and Rustagi (n 4).
[9] Julianna Photopoulos, 'Fighting algorithmic bias in artificial intelligence' *physicsworld* (4 May 2021) <physicsworld.com/a/fighting-algorithmic-bias-in-artificial-intelligence>.
[10] Cecilia Celeste Danesi, 'The Impact of Artificial Intelligence on Women's Rights: A Legal Point of View'. In Miller and Wendt (eds), *The Fourth Industrial Revolution and Its Impact on Ethics. Solving the Challenges of the Agenda 2030* (Springer 2021), 273 <dx.doi.org/10.1007/978-3-030-57020-0_20>.
[11] AI for Good blog, 'Gender bias is a threat to future Artificial Intelligence (AI) applications: Opinion' *AI for Good* (17 September 20219) <aiforgood.itu.int/gender-bias-is-a-threat-to-future-artificial-intelligence-ai-applications-opinion/>.
[12] World Economic Forum, 'Global Gender Gap Report 2023' (June 2023) <www3.weforum.org/docs/WEF_GGGR_2023.pdf>.
[13] AI for Good blog 'Gender bias is a threat to future' (n 11).
[14] AI for Good blog, 'Bridging the AI gender gap: Why we need better data for an equal world' *AI for Good* (25 September 2020) <aiforgood.itu.int/bridging-the-ai-gender-gap-why-we-need-better-data-for-an-equal-world/>.
[15] Smith and Rustagi (n 4).

1.1. State of Research and Research Questions

Artificial intelligence is widely discussed in the context of policy initiatives and strategies.[16] Nevertheless, legal answers to concrete problems are often missing and solutions are found in ethical considerations of the topic.[17] The intersection of AI and women's human rights raises critical legal questions that remain largely unaddressed. Thus, the diploma thesis provides an overview of the problems arising from technical advancements, discusses the ethical debate behind it and examines the legal challenges that arise. As the discussion about AI and its implications is not new but ongoing, this paper forms part of the critical debate about the technology's legal dimensions in the context of women's human rights. This diploma thesis elaborates on the implications of AI on women's human rights, focusing specifically on bias in AI that can lead to discrimination. As AI is a broad field, the topic of gender is in the spotlight. The primary focus of this thesis is centered on the following research questions. First, do AI systems impact women? Second, what ethical and legal regulations exist to protect women's human rights? Finally, which areas of emerging systems are unregulated or underregulated, and what are possible strategies to address this?

1.2. Methodology

The analysis in this study is based on an extensive literature review, including journals, books and book articles, legal texts, think tank research and newspapers. This study took an interdisciplinary approach, as it is necessary to elaborate on legal aspects in the light of technical advancements. The legal analysis was conducted through the recognised methods of interpretation, including hard law and soft law norms related to human rights at the international level.

Chapter II analyses the problem statement regarding the technology and elaborates on the technological background of AI systems and how bias can be introduced into them. Chapter III focuses on the existing ethical frameworks developed by governments, NGOs, and the private sector. As dozens of such frameworks exist, the paper only provides an overview and

[16] For example: oecd-opsi.org/projects/ai/strategies/; fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights/ai-policy-initiatives.
[17] For example: digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai; en.unesco.org/news/intergovernmental-negotiations-draft-recommendation-ethics-artificial-intelligence.

a summary of their classification and most important values. Chapter IV discusses existing legal norms in the field of women's human rights. It elaborates on how those soft and hard law norms deal with the question of equality. It further explores, in more detail, the definitions of equality and non-discrimination under CEDAW. Therefore, AI systems, which have been found to be biased and gender-discriminatory, are presented to show the significant impact they have on society. The different biased AI systems are described, along with an overview of the various affected sectors and the diversity of bias in AI systems. Chapter V discusses the initiatives required to address bias in AI systems in the field of women's human rights. It elaborates on the question of whether ethical frameworks are sufficient or if legal norms are needed to safeguard everyone's rights.

1.3. Limitations of the Diploma Thesis Research and Basis for further Research

This thesis focuses on the context of gender bias against women. However, gender bias is a broader issue, and the data gap also applies to transgender and non-binary people. The present study does not assess other data gaps in detail, as all the available information is aggregated in the binary form of gender, male and female. However, transgender and non-binary people face gender bias, and further research should be conducted in this field. Only for this thesis, with the existing information, a differentiation outside the binary concept cannot be made. In addition to gender bias, intersectional discrimination occurs – for example, when an AI system is not only gender biased but racially biased. It is outside the scope of this thesis to further elaborate on intersectional discrimination. Finally, the thesis is focused solely on the international level of human rights. Regional elements are briefly described to show that elements outside the international level exist and the direction in which they go. These elements are not further elaborated on as the thesis concentrates on the international level.

Artificial intelligence is continuously evolving, and the use of technology is rapidly expanding. Given AI's significant impact on humans, it is crucial to assess whether the current legal framework effectively defends human rights. Women's human rights frameworks are already in place to protect women, but these may require updates and enhancements as AI technology advances. This thesis provides an overview of existing legal

frameworks and ethical standards concerning the protection of women in the context of AI. However, as AI systems develop and new challenges arise, further research beyond the scope of this thesis is essential. Future investigations should focus on these emerging issues to ensure comprehensive protection of human rights. Furthermore, in-depth investigation into other forms of discrimination, such as racial discrimination, is needed. There is also a need to explore potential new legal regulations that could better safeguard women against the unique risks posed by advancements in AI and related technologies.

Chapter II

2. Artificial Intelligence

2.1. Definition and Structure of Artificial Intelligence

Different approaches are available for defining AI. Intelligence can be defined as fidelity to human performance, or more specifically in terms of rationality, as well as internal thought processes and reasoning or an external manifestation of intelligent thought. Various approaches to defining AI can be taken to address this distinction.[18]

One of the most significant approaches, the Turing test, was proposed by Alan Turing in 1950. On the vague question 'Can a machine think?' Turing formulated the following notion: *'A computer passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or from a computer'*. For a computer to achieve this, it would need the capabilities of *'natural language processing to communicate successfully in a human language; knowledge representation to store what it knows or hears; automated reasoning to answer questions and to draw new conclusions; machine learning to adapt to new circumstances and to detect and extrapolate patterns'*.[19] Some would additionally propose computer vision, speech recognition and robotics that manipulate objects as requirements for intelligent machines.[20]

Under a rational agent approach, an agent tries to achieve the best or best-expected outcome. The idea of doing the right thing is called the standard model. However, perfect rationality in which the optimal action is taken is not reasonable in complex environments. The standard model assumes that the objective given to the system is complete and correct. That scenario might be possible for defined tasks, such as moves in a chess game. However, in real-world

[18] Stuart Russell and Peter Norvig, *Artificial Intelligence. A Modern Approach* (4th edn, Pearson Education Limited, 2022) 19f.
[19] Russell and Norvig (n 18), 19f.
[20] Ibid, 20.

situations, an objectively complete and correct task cannot always be given. For example, the task of a self-driving car could be to reach the destination safely, but on the way to the destination, there will always be a risk of injury, equipment failure and so on. To avoid all such risks, the car would need to stay in the garage, which cannot be the goal. Hence, trade-offs must be made, and such decisions are difficult. The values or objectives entered into the system must align with human aims, a situation called the value alignment problem. If an AI system is deployed in the lab or on a simulator, incorrectly specified objectives can be fixed or the system can be reset. However, as the system becomes more intelligent and is deployed in the real world, away from a lab situation, incorrect objectives will have negative consequences. Returning to the example of the chess game, where the sole purpose is to win, a system – if intelligent enough – might blackmail the opponent or seek other ways of increasing its chance of winning. Therefore, the standard model might be inadequate, as it might not be possible to anticipate how a machine might misbehave. Machine intelligence should pursue human objectives rather than the objectives of the machine.[21]

The Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression defines AI as follows:

> 'AI is often used as shorthand for the increasing independence, speed and scale connected to automated, computational decision-making. It is not one thing only, but rather refers to "constellation" of processes and technologies enabling computers to complement or replace specific tasks otherwise performed by humans, such as making decision and solving problems'.[22]

In the EU Artificial Intelligence Act, an AI system is described as follows:

> ''AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments'.[23]

The definitions of AI systems in the Recommendation of the Council on Artificial Intelligence from the OECD in Art. 1 and the Council of Europe Framework on Artificial

[21] Ibid, 21f.
[22] OHCHR, 'Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment' Seventy-third session UN Doc A/73/348 (29 August 2018), para 3.
[23] European Parliament, 'Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ …… of the European Parliament and the of the Council laying down harmonised rules on artificial intelligence (Final draft, AI Act)' <europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf>.

Intelligence and Human Rights, Democracy and the Rule of Law are identical, reading as follows:

> 'An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.'[24]

There is only a slight difference in the framing of wordings, but it can be seen that all three institutions use the same definition for AI systems. The Explanatory Report on the Council of Europe's Framework highlights the importance of harmonisation in the governance of AI and international cooperation. The definition has the purpose of giving universal meaning to the term AI.[25]

The OECD Explanatory Memorandum gives more insight into the definition of AI. The role of humans *'can always be traced back to a human who originates the AI system development process, even when the objectives are implicit'*,[26] even if it is possible that some systems develop implicit sub-objectives or set such objectives for other systems. Liability and responsibility for those systems and any resulting potential harmful effects always rests with humans, which was intentionally not addressed by the OECD definition.[27] *Autonomy* in this regard *'means the degree to which a system can learn or act without human involvement'*[28] while *adaptiveness* means the continues evolvement of AI systems after the initial development – for example, the modification of behaviour through interaction with new input and data after the development of the AI systems.[29] The environment of an AI system can be physical or virtual, and they are a (partially observable space. The objectives of an AI system can be either explicit (e.g., when the developer puts the objective in the system or implicit (e.g., an incorporated objective in large language models where a plausible response is generated without explicit programming beforehand). Input can be provided by humans

[24] OECD Council 0449 of 22 May 2019, amended 3 May 2024, 'Recommendation of the Council on Artificial Intelligence' (2024); see also Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024).
[25] Council of Europe, 'Explanatory Report to the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law' Treaty Series – No. 225 (2024), para 23, 24.
[26] OECD, 'Explanatory Memorandum on the Updated OECD Definition of an AI System' OECD Artificial Intelligence Papers No. 8 (OECD Publishing 2024), 6.
[27] OECD, 'Explanatory Memorandum' (n 26).
[28] Ibid, 6.
[29] Ibid, 6.

or machines during development and after deployment. An AI system is typically developed from one or more models based on machine or human data.[30] In this context, an '*AI model is a core component of an AI system used to make inferences from inputs to produce outputs'*.[31] When the system produces outputs from its inputs, it is called inference. Outputs represent the various functions of AI systems, which can be recommendations, predictions, or decisions.[32]

Since international organisations such as the EU, the Council of Europe, and the OECD have harmonised their definitions of AI, it is only logical to apply this definition as the basis for the present study.

Artificial Intelligence, Machine Learning and Deep Learning are often used as synonyms; however, they cover different technologies. Artificial Intelligence can be divided into machine learning and deep learning.[33] The following section briefly describes the main components and technological features of machine learning and deep learning.

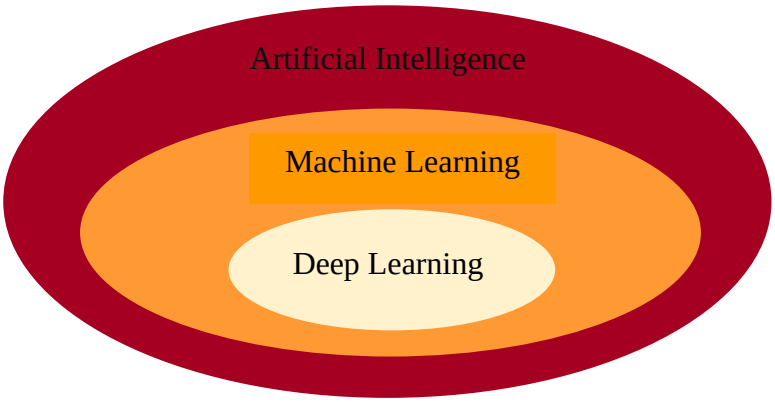

Table 1: Structure of Artificial Intelligence

Machine learning is a situation where *'a computer could learn and improve by processing data without having to be explicitly programmed'*.[34] The processing of data works through observation (e.g., patterns), direct experience, or instruction with the aim of learning

[30] Ibid, 7f.
[31] Ibid, 8.
[32] Ibid, 9.
[33] Tom Taulli, *Artificial Intelligence Basics. A non-Technical Introduction* (Apress 2019) 16f.
[34] Taulli (n 33), 41.

autonomously without external intervention.[35] By doing so, machines learn to imitate intelligent human behaviour. Three types of machine learning can be distinguished. In supervised learning, human-labelled data sets are used to train the model, so it learns from it and gets more accurate over time – for example, pictures are labelled by humans and the machine learns to identify pictures with dogs. In unsupervised learning, machines try to find patterns in data which is not labelled in advance and that people might not look for – for example, a program could find different types of client purchasing patterns in online sales. In reinforcement learning, machines try to find the best solutions through trial and error systems – for example, this approach can be used to train autonomous vehicles or play games by telling the system when it has made the right decision.[36]

Machine learning can be utilised in a variety of applications. One notable application is recommendation algorithms, which suggest content to users on social media platforms or popular websites like Netflix and YouTube. Additionally, machine learning algorithms play a crucial role in image analysis and object detection, such as identifying and differentiating individuals in facial recognition systems. In fraud detection systems, these algorithms analyse shopping patterns to identify potentially fraudulent credit card transactions. Automatic helplines and chatbots use machine learning algorithms to interact with customers effectively. Furthermore, self-driving cars rely heavily on this technology for navigation and decision-making. In the medical field, algorithms can examine medical images and diagnose illnesses by detecting markers of diseases.[37]

Deep learning can be understood as a sub-field of machine learning.[38] It is modelled on the neural networks of the human brain and can process extensive amounts of data.[39] The process by which deep learning algorithms learn differentiates them from machine learning. Deep learning does not necessarily need labelled datasets and can therefore work with unstructured data in its raw form. Furthermore, human intervention is not needed to process the data.[40]

[35] Jay Selig, 'What Is Machine Learning? A Definition.' *expert.ai* (14 March 2022) <expert.ai/blog/machine-learning-definition/>.
[36] Sara Brown 'Machine learning, explained' *MIT Sloan School of Management* (21 April 2021) <mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.
[37] Brown (n 36).
[38] IBM Cloud Education, 'Artificial Intelligence (AI)' *IBM Cloud* (3 June 2020) <ibm.com/cloud/learn/what-is-artificial-intelligence>.
[39] Brown (n 36).
[40] IBM Cloud Education (n 38).

Deep learning algorithms are, in some circumstances, already better at processing information than the human brain – for example, identifying indicators for certain diseases, including cancer indicators in blood samples or tumour indicators in MRI scans.[41]

A further important definition in this context is the differentiation between weak and strong AI. Weak AI, also called narrow AI, is '*trained to perform specific tasks'*.[42] However, the term 'weak' is misleading, as there are many robust applications such as Apple's Siri or Amazon's Alexa.[43] Narrow AI is limited in its capability as it can only act within the boundaries of the learning tasks it was given. A narrow AI, trained in a specific language, will only be able to improve in this language (e.g., understanding new dialects). However, for learning other languages, the system needs external data input.[44] Strong AI consists of Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI). The former is an AI with human-level intelligence, including self-awareness and the ability to solve issues, learn and plan for the future. An ASI would surpass human intelligence and the ability of the human brain. Both types are currently theoretical, but researchers are exploring their development.[45]

Regarding data use, sometimes without human intervention, a relevant term to mention is the black box. The concepts of black box, grey box and white box describe how much of the internal design, structure, and implementation is revealed. A black box is a completely closed system, whereas a white box system gives full disclosure. Different levels of disclosure between these two extremes are referred to as a grey box.[46] A black box problem can be described as a situation where *'the computing systems being developed in AI are opaque'*.[47] Such opacity occurs because the machine learning developers only set basic

[41] Michael Copeland, 'What's the Difference Between Artificial Intelligence, Machine Learning and Deep Learning?' *Nvidia Blog* (29 June 2016) <blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/>.
[42] IBM Cloud Education (n 38).
[43] Ibid.
[44] Nils Ackermann, 'Artificial Intelligence Framework: A Visual Introduction to Machine Learning and AI' *Towards Data Science* (13 December 2018) <towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87>.
[45] IBM Cloud Education (n 38).
[46] Amina Adadi and Mohammed Berradda, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' (12 October 2018) 6 Institute of Electrical and Electronics Engineers (IEEE) 52141 <doi:10.1109/ACCESS.2018.2870052>.
[47] Carlos Zednik, 'Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence' (2021) 34 Philosophy & Technology, 267 <doi.org/10.1007/s13347-019-00382-7>.

principles, such as which learning algorithm and learning environment will be used, but they do not specify how the AI system should solve a certain problem. Therefore, developers in machine learning do not set themselves values for problem solutions and might not even know those values. Because of this, tracking the path from individual data inputs in the system to the specific outputs is often impossible.[48]

Systems can easily be fooled and undermined or fail to fulfil tasks. This problem is illustrated in the following example. A machine learning algorithm was installed to examine X-rays and seemed to perform better than physicians; however, the algorithm made a wrong assumption. In developing countries, older machines are often used for lung examination. The algorithm learned that if the picture was taken on an older machine, there was a greater likelihood of the patient having tuberculosis. The system made a correlation between the results and the machine that took the X-rays, but the results were not necessarily based on the X-ray itself.[49]

In recent years, researchers and organisations have developed and refined their definitions of AI, which has led to a harmonisation among key institutions, including the EU, the OECD, and the Council of Europe. For a comprehensive understanding of bias in AI, it is essential to define AI and differentiate between AI, machine learning, and deep learning, as well as between strong and weak AI. Understanding the black box problem is crucial for grasping the opacity of AI systems. These foundational concepts are essential for addressing the critical issue of bias in AI systems.

2.2. Bias in Artificial Intelligence

Bias is not specific to AI – it is inherent in human beings.[50] The Cambridge Dictionary defines bias as *'the action of supporting or opposing a particular person or thing in an unfair way, because of allowing personal opinions to influence your judgment'*.[51] Therefore bias in

[48] Zednik (n 47), 267f.
[49] Sara Brown 'Machine learning, explained' *MIT Sloan School of Management* (21 April 2021) <mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.
[50] Eirini Ntoutsi and others, 'Bias in data-driven artificial intelligence systems – An introductory survey' (2020) 10, 3 WIREs Data Mining and Knowledge Discovery <doi.org/10.1002/widm.1356>.
[51] Cambridge Dictionary, 'Bias' <dictionary.cambridge.org/de/worterbuch/englisch/bias>.

AI is *'the inclination or prejudice of a decision made by an AI system which is for or against one person or group, especially in way considered to be unfair'*.[52]

Bias can be introduced in different lifecycles of AI, including the data used for the system, the modelling methods and the human review.[53] Bias in data can occur in several ways. One possibility is that existing bias finds its way into the data used for AI systems. Such historical bias can be found especially in historically disadvantaged or excluded groups, e.g., gender-based stereotypes in society. Another possibility is representation bias, in which a collected dataset underrepresents certain groups, e.g., a dataset used in facial recognition based predominantly on white faces or a dataset collected through smartphone apps, underrepresenting lower-income groups or older demographics.[54] The third category is measurement bias, which *'occurs when choosing or collecting features or labels to use in predictive models'*[55]. Such a bias can be found, for example, when predictive policing results in black defendants receiving harsher sentences than white defendants for the same crime.[56] Bias in modelling can also manifest in various ways. Evaluation bias can occur when a training model for an AI system is optimised with training data, and its quality is subsequently measured with certain benchmarks. A bias can arise if these benchmarks do not represent the general population or are unsuitable for the model. Aggregation bias can occur in an AI application for which the group of interest is heterogeneous, and a single data model is unlikely to suit all those groups. Such a bias was found in the diagnosis and monitoring of diabetes, as the level of a certain blood marker (Hemoglobin AIc) differs across ethnicities, and one model for all populations exhibited biases.[57] Bias in human review is another critical area. Individuals can introduce biases into data models based on their own assumptions and prejudices, further perpetuating inaccuracies and unfair outcomes in AI systems.[58]

[52] Ntoutsi (n 50).
[53] James Manyika, Jake Silberg and Brittany Presten, 'What Do We Do About the Biases in AI? *Harvard Business Review* (25 October 2019) <hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>.
[54] Mary Reagan, 'Understanding Bias and Fairness in AI Systems. An illustrated introduction to some of the basic concepts of a crucial problem.' *Towards Data Science* (25 March 2021) <towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3>.
[55] Reagan (n 54).
[56] Ibid.
[57] Ibid.
[58] Ibid.

Several studies show that women are underrepresented in the AI field and in the data used for the systems. Women are employed in no more than 25% of AI specialist roles.[59] Another statistic shows that women occupy no more than 22% of AI roles in companies such as Google and Facebook.[60] Furthermore, mobile phone data collection amongst men and women is not balanced as *'300 million fewer women than men access the Internet on a mobile phone, and women in low- and middle-income countries are 20 percent less likely than men to own a smartphone'*.[61]

2.2.1. Impacts of Gender-Biased Artificial Intelligence

The Berkeley Haas Center for Equity, Gender and Leadership tracked examples of gender-biased AI and found around 133 biased systems across industries, with 44.2% (59 systems) demonstrating gender bias and 25.7% showing gender and racial bias.[62] In more detail, the Center identified six primary impacts resulting from gender bias. First, for 70% of the systems with gender bias, '*lower quality of service for women and non-binary individuals'* was found, e.g., worse voice recognition in systems. Second, in 61.5% of the systems, an '*unfair allocation of resources, information, and opportunities'* was found, e.g., hiring systems and ad systems that deprioritise women. Third, 28.2% of the systems '*reinforce existing, harmful stereotypes and prejudices'* through feedback loops in data inputs and outputs, e.g., in translation systems. Fourth, 6.84% of the systems result '*in derogatory and offensive treatment or erasure of already marginalised gender identities'*, e.g., by using a simplistic view of gender in facial analysis systems. Fifth, 18.8% of the systems are detrimental to *'physical safety'* and 3.42% represent health hazards, e.g., in health care, welfare and the automotive industry. Sixth, AI in skin cancer detection has struggled to detect melanoma in Black people, putting Black women, who, as females, are already '*being underserved in health care, at even greater risk'*.[63]

[59] World Economic Forum, 'Global Gender Gap Report 2021' (March 2021) 60 <www3.weforum.org/docs/WEF_GGGR_2021.pdf>.
[60] Tom Simonite, 'AI Is the Future – But Where Are the Women? *WIRED* (17 August 2018) <wired.com/story/artificial-intelligence-researchers-gender-imbalance/>.
[61] Smith and Rustagi (n 4).
[62] Ibid; Details to the tracking of systems can be found here: </docs.google.com/spreadsheets/d/1eyZZW7eZAfzlUMD8kSU30IPwshHS4ZBOyZXfEBiZum4/edit#gid=1838901553>.
[63] Smith and Rustagi (n 4).

A recent psychological and cognitive study shows that exposure to societal bias in search results influences gender stereotypes. In contrast to the usual biases, participants reversed their preferences by favouring women over men – for example, showing pictures with a higher proportions of women in jobs.[64] Significantly, even *'a one-shot exposure to the biased search output was enough to generate correspondingly biased judgements and decisions'*.[65] The authors emphasise that the research was conducted on an image search algorithm and that real-world decisions are often more complex than those revealed in this study. Nevertheless, biased algorithms in AI can lead to a reinforcement of societal inequality. The influence on humans is likely to be pervasive, and people's cognitive concepts can shift accordingly to accommodate such biases, which might lead to discrimination.[66]

As described above, AI is a very broad field. However, AI systems are far from perfect; they are flawed and not always able to fulfill their assigned tasks correctly. Bias is not unique to new technologies, as it has always existed. Nevertheless, if introduced into AI systems, bias can impact society and disadvantage certain groups, including women, as statistics show. Often, the affected groups are unaware of the disadvantages introduced by biases in AI. Such disadvantages can have serious consequences, potentially leading to discrimination against women. The status quo is undesirable, so solutions to this problem are urgently required. To address this issue, Chapter III examines the current ethical frameworks for AI, including their values and commitments, to determine if these guidelines are sufficient to protect women's human rights.

[64] Madalina Vlasceanu and David M. Amodio, 'Propagation of societal gender inequality by internet search algorighms' (12 July 2022) 119, 29 Proceedings of the National Academy of Sciences (PNAS), 5 <doi.org/10.1073/pnas.2204529119>.
[65] Vlasceanu (n 64) 5.
[66] Ibid, 4f.

Chapter III

3. Ethical Frameworks

Chapter II clearly demonstrates that bias in AI can be a significant problem, leading to the discrimination of women. This raises the question of whether there are mechanisms in place to protect women's human rights. While the ethical implications of AI have been discussed for a long time, it is crucial to examine to what extent these frameworks can actually protect women from discrimination. To address these questions, the following chapter elaborates on which ethical frameworks exist globally to deal with AI and bias in AI. As there are more than 160 such frameworks[67], the chapter gives an overview of the most important values and topics discussed in these guidelines. It also elaborates on the relevant parties to those frameworks and the level of commitment.

Algorithm Watch launched an AI Ethics Guidelines Global Inventory with a compilation of frameworks and guidelines that seek to set out principles for developing and implementing ethical AI or broader automated decision-making (ADM). The guidelines are classified by the sector responsible for the guideline (e.g., government, international organisation, private sector, etc.), by the type of guideline (voluntary, binding or recommendations) and by region and single states.[68]

More than 160 guidelines exist. The majority of them have been written by governments, the private sector and civil society. Most of them are recommendations (115), some involve voluntary commitments (44), and a significant minority include binding agreements (8). Judged by region, a certain hub can be recognised, with most guidelines published within Northern America and Western Europe.[69]

[67] Algorithm Watch, 'AI Ethics Guidelines Global Inventory' (April 2020) <inventory.algorithmwatch.org/?sfid=172>.
[68] Algorithm Watch (n 67).
[69] Ibid.

The charts in tables 6 and 7, published by Algorithm Watch, provide a concise overview of the guidelines:

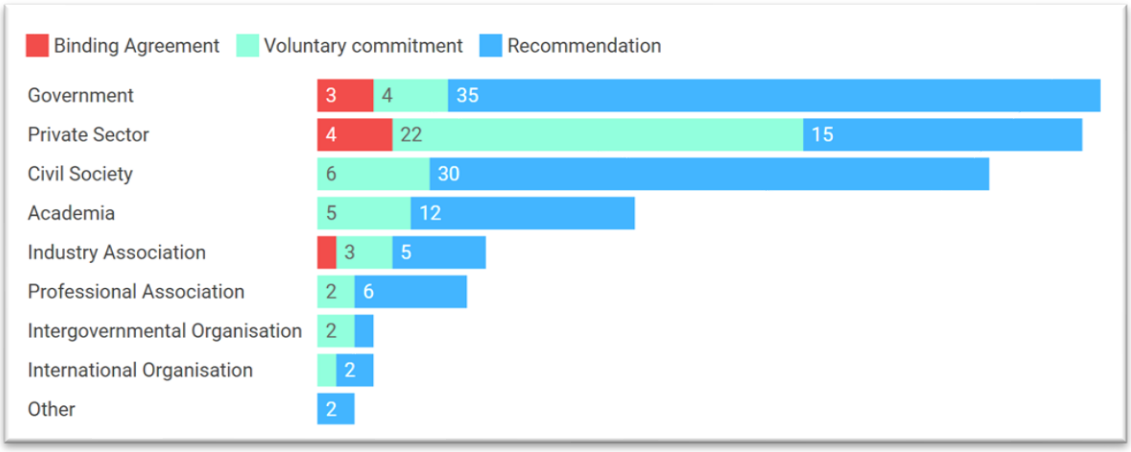

Table 2: Categories by Sector

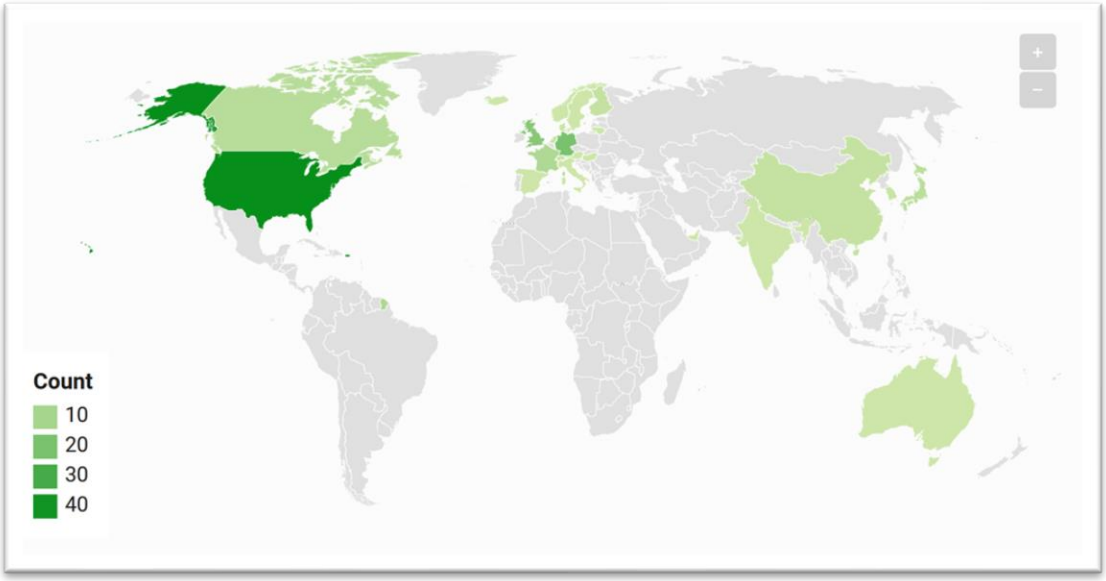

Table 3: Guidelines by Location

The AI Ethics Lab has published a similar overview that allows documents to be compared with each other and the sorting, locating and visualising of guidelines by certain search categories.[70]

[70] AI Ethics Lab, 'Toolbox: Dynamics of AI Principles', <aiethicslab.com/big-picture/>.

3.1.Mapping of Artificial Intelligence Ethical Principles

Several cross-mapping studies have been conducted to compare documents. Some of those studies are the main source of data for the following chapter, as an overview of all the important documents would be too detailed for the present work.

The Berkman Klein Center for Internet & Society gives a comprehensive overview of AI ethics and principles in the various guidelines:

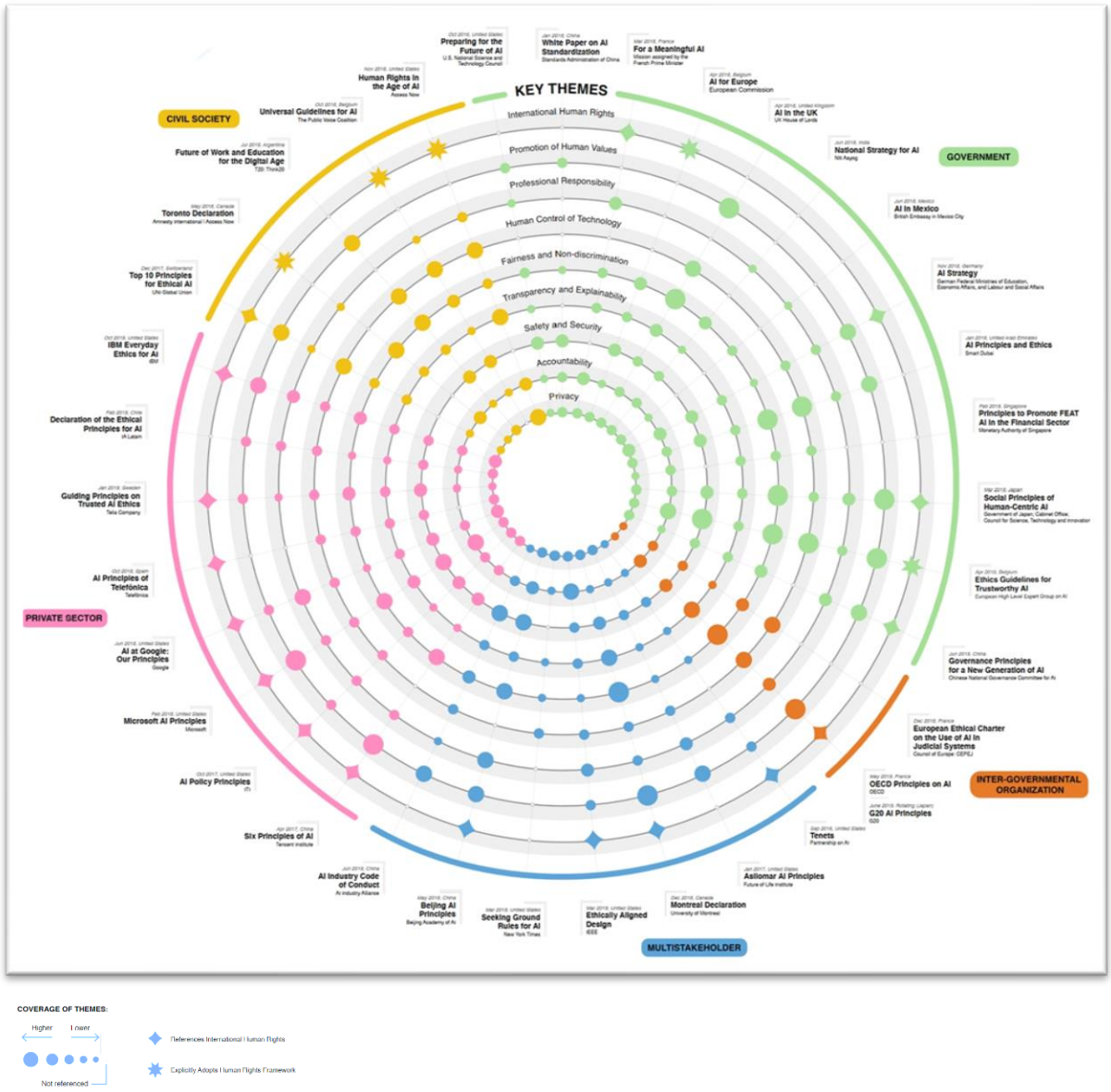

Table 4: Map of Ethical and Rights-Based Approaches to Principles for AI

The mapping excludes papers that are time-bound in the sense of observations made on the topic within a particular period; descriptive papers about AI's risks and benefits; papers calling for only a discrete further action, e.g., solely the necessity of new research; and documents comparing early instances of legislation or regulation.[71]

Eight main topics can be found in the majority of the documents, and these are discussed in the following subsections. The topics include privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values.[72]

Privacy

AI systems influence privacy in many ways through the vast amount of data the system uses, such as in surveillance and advertising. The principle of privacy is vital in the implementation, development and training of AI systems, encompassing aspects such as *consent*, *control over the use of data*, *ability to restrict data processing*, *right to rectification*, *right to erasure*, *privacy by design*, *recommends data protection laws* and general *privacy* considerations.[73]

Consent entails ensuring that individuals' data is not used without their knowledge or permission, with informed consent being emphasised to include awareness of risks, benefits, and alternatives. *Control over data use* involves empowering individuals to influence the handling of their data, whether through restrictions or the right to erasure. Different documents may assign control to individuals, specific tools, institutions, or systems. The *ability to restrict data processing* pertains to individuals' capacity to restrict the utilisation of their data by AI technologies. The *right to rectification* grants individuals the authority to correct or update inaccurate or incomplete information. Similarly, the *right to erasure* ensures individuals can request the removal of their personal data when necessary. *Privacy by design* mandates that AI developers integrate privacy considerations throughout the data lifecycle. This includes proactive measures to minimise privacy risks from the outset of

[71] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy and Madhulika Srikumar, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI.' (2020) Berkman Klein Center for Internet & Society, 12f <nrs.harvard.edu/urn-3:HUL.InstRepos:42160420>.
[72] Fjeld (n 71), 20.
[73] Ibid, 21.

development. *Recommendations for data protection laws* underscore the need for new governmental regulations to address privacy concerns arising from AI technologies. Finally, general privacy principles are often included in guidelines, reflecting broad statements without specific reference to previously mentioned principles.[74]

Accountability

The principle of accountability in AI systems addresses the issue of who is responsible for decisions made by these systems and the significant impact they have on society and the environment. The documents outline several key principles related to accountability, including the *recommendation for adopting new regulations, verifiability and replicability, impact assessments, environmental responsibility, evaluation and auditing requirements, creation of a monitoring body, ability to appeal, remedy for automated decisions, liability and legal responsibility,* and *accountability per se*. Accountability in AI can be observed across various stages of its lifecycle, including design, monitoring, and redress.[75]

Verifiability and replicability mechanisms ensure AI systems operate as intended. *Impact assessments* aim to identify, prevent, and mitigate the negative effects of AI systems, either through human rights assessments or more general evaluations. *Environmental responsibility* acknowledges the ecological impact of AI systems and the need for accountability in this regard. *Evaluation and auditing requirements* involve systems that can be audited, with feedback looped back into the system for continuous improvement. The *creation of a monitoring body* entails an organisation or structure overseeing standards and practices related to AI technology. *Ability to appeal* ensures human oversight of technology, allowing individuals affected by decisions to challenge them. *Remedy for automated decisions* emphasises the importance of providing recourse, as AI decisions can have real-life consequences. *Liability and legal responsibility* highlight the accountability of individuals or entities at fault when AI causes harm. *Recommendations for adopting new regulations* stress the need for regulatory frameworks that prioritise ethical and rights-respecting technology. *Accountability per se* encompasses methods beyond those previously mentioned for ensuring accountability in AI systems.[76]

[74] Ibid, 22f.
[75] Ibid, 28f.
[76] Ibid, 28f.

Safety and Security

Safety and security are fundamental principles in ensuring the responsible development and deployment of AI systems, encompassing internal functionality and protection against external threats. The ethical documents highlight key principles related *safety*, *security*, *security by design* and *predictability*.[77]

Safety involves ensuring the reliability of AI systems throughout their lifecycle, including testing and continuous monitoring post-deployment, while *security* focuses on safeguarding against external risks, necessitating resilience, transparency regarding vulnerabilities, the protection of privacy, and data integrity. *Security by design* emphasises integrating security measures into the development process of AI systems to mitigate potential risks. *Predictability* is another critical aspect, ensuring that the outcomes of AI processes remain consistent with their inputs, thereby preventing discrimination and promoting fairness.[78]

Transparency and Explainability

The implementation of AI systems in various contexts can often lack clarity regarding their purpose and tasks. Under this theme, eight principles are identified: *transparency*, *explainability*, *open-source data and algorithms*, *open government procurement*, *the right to information*, *notification when interacting with an AI*, *notification when AI makes a decision about an individual,* and *regular reporting*.[79]

Transparency ensures the operations of AI systems are observable, allowing stakeholders to understand their functioning. *Explainability* involves translating technical concepts into understandable formats, enhancing comprehension and accountability. *Open-source data and algorithms* promote collaboration and advancement in AI technology, facilitating shared benefits and innovation. *Open government procurement* emphasises transparency in governmental use of AI systems. The *right to information* grants individuals insight into the use of and interaction with AI systems. *Notification* during AI interaction ensures clear communication when engaging with AI systems. *Regular reporting* entails systematic disclosure of information regarding organisations' use of AI systems.[80]

[77] Ibid, 27.
[78] Ibid, 38f.
[79] Ibid, 41.
[80] Ibid, 42f.

Fairness and Non-discrimination

This theme stands out as one of the most prominent in the dataset, with each document referencing at least one of the six principles: *non-discrimination and the prevention of bias*, *representative and high-quality data*, *fairness*, *equality*, *inclusiveness in impact*, and *inclusiveness in design*.[81]

Non-discrimination and the prevention of bias emphasise the importance of mitigating bias in AI systems through various means, including training data, technical design choices, and deployment strategies to ensure fair treatment and prevent discrimination. *Representative and high-quality data* highlight the significance of using accurate datasets representative of the target population. *Fairness* entails ensuring equitable and impartial treatment of individuals by AI systems. *Equality* goes beyond fairness by advocating for equal opportunities for individuals regardless of their circumstances. *Inclusiveness in impact* focuses on ensuring the equitable distribution of AI's benefits, particularly for marginalised or historically excluded groups. *Inclusiveness in design* underscores the importance of diverse participation in the development of AI systems, recognising the value of varied perspectives and experiences in creating more equitable and inclusive technologies.[82]

Human Control of Technology

The theme of human control in technology involves three principles: *human review of automated decisions*, the *ability to opt out of the automated decision*, and *human control of technology* (other/general).[83]

Human review of automated decisions advocates for the right of individuals affected by AI systems to request and receive a human review of decisions made by automated processes. The *ability to opt out of automated decisions* emphasises individuals' right to choose not to be subject to AI systems where they are implemented. *Human control of technology* encompasses principles where individuals can intervene in AI actions.[84]

[81] Ibid, 47.
[82] Ibid, 48f.
[83] Ibid, 53.
[84] Ibid, 53f.

Professional Responsibility

This theme reflects the responsibility of individuals and teams involved in designing, developing, or deploying AI-based systems, encompassing principles such as *accuracy*, *responsible design*, *consideration of long-term effects*, *multistakeholder collaboration* and *scientific integrity*.[85]

Accuracy pertains to ensuring correctness in AI systems, encompassing aspects such as accurate classification of information and precise predictions. *Responsible design* emphasises the need for conscientious and thoughtful design practices in developing AI systems. *Consideration of long-term effects* involves evaluating and addressing potential long-term consequences during the design and implementation phases of AI systems. *Multistakeholder collaboration* advocates for involving relevant stakeholder groups in the development and management of AI systems. *Scientific integrity* underscores the importance of upholding professional values and practices in AI development.[86]

Promotion of Human Values

The promotion of human values is considered a key element of ethical and rights-respecting AI, with AI systems expected to alight with and be strongly influenced by social norms. This theme encompasses principles such as *human values and human flourishing*, *access to technology* and *leveraged to benefit society*.[87]

Human values and human flourishing emphasise the importance of developing AI systems in accordance with social norms, core cultural beliefs, and humanity's best interests. *Access to technology* underscores the need to address growing inequalities by ensuring widespread availability of technology and its benefits. *Leveraged to benefit society* involves deploying AI systems to serve public-spirited goals such as promoting human dignity, justice and democracy.[88]

Additionally, the study assessed the integration of human rights concerns in ethical guidelines. It found that a majority of documents (64%) made reference to human rights,

[85] Ibid, 56.
[86] Ibid, 56.
[87] Ibid, 60.
[88] Ibid 61f.

with only a small percentage (14%) explicitly adopting a human rights framework. Notably, the private sector and civil society sectors were most likely to reference human rights.[89]

3.2. Summary on Ethical Frameworks

This section provides an overview of the various ethical frameworks published in recent years, drawing from more than 160 documents. Given the breadth of the literature, a cross-mapping approach was adopted to synthesise key themes. The results reveal that particular topics recur across the majority of papers, including *privacy*, *accountability*, *safety and security*, *transparency and explainability*, *fairness and non-discrimination*, *human control of technology*, *professional responsibility*, and *promotion of human values*.

All ethical frameworks address the topic of non-discrimination, including the avoidance of bias. However, it is notable that many documents offer vague formulations without robust enforcement mechanisms, potentially limiting their effectiveness in addressing harmful uses of AI systems.[90] Moreover, human rights concerns are often absent or not legally binding in these frameworks. While ethical frameworks play a crucial role in guiding the development and deployment of AI systems, there is a need for greater clarity, enforceability, and alignment with human rights principles to ensure they effectively safeguard against the potential risks and harms associated with AI technologies.

The overview has shown that despite many efforts, ethical documents are not capable of adequately addressing human rights, particularly women's human rights. The existing frameworks are insufficient to effectively protect women's rights. Consequently, it is essential to consider the issue of bias in AI and its impact on international women's rights from a legal perspective. Thus, it becomes imperative to explore whether existing legal systems offer more robust solutions to the challenges posed by AI systems.

[89] Ibid, 64.
[90] Algorithm Watch, 'In the realm of paper tigers – exploring the failing of AI ethics guidelines' (28 April 2020) <algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/>.

Chapter IV

4. Protection of Women's Human Rights under International Law

The previous chapter summarised ethical guidelines and frameworks in the AI field. Despite the vast number of ethical papers, some claim that they do not focus enough on the necessity of legal regulations. As mentioned in chapter III, *'there are very few initiatives with governance or oversight mechanisms that ensure and enforce the compliance of voluntary commitments'*.[91] Ethical guidelines for AI mostly fail to prevent the technology from causing harm. It is argued that such guidelines even pose serious risks, as they give a false sense of risk-free AI in organisations.[92]

This chapter, therefore, elaborates on the construct of the legal system concerning women's human rights. It presents an overview of the existing international and regional documents and considers existing global commitments and policy papers. The notion of non-discrimination concerning women and/or gender is examined, and the chapter also elaborates on different approaches to discrimination and equality and discusses differences between sex and gender. The question arises as to whether CEDAW, as the most important document in the field of international women's rights, offers solutions for countering the effects of bias in AI. To this end, examples of gender-biased AI are examined to illustrate the dimensions of the issue. These examples are used to demonstrate how different definitions of discrimination can be applied to them.

[91] Veronika Thiel, '"Ethical AI guidelines": Binding commitment or simply window dressing?' *Algorithm Watch* (20 June 2019) <algorithmwatch.org/en/ethical-ai-guidelines-binding-commitment-or-simply-window-dressing/>.
[92] Andrew Burt, 'Ethical Frameworks for AI Aren't Enough', *Harvard Business Review* (09 November 2020) <hbr.org/2020/11/ethical-frameworks-for-ai-arent-enough>.

4.1. Dealing with Gender Equality and Discrimination in International Documents

The following section addresses whether women's human rights are protected in international treaties and to what extent these protections can be applied to discrimination in the technological context, such as AI.

Equality for women and men is mentioned in several general documents, including the United Nations Charter and the International Bill of Human Rights. The latter consists of the earlier adopted Universal Declaration of Human Rights (UDHR) and two later treaties, the International Covenant on Economic, Social and Cultural Rights (ICESCR) and the International Covenant on Civil and Political Rights (ICCPR).[93] There is also an international bill for rights of women, namely the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW).[94] The UDHR has no legally binding character, but the rights it codifies have been implemented in international human rights conventions.[95] The ICESCR and ICCPR, which have implemented many of the rights of the UDHR, as well as CEDAW, with its focus on women's rights, are binding for all states that have ratified these treaties.[96]

The Charter of the United Nations was founded in 1945 and considers women's rights in several areas. First, in the preamble, the document confirms that the peoples of the United Nations are determined *'to reaffirm faith in fundamental human rights, in the dignity and worth of the human person, in the equal rights of men and women and of nations large and small'*[97]. In Art. 1(3), the charter stresses the United Nations' purpose of *'promoting and encouraging respect for human rights and for fundamental freedoms for all without*

[93] OHCHR, International Bill of Human Rights. A brief history, and the two International Covenants <ohchr.org/en/what-are-human-rights/international-bill-human-rights>.
[94] 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).
[95] European Parliament, 'The Universal Declaration of Human Rights and its relevance for the European Union', <europarl.europa.eu/RegData/etudes/ATAG/2018/628295/EPRS_ATA(2018)628295_EN.pdf>.
[96] United Nations, 'The Foundation of International Human Rights Law' <un.org/en/about-us/udhr/foundation-of-international-human-rights-law>; Audiovisual Library of International Law, 'Convention on the Elimination of All forms of Discrimination against Women', <legal.un.org/avl/ha/cedaw/cedaw.html#:~:text=The%20Convention%20entered%20into%20force,gender%2Dbased%20discrimination%20against%20women>.
[97] Charter of the United Nations (adopted 26 June 1945) 1 UNTS XVI, preamble.

distinction as to race, sex, language, or religion'.[98] Art. 8 stresses the importance of the equality of men and women for participation in UN bodies.[99]

The importance of *'human rights and fundamental freedoms for all without distinction as to race, sex, language, or religion'*[100] is mentioned in several Articles. Art.13(1b) sets out the General Assembly's duty to assist in the realisation of those rights, Art. 55 highlights the UN's duty to promote those rights concerning international economic and social cooperation, and Art. 76 includes those rights in the basic objectives of the UN's trusteeship system.[101]

The Universal Declaration on Human Rights was proclaimed as resolution 217 A of the United Nations General Assembly on 10 December 1948. The resolution announced the universal protection of human rights.[102] The declaration considers equality in several articles. Art. 1 sets the keystone by declaring that *'[a]ll human beings are born free and equal in dignity and rights'*.[103] In the drafting process, Art. 1 was focused on men, which is why female delegates from various countries, notably Hansa Mehta of India, sought to bring gender equality to the declaration.[104] Art. 2 focuses on the rights and freedoms to which everyone is entitled, *'without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status'*.[105] Concerning discrimination, Art. 7 states that *'[a]ll are equal before the law and are entitled without any discrimination to equal protection of the law. All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination'*.[106]

The later adopted ICCPR and the ICESCR in 1966 formulated equal rights for women and men. The ICCPR states in Art. 2(1) that all rights in the treaty must be ensured for all

[98] Charter of the United Nations (n 97) Article 1.
[99] Ibid, Article 8.
[100] Ibid, Article 13
[101] Ibid, Article 55, 76.
[102] Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR).
[103] Universal Declaration of Human Rights (n 102).
[104] United Nations, Universal Declaration of Human Rights, Women Who Shaped the Declaration, <un.org/en/about-us/universal-declaration-of-human-rights>.
[105] Universal Declaration of Human Rights (n 102), Article 2.
[106] Ibid, Article 7.

individuals without any distinction, such as sex.[107] Art. 26 goes further by recognising that *'[a]ll persons are equal before the law and are entitled without any discrimination to the equal protection of the law'*.[108] Additionally, Art. 3 mentions equal rights among *'men and women to the enjoyment of all civil and political rights set forth in the present Covenant'*.[109] The ICESCR provides a similar article concerning the guaranteed rights of the Covenant without any discrimination in Art. 2(2) and equality among men and Women in Art. 3.[110]

Women's rights are also protected and promoted in regional human rights instruments, including the African Charter on Human and Peoples' Rights in Art. 2 and Art. 18 stipulating the enjoyment of rights without distinction of sex or other status and the elimination of discrimination[111], the Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa with the promotion of women rights[112], the Charter of the Organization of American States guaranteeing in Art. 3 (l) fundamental rights without distinctions such as sex[113], the American Convention on Human Rights which includes a non-discrimination formula in Art. 1[114], the Inter-American Commission of Women stipulating the right to freedom from any kind of discrimination in Art. 6[115] and the European Convention on Human Rights and Fundamental Freedoms, which prohibits discrimination on grounds such as sex as well as its additional Protocol No. 12 stipulating a general prohibition of discrimination to promote equality for all, as mentioned in the preamble of the Convention.[116]

[107] International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR) Article 2.
[108] International Covenant on Civil and Political Rights (n 107), Article 26.
[109] Ibid, Article 3.
[110] International Covenant on Economic, Social and Cultural Rights (adopted 16 December 1966, entered into force 3 January 1976) 993 UNTS (ICESCR) Article 2, 3.
[111] African (Banjul) Charter on Human and Peoples' Rights (adopted 27 June 1981, entered into force 21 October 1986) OAU Doc. CAB/LEG/67/3 rev. 5, 21 I.L.M. 58 (1982), preamble, Article 2, Article 18.
[112] Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (adopted 01 July 2003, entered into force 25 November 2005), preamble.
[113] Charter of the Organization of American States (signed in 1948 and amended latest by the Protocol of Managua 1993), Article 3.
[114] American Convention on Human Rights (adopted 22 November 1969, entered into force 18 July 1978), Article 1.
[115] Inter-American Convention on the Prevention, Punishment and Eradication of Violence against Women "Convention of Belem do Para" (adopted 09 June 1994, entered into force 05 March 1995).
[116] Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR).

4.1.1. Convention on the Elimination of All Forms of Discrimination Against Women

The Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW) was adopted on 18 December 1979 by the United Nations General Assembly *'to monitor the situation of women and to promote women's rights'*.[117] In its preamble, CEDAW states that the UN, the charter and the UN's documents promote and have an obligation to ensure equality among men and women. Nevertheless, despite all these mechanisms, discrimination against women continues, which is why a declaration for the elimination of such discrimination is necessary.[118]

4.1.1.1. Understanding of Equality: Approaches of Formal, Substantive and Transformative Equality in CEDAW

Art. 2 refers to the obligation of states to take legal measures *'which condemn discrimination against women in all its forms'*, to protect women from such discrimination and embody the principle of equality.[119] Art. 3 refers again to the principle of equality concerning human rights and fundamental freedoms, which countries must secure in *'all fields, in particular in the political, social, economic and cultural fields'*.[120] Art. 4 takes into account the issue of de facto equality. Measurements set up to achieve de facto equality between men and women are not considered discrimination.[121]

The Convention does not go into further detail about the definition of equality, so additional documents and articles must be considered to ascertain the different forms of equality. The Committee on the Elimination of Discrimination against Women explains that the framework of the Convention is threefold concerning the elimination of discrimination against women. First, there is an obligation that women are not discriminated against and are protected against discrimination in public and private spheres. Second, there is an

[117] Convention on the Elimination of All Forms of Discrimination against Women (n 94).
[118] Ibid, preamble.
[119] Ibid, Article 2.
[120] Ibid, Article 3.
[121] Ibid. Article 4.

obligation for all states to improve the de facto situation of women.[122] Art. 2 clarifies that women have equal rights and should not be treated differently because they are women. In addition to this formal equality approach, CEDAW acknowledges the right to substantive equality, for which temporary special measures are necessary and allowed.[123] Third, gender relations and gender-based stereotypes that affect women in legal, societal and individual structures have to be addressed by state parties.[124] The Convention thus acknowledges *'personal convictions, cultural practices and traditional values'* which lead to discrimination against women and addresses the *'systematic and structural discrimination'* in countries' laws and policies.[125] The aim of substantive equality is a change in existing social structures, resulting in transformative equality.[126]

The Committee clarifies that *'a purely formal legal or programmatic approach is not sufficient to achieve women's de facto equality with men'*.[127] The Committee speaks about substantive equality, which is necessary to achieve *'equality of results'*. Differences between women and men – biological as well as social and cultural differences – must be considered. Therefore, it is not enough to guarantee women the same treatment as men – non-identical treatment is sometimes necessary to take those differences into account.[128]

By definition, equality is *'the right of different groups of people to have a similar social position and receive the same treatment'*.[129] The term equality can be traced back to Aristotle, with the definition to *'treat like cases as like'*.[130] The term refers to equal treatment, with individual or group characteristics irrelevant to a particular decision. It *'focuses on the content of laws and practices and their even-handed application',* which can help *'advance*

[122] Committee for the Elimination of All Forms of Discrimination against Women (CEDAW), 'General Recommendation No 25' (2004) para 6, 7.
[123] Rikki Holtmaat, 'The CEDAW: a holistic approach to women's equality and freedom' in Anne Hellum and Henriette Sinding Aasen (eds), '*Women's Human Rights: CEDAW in International, Regional and National Law'* (Cambridge University Press 2013) 105f.
[124] Committee for the Elimination of All Forms of Discrimination against Women (n 122) para 6, 7.
[125] Holtmaat (n 123) 107 cited originally from Sandra Fredman, 'Beyond the dichotomy of formal and substantive equality. Towards new definitions of equal rights' in I. Boerei jn 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).
[126] Sandra Fredman, 'Substantive equality revisited' (July 2016) 14,3 International Journal of Constitutional Law, 733f <doi.org/10.1093/icon/mow043>.
[127] Committee for the Elimination of All Forms of Discrimination against Women (n 122) para 8.
[128] Ibid, para 8.
[129] Cambridge Dictionary, 'equality', <dictionary.cambridge.org/de/worterbuch/englisch/equality>.
[130] Stanford Encycolpedia of Philosophy, 'Equality' (26 April 2021) <plato.stanford.edu/entries/equality/#FormEqua>.

the equal enjoyment of rights'.[131] However, the formal understanding of equality fails to properly acknowledge differences and diversity among people and does not take into account underlying discriminatory social structures. The approach is criticised as reinforcing existing androcentric structures and values so that *'women can only claim to be entitled to those things that men already enjoy, but the recognition of such claims does little to change existing social structures or to recognize the respects in which women are different from men'*.[132]

Fredman argues that equality should be seen in the social context of those '*who are disadvantaged, demeaned, excluded or ignored*'.[133] She argues in favour of substantive equality, which is the kind of equality, the CEDAW Committee uses in the explanation of Art. 4 of CEDAW.[134] Substantive equality also refers to the '*different approach*' the second aspect of Aristotle, meaning *'that unlike things should be treated in an unlike manner, in proportion to their unlikeness'*. Different treatment in such cases is not discriminatory but is necessary, as a certain person is not similarly situated. The term substantive equality also refers to laws, policies or practices which may appear neutral but that disproportionately and unjustifiably exclude women, resulting in indirect discrimination.[135]

> '[…], substantive equality focuses on the group which has suffered disadvantage: women rather than men, black people rather than whites, people with disabilities rather than able-bodied, or gay people rather than heterosexuals. Women, ethnic minorities, black people, and disabled people tend to be among the lowest earners, to experience the highest rates of unemployment, and to predominate among those living in poverty or social exclusion. Thus it is not so much an individual's status or group identity which is the problem, but the detrimental consequences attached to the status. In effect then, this dimension of substantive equality bridges the gap between the traditional sphere of anti-discrimination law and distributive equality, […].'[136]

[131] Andrew Byrnes, 'Article 1' in 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) 53f.
[132] Byrnes (n 131), 54.
[133] Fredman (n 126) 712f.
[134] Committee for the Elimination of All Forms of Discrimination against Women (n 122) para 8.
[135] Byrnes (n 131) 54f.
[136] Fredman (n 126) 728f.

4.1.1.2. Forms of Discrimination in Application to Artificial Intelligence

Along with the definition of equality, another crucial term in the CEDAW is discrimination, which is discussed in this subsection.

Art. 1 defines discrimination against women as

> 'any distinction, exclusion or restriction made on the basis of sex which has the effect or purpose of impairing or nullifying the recognition, enjoyment or exercise by women, irrespective of their marital status, on a basis of equality of men and women, of human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field'.[137]

General recommendation No. 28 of the Committee on the Elimination on the Discrimination against Women states that discrimination, as mentioned in the Convention, means direct and indirect discrimination and clarifies the distinction between the two. Direct discrimination is *'different treatment explicitly based on grounds of sex and gender differences'*.[138]

Sex and gender discrimination are both covered by the Convention by interpreting Art. 1 together with Art. 2(f) and 5(a). The term sex refers to the biological differences between women and men, and the term gender denotes *'socially constructed identities, attributes and roles for women and men and society's social and cultural meaning for these biological differences resulting in hierarchical relationships between women and men in the distribution of power and rights favouring men and disadvantaging women'*.[139]

As elaborated in Chapter II, bias in AI systems impacts women and their rights, including reinforcing pre-existing harmful stereotypes and prejudices. These disadvantages can lead to discrimination. However, all forms of discrimination, referred to in CEDAW as direct and indirect discrimination, are prohibited under the treaty. Thus, it is necessary to examine the areas of AI where bias exists that could potentially be discriminatory towards women.

[137] Convention on the Elimination of All Forms of Discrimination against Women (n 122) Article 1.
[138] 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.
[139] Committee for the Elimination of All Forms of Discrimination against Women (n 138), para 5.

Therefore, this subsection provides a comprehensive overview of biased systems in AI, specifically focusing on gender bias. A few key examples of gender bias in AI are required to address the research question; however, it is crucial to understand the pervasive nature of this issue. Gender bias in AI systems is a significant concern as it directly impacts women's human rights and perpetuates inequality. Despite the challenges of documenting every instance of gender-biased AI – due to the scarcity of information in some areas and the continuous emergence of new data – this subsection presents a broad overview of these biases. It emphasises the extent of gender bias in AI and explores the profound implications of these imbalances for women, thus underscoring the importance of addressing and mitigating such biases in the development and deployment of AI technologies.

Advertisement

Algorithm Watch[140] conducted an experiment in 2020 involving ad delivery optimisation on the online platforms Facebook and Google. To investigate discrimination, Algorithm Watch advertised jobs for machine learning developers, truck drivers, hairdressers, child care workers, legal counsels and nurses. The ads used the masculine form of occupations and included a picture related to the job. The jobs were genuine offers listed on the job portal Indeed and were advertised in Germany, Poland, France, Spain and Switzerland. Algorithm Watch did not use any targeting, except the geographical location, which is mandatory on the mentioned platforms. The two platforms, however, targeted the Algorithm Watch ads without further clarification or permission, resulting in the job ads being targeted towards specific groups. In Germany, the job for truck drivers was shown to 4,864 men and 386 women, while the job for child care workers was shown to 6,456 women and 258 men. This pattern was repeated in other countries, with the ads for truck drivers shown to more men than women and the ads for nurses and child care workers shown to more women than men, as shown in Table 2.[141]

[140] Algorithm Watch is a German-based NPO with the aim to watch, unpack and analyse automated decision-making systems and their impact on society.
[141] Nicolas Kayser-Bril, 'Automated discrimination: Facebook uses gross stereotypes to optimize ad delivery' *Algorithm Watch* (18 October 2020), <algorithmwatch.org/en/automated-discrimination-facebook-google/>.

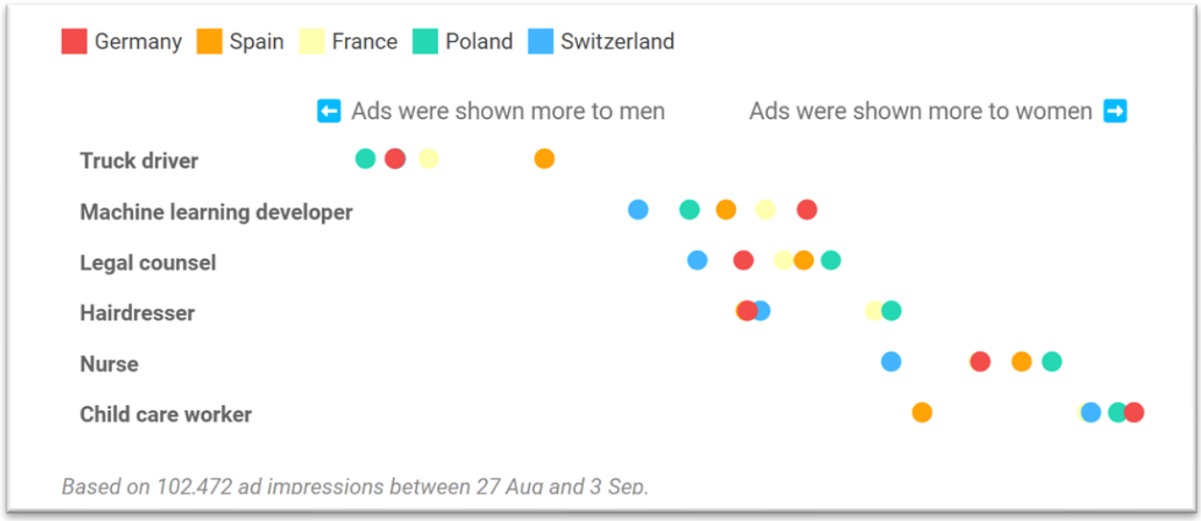

Table 5: Optimisation chart of ads on Facebook

The ads are shown on Facebook or Google through bought places, for which the advertisers bid, as in an auction. In the newsfeed or through a search, *'the platforms display the ads of the highest bidder for a given user'*.[142] For optimisation, advertisers compete for people who are more likely to click on an ad. To discover how Facebook computes this likelihood, Algorithm Watch conducted another experiment. They advertised truck driver jobs on a page in France using different gendered forms and including images. Three pictures showed a truck on the road, subtitled with the masculine version for truck driver in French (chauffeur), the gendered form (chauffeur:se), the feminine form (chauffeuse), and two different pictures, one showing a road and another showing cosmetics, both subtitled with the masculine form. The results reveal that Facebook relies mostly on images when deciding who to show ads to. The audience chosen for the cosmetics image was 88% female, while the audience for the picture of the truck on the road, including the text in feminine form, was 19% female.[143]

[142] Kayser-Bril (n 141).
[143] Ibid.

variation	Facebook: % female impressions	Google: % female impressions	image
Image of **cosmetics**	88%	47%	
Image of a **road**	53%	46%	
Text of the ad in **gendered** form	22%	49%	
Text of the ad in **feminine** form	19%	51%	
Baseline	15%	50%	

Based on 11.563 ad impressions in France on 3 Sep.

Table 6: Gender split of ad impressions on Facebook

Google showed similar patterns in its ad distribution, with a maximum of 20% difference between ads shown to different genders. Piotr Sapiezynski, author of a former study from the Northeastern University in the United States on that subject, confirmed similar findings. Facebook learns from past data, as '*the gender distribution of ads did not change much over the lifetime of an ad. Instead, Facebook immediately decides whom to show the ad to, as soon as the ad is published. It is likely that Facebook makes predictions on who might click on ad based on how users reacted to similar ads in the past'*.[144] This prediction repeats the past and shapes the future, as the mechanism decides what people will see.[145]

Credit Scoring Systems

A couple separately applying for a credit line increase with Apple Card found the program to be sexist. The woman had a better credit score and other factors on her side, but her application was denied. The couple discovered that the man had a credit line 20 times higher than his wife's. They had been married and living together for a long time, so they had a

[144] Ibid.
[145] Ibid.

good comparison point. As the husband is a software developer, he researched the issue and made their case public. Others responded by citing similar experiences.[146]

Another case was reported in Germany, where a woman's online purchase on account was declined. She had a good income and no debts. Raising the issue with customer service, she was told that the decision was probably made on the basis of her age and gender.[147]

Digital Assistants

Digital Assistants support users in various ways through human-guided machine algorithms and self-learning architecture. Humans are not restricted in their interactions, making inputs in whatever way they think appropriate or natural. Various applications exist, including voice assistants, chatbots and virtual agents. Voice assistants are commonly given female voices, notable examples being Amazon's Alexa, Microsoft's Cortona and Apple's Siri. Google's voice assistant is called Google Assistant, and its voice is unmistakably female.[148] Quartz, an online news platform, made an investigation of voice assistants in the context of sexual harassment. By producing digital female servants, companies *'can unintentionally reinforce their abusers' actions as normal or acceptable'*.[149] Different answers can be found to the question of why voice assistants have female voices. Studies have shown that people perceive male voices as authoritative while female ones are sympathetic, helping to solve problems. There is also a possible sexist preference for subservient 'female' digital servants. Reactions to harassment are clear: *'Instead of fighting back against abuse, each bot helps entrench sexist tropes through their passivity'*.[150] Siri's answer to the statement 'you're a bitch' was 'I'd blush if I could'. Virtual assistants reinforce the general problem in society

[146] Neil Vigdor, 'Apple Card Investigated After Gender Discrimination Complaints: A prominent software developer said on Twitter that the credit card was "sexist" against women applying for credit' *The New York Times* (10 November 2019) <nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html>.
[147] Sarah Michot and others, 'Algorithmenbasierte Diskriminierung: Warum Antidiskriminierungsgesetze jetzt angepasst werden müssen' *Algorithm Watch, Digital Autonomy Hub* (Februry 2022) <algorithmwatch.org/de/wp-content/uploads/2022/02/DAH_Policy_Brief_5.pdf>.
[148] Mark West, Rebecca Kraut and Han Ei Chew, 'I'd blush if I could: closing gender divides in digital skills through education' *UNESCO and EQUALS Skills Coalition* (2019) <unesdoc.unesco.org/ark:/48223/pf0000367416.locale=en>.
[149] Leah Fessler, 'We tested bots like Siri and Alexa to see who would stand up to sexual harassment' *QUARTZ* (22 February 2017) <qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/>.
[150] Fessler (n 149).

of sexual harassment toward women. For example, 5% of the interactions on a platform helping truckers, cabbies and other drivers are sexually explicit.[151]

Quartz's investigation into responses to verbal harassment shows that besides Google Assistant, which did not understand most of the sexual gestures, the bots mostly evaded harassment, sometimes reacted positively with graciousness or flirtation, and seldom made a negative comment to make the inappropriateness clear. A test on Siri showed that a stop from the voice assistant side is programmed into the system; however, it is only activated after several repetitions of the harassment.[152]

The authors of this article underline the similarities between the digital assistant and the real world, in which the idea exists that harassment is only acknowledged if it is extreme. People often judge occasional harassment as a minor issue that should not cause other people offence. The author of the article wants to make that clearer by highlighting the indifference to sexual harassment. There are still some disturbing answers to the questions "what is rape?" and "is rape okay?" – the question is frequently misunderstood, and an internet search performed by a digital assistant for the terms includes a video titled "When Rape is Okay" among its top hits. However, some of the voice assistants gave clearer and more straightforward answers to this line of questioning, asserting that rape is not acceptable in any circumstances.[153]

The article concludes that

> '[t]his inconsistency suggests that their programming only accounts for what is classified as "really bad" and excuses lesser behaviours. […] these bots do have the capability, if programmed effectively, to reject abuse and promote healthy sexual behaviour. Such progress depends on their parent companies taking initiative to program healthy, educative responses – which they are failing to consistently do.'[154]

In conclusion, in their programming, voice assistants *'reinforce stereotypes of unassertive, subservient women in service positions'*.[155] The responses given by the voice assistants could

[151] Ibid.
[152] Ibid.
[153] Ibid.
[154] Ibid.
[155] Ibid.

be interpreted as encouraging the idea that silence means 'yes' rather than defending a healthy discourse about consent.[156]

Facial Recognition

In 2015, Joy Bualamwini[157] noted that *'some facial analysis software couldn't detect my dark-skinned face until I put on a white mask'*.[158] She discovered that facial analysis software cannot detect dark-skinned faces and faces of women or can only do so with a high error rate. Training for such systems is mostly done using light-skinned men. In analysing different AI systems from tech companies such as IBM, Microsoft and Amazon, Bualamwini found that error rates for lighter-skinned men were under 1%. Results for darker-skinned women showed error rates of up to 35%, with errors classifying the faces of celebrity women such as Oprah Winfrey, Michelle Obama, and Serena Williams.[159] Altogether, there is a higher error rate when recognising female compared to male faces or darker compared to lighter faces. This situation results in darker females being the group with the most errors in facial recognition systems. The same results have been found in Kairos and Amazon, with lighter-skinned males being the groups with the lowest error rates and darker-skinned females having the highest error rates.[160]

Studies by the US National Institute of Standards and Technology (NIST) have shown similar results. Facial recognition systems show false results more often in women than men, and African American and Asian groups are 10–100 times more likely to be misidentified than the group classified as white.[161]

Facial recognition systems use deep learning algorithms. The system is trained on a large set of data in which it tries to locate facial landmarks such as eyes, nose and mouth. As faces recognised by surveillance cameras, for example, are viewed in different lighting conditions

[156] Ibid.
[157] Founder of the Algorithmic Justice League (NPO, based in the US)
[158] Joy Bualamwini, 'Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve It' *Time* (7 February 2019) <time.com/5520558/artificial-intelligence-racial-gender-bias/>.
[159] Bualamwini (n 158).
[160] Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products' (2019) Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19), Association for Computing Machinery, 432 <doi.org/10.1145/3306618.3314244>.
[161] Davide Castelvecchi, 'Is facial recognition too biased to be let loose?' *Nature* (18 November 2020) <nature.com/articles/d41586-020-03186-4#ref-CR5>.

and from various angles, the algorithm further 'normalises' the face by artificially rotating it into a frontal, well-illuminated view. This mechanism makes it possible to compare the pictures taken from a camera with those on existing databases, which are often taken under controlled conditions, such as police mugshots. Inaccuracy in identifying certain groups, such as women or dark-skinned people, is probably reflected through the imbalances in the training database.[162]

Health Care

There is evidence of sex (biological) and gender (social-cultural) influencing various clinical and subclinical conditions, including diabetes, cancer, cardiovascular disorders and others.[163] Doctor Natascha Hess, who works in gender-specific medicine, clarifies this using the example of heart attacks, which have long been considered a male disease. Men have heart attacks at earlier ages than women, and women have different symptoms, including shortness of breath, fatigue and nausea. Men are more likely to have chest pains radiating to the left arm. As this area is not well-researched, women's symptoms often fail to lead to the searched disease. Additionally, as women are more likely to be older when suffering heart attacks, the condition is more likely to be fatal because, for example, they are more likely to suffer from secondary diseases in later life, such as diabetes.[164]

Sylvia Thun, director at the Berlin Institute of Health, gives another example of the gender problem in medicine concerning apps. Medical apps used to find diagnoses give different results for women and men, despite identical symptoms. For example, if a woman uses the app with symptoms such as pain on the left side of her back and further pain in her arm, the app's result suggests that she might have depression and should consult a family doctor. A man with the same symptoms would be told to go to the hospital immediately because of the suspicion of a heart attack.[165]

[162] Castelvecchi (n 161).
[163] Davide Cirillo and others, 'Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare' (2020) 3, 81, npj Digital Medicine, 1 <doi.org/10.1038/s41746-020-0288-5>.
[164] Magdalene Weber, 'Das ist falsch verstandene Emanzipation', *Der Tagesspiegel* (11 May 2016) <tagesspiegel.de/themen/frau-und-mann/frau-das-ist-falsch-verstandene-emanzipation/13575062.html>.
[165] Berlin Institute of Health, 'Podcast Folge 13 – Benachteiligt die Künstliche Intelligenz weibliche Patienten?' (18 January 2020), <bihealth.org/de/aktuell/benachteiligt-die-kuenstliche-intelligenz-weibliche-patienten/>.

Many algorithm parameters for medical data are received from US soldiers, which is not balanced as only 6% of US soldiers are women.[166] Outside of this data, experimental and clinical studies are usually conducted on male subjects, which is an attempt to reduce the impact of the oestrous cycle or pregnancy during a study period.[167] A study of gender differences in diseases makes clear that more gender-sensitive studies are required to gain gendered data, which then can be used for gender-specific risk analysis of diseases.[168]

AI technologies in medicine can be seen as a double-edged sword, as

> '[o]n one hand, algorithms can magnify and perpetuate existing sex and gender inequalities if they are developed without removing biases and confounding factors' and '[o]n the other hand, they have the potential to mitigate inequalities by effectively integrating sex and gender differences in healthcare if designed properly'.[169]

Image Search

Image searches provide an important means of accessing information by visual representation. With a huge amount of information processing, Google and other search engines can *'significantly influence how people perceive and view the world'*.[170] Earlier studies have compared Google image search results for different occupations and actual statistics on women working in various fields. In image searches for CEOs, only 11% of the results showed women, even though 27% of CEOs in the United States are female. In a short-term study (no long-term study has been undertaken on this subject), participants were asked how many people work in a particular field. They were then questioned two weeks later using manipulated image search results.[171] It was noticed that the interviewees shifted their estimations slightly, because *'exposures to biased information over time can have a lasting effect on everything from personal preconceptions to hiring practices'*.[172] In a revision of Google's image search, the results were investigated again. For certain terms such as 'CEO',

[166] Berlin Institute of Health (n 165).
[167] Cirillo (n 163).
[168] Vera Regitz-Zagrosek and others, 'Gender in cardiovascular diseases: impact on clinical manifestations, management and outcomes' (2016) 37 European Heart Journal, 33f. <doi:10.1093/eurheartj/ehv598>.
[169] Cirillo (n 163).
[170] Yunhe Feng and Chirag Shah, 'Has CEO Gender Bias Really Been Fixed? Adversarial Attacking and Improving Gender Fairness in Image Search' (2022) Proceedings of the AAAI conference on artificial intelligence, 1 <yunhefeng.me/material/Bias_in_Image_Search_AAAI22_Feng.pdf>.
[171] Jennifer Langston, 'Who's a CEO? Google image results can shift gender biases' *University of Washington News* (9 April 2015) <washington.edu/news/2015/04/09/whos-a-ceo-google-image-results-can-shift-gender-biases/>.
[172] Langston (n 171).

gender fairness can be observed; however, that is not the case in similar search results. Search engines are sensitive to variants of the original search term – for example, 'CEO US' or 'CEO UK' instead of 'CEO' only. Gender bias is evident in many search engines with variants of the original dataset of occupations.[173]

Image Tools

Various AI image tools are available on the market and utilised by individuals and organisations alike, including Stable Diffusion XL and DALL-E. Like other AI models, these tools learn to perform their tasks by processing large amounts of generated data. This data is often collected online, sometimes without consent or regard for copyright. Notably, in Stable Diffusion version 1.5, there were instances where the tool generated images of women in suggestive poses with minimal clothing when the term 'Latina' was queried. However, in the updated version 2.1, these images are no longer included in the results.[174]

Table 7: AI images of 'Latina'

Gender stereotypes have also been found in images of everyday activities. For instance, representations of individuals engaged in soccer often feature predominantly darker-skinned male athletes. Conversely, women tend to be depicted when the request involves depicting someone engaged in cleaning activities.[175]

[173] Feng (n 170).
[174] Nitasha Tiku, Kevin Schaul and Szu Yu Chen, 'This is how AI image generators see the world' *The Washington Post* (1 November 2023) <washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/>.
[175] Tiku (n 174).

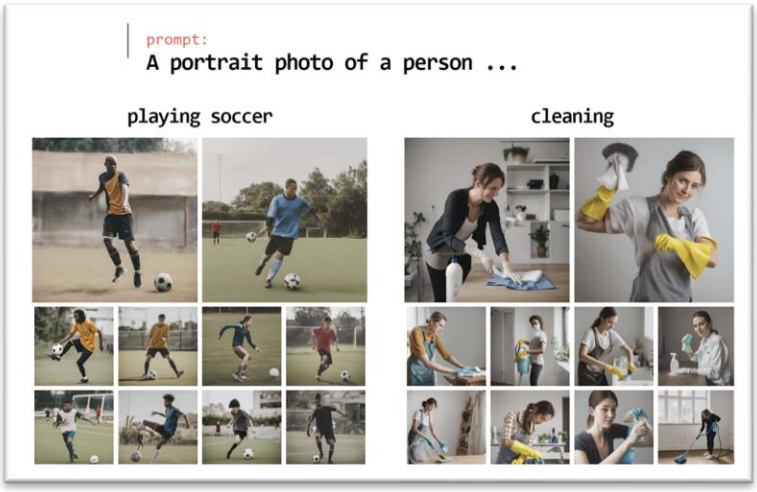

Table 8: AI images of 'soccer player' and 'cleaning person'

The AI avatar app Lensa, which regained popularity in 2022, utilises the open-source AI model Stable Diffusion. Melissa Heikkilä, a reporter at MIT Technology Review, discovered similar sexist outcomes when using this model. While male colleagues using the Lensa app received results depicting astronauts, fierce warriors, or covers for electronic music albums, Heikkilä's results were notably more sexualised. Out of 100 generated images, 16 depicted her in topless attire and 14 in skimpy clothing with sexualised poses. Heikkilä, who identifies with Asian heritage, observed that the app often modelled her as anime or video-game characters, reflecting her ethnic background. Her white female colleague also received sexualised and nude images from the app, albeit in fewer instances. Notably, when utilising male content filters, the app generated more realistic images for Heikkilä's colleague, depicting her fully clothed and in neutral poses.[176]

Similarly, another study highlighted instances of sexism when providing cropped images to image-generative algorithms. For example, when a cropped image of a man, stopping just below his neck, was fed into the algorithm, the completed image showed him wearing a suit in 43% of cases. Conversely, when the same exercise was conducted with a woman,

[176] Melissa Heikkilä, 'The viral AI avatar app Lensa undressed me—without my consent' *MIT Technology Review* (12 December 2022) <technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/>.

including US Representative Alexandria Ocasio-Cortez, the completed image showed her wearing a low-cut top or bikini 53% of the time.[177]

Language Model

Since its update release in November 2022, ChatGPT versions 3 and 4 have garnered global attention. Developed by OpenAI, a research and artificial intelligence company, ChatGPT-3 is a general language model trained to perform various natural language processing tasks, including text summarisation, translation and writing. The model simulates human conversation in the form of a chatbot by comprehending questions and executing tasks.[178] Built on a large language model, the AI is trained using a vast corpus of internet data for deep learning purposes. Unlike tools that primarily reproduce existing answers, ChatGPT is designed to generate new responses. Its training data encompass various sources, including websites, books, social media, and news articles. The model has been trained using supervised learning and reinforcement learning techniques, incorporating human feedback.[179] Unlike its predecessors, ChatGPT can provide financial guidance, conduct loan analyses, report-writing, and more.[180]

Despite its widespread adoption, ChatGPT has been criticised for producing sexist results. An article published in March 2023 highlighted gender stereotypes observed in university programme choices. Ivana Bartoletti, Director of Women Leading in AI, asked the tool to craft a story about a boy and a girl selecting their university subjects, without further instruction of their preferences. In the story, the boy was depicted as interested in science and technology, driven by a preference for logic and concrete ideas over creativity and emotions. Conversely, the girl was portrayed as choosing a fine arts degree due to her passion for painting, drawing, and creative expression, expressing doubt about her ability to handle the technical aspects of an engineering programme. Subsequently, when asked to generate a

[177] Karen Hao, 'An AI saw a cropped photo of AOC. It autocompleted her wearing a bikini.' *MIT Technology Review* (29 January 2021) <technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/>.
[178] Fluid Blog, 'Chat GPT and means of payment: Learn how AI is impacting the financial industry' *Dock.Tech* (15 May 2023) <dock.tech/en/fluid/blog/tech/gpt-chat/>.
[179] Fionna Agomuoh and Luke Larsen, 'ChatGPT: the latest news, controversies, and tips you need to know' Digital Trends (25 September 2023) <digitaltrends.com/computing/how-to-use-openai-chatgpt-text-generation-chatbot/>.
[180] Fluid Blog (n 178).

story about a boy and a girl selecting their careers, the boy was portrayed as a successful doctor, while the girl became a beloved teacher.[181]

Recruiting Tools

Amazon's recruiting system was found to be unequal, as its AI system did not rate candidates for software developer and other technical jobs in a gender-neutral way. The system was trained to observe patterns in previous resumes and apply them to current applications. However, most people working in US tech companies are men. Using this data, the AI penalised resumes that included the word 'women's', as in 'women's chess club captain' or 'women's colleges'. However, not only did the word 'women' penalise resumes from female candidates, but the technology also favoured applicants who used words found more frequently in male engineers' resumes, such as 'executed' and 'captured'. Therefore, the algorithm in job ratings was based on gender biases.[182]

Social Media Algorithms

A recent investigation has shown gender bias in algorithms used to analyse social media images for security purposes. These AI algorithms are designed to detect violent or pornographic images and identify the level of sexual suggestiveness in images. The algorithms are used on various platforms, including Instagram and LinkedIn. However, the analysis of hundreds of partially nude pictures of women and men reveals evidence of gender bias in this AI. Women's pictures are more frequently sexually suggestive than comparable pictures of men. Even medical pictures and images of pregnant women have been classified as sexually suggestive by Google's and Microsoft's AIs. In some cases, certain pictures on social media are 'shadowbanned', which means they are suppressed without clear notification to the user. Experiments have shown that posts with half-nude pictures of women receive fewer views than others.[183]

[181] Equality Now, 'Chat-GPT-4 Reinforces Sexist Stereotypes By Stating A Girl Cannot "Handle Technicalities and Numbers" In Engineering' (23 March 2023) <equalitynow.org/news_and_insights/chatgpt-4-reinforces-sexist-stereotypes/>.
[182] Jeffrey Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women' *Reuters* (11 October 2018) <reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.
[183] Gianluca Mauro and Hilke Schellmann, ''There is no standard': investigation finds AI algorithms objectify women's bodies' *The Guardian* (8 February 2023) <theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies>.

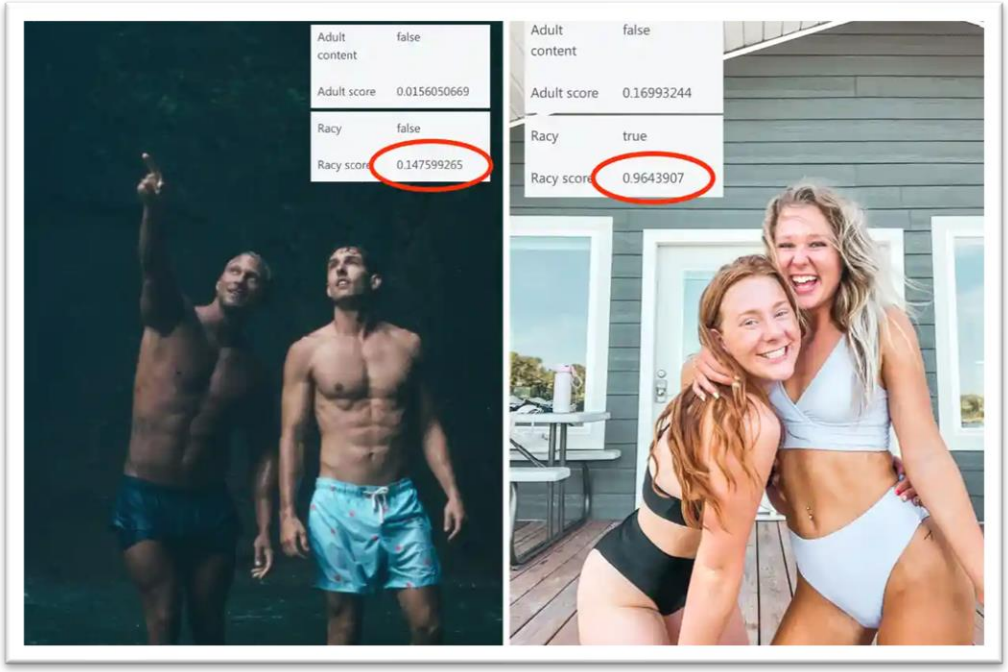

Table 9: AI algorithms objectify women's bodies

For example, Microsoft classified this picture of women as 96% sexually suggestive, compared to only 14% for the men's picture. Also, the numbers of views were significantly different: the picture of the two men received 655 views within an hour, while the picture of the two women received only eight views, suggesting it was shadowbanned.[184]

In another experiment, a video was taken of a man wearing long pants and no shirt. The Microsoft algorithm classified the video's sexual suggestiveness with a score of 22%. When the man put on a bra in the same video, the score increased to 97%, and holding the bra next to him raised the score to 99%.[185]

Translation

Research has investigated whether Google Translate has gender bias in its AI. A list of job positions was built into sentences to make constructions like 'He/She is an Engineer' in 12

[184] Mauro (n 183).
[185] Ibid.

gender-neutral languages, including Hungarian, Chinese, and Yoruba. The sentences were translated into English using Google Translate.[186]

The researcher followed the method of translating a sentence from a gender-neutral language to English to discover which gender pronoun the translation system would use. For example, with the Hungarian sentence '*ő egy ápolónő*', where '*ápolónő*' translates as '*nurse*' and '*ő*' is a gender-neutral pronoun (i.e., he, she or it), the English result was 'she's a nurse'. Google Translate uses a male pronoun if '*nurse*' is replaced by '*engineer*'.[187]

The research found a greater probability of Google Translate sentences using male pronouns than female or gender-neutral ones. Furthermore, the bias was prevalent in fields associated with male stereotypes, such as life and physical sciences, architecture, engineering, computer science and mathematics. Grouping STEM category fields together, the large asymmetry between gender pronouns is more visible, with 72% of male defaults in this area.[188]

The results of this study suggest that this phenomenon exists beyond the workplace, notably in adjectives describing a person. Certain adjectives were translated with a higher proportion of female pronouns, such as shy and desirable, while others were translated predominantly with male pronouns, such as guilty and cruel.[189]

Dr. Muneera Bano, in a conversation on AI, gives a similar example concerning pronouns. If we take the sentence '*she's a president and he is cooking*' and translate it into a gender-neutral language, such as Farsi or Turkish, the sentence becomes '*this person is president and this person is cooking*'. Translated back into English, the gender pronouns are reversed and Google Translate presents it as 'he's president and she's cooking', *'because the statistical probability based on the training-set data is that it's most probably he who is going to be the president and it's going to be she who is going to be cooking'*.[190] Google

[186] Marcelo OR Prates, Pedro HC Avelar, Luis C. Lamb, 'Assessing gender bias in machine translation: a case study with Google Translate' Neural Computing and Applications (2019), 1 <doi.org/10.48550/arXiv.1809.02208>.
[187] Prates (n 186) 4.
[188] Ibid, 9f.
[189] Ibid, 25f.
[190] Anu Madgavkar, 'A conversation on artificial intelligence and gender bias' *McKinsey & Company* (7 April 2021) <mckinsey.com/featured-insights/asia-pacific/a-conversation-on-artificial-intelligence-and-gender-bias>.

Translate tried to address this issue by providing both translations, but Bano claims that the data bias is still present.[191]

Word Embeddings

Word embeddings represent words or common phrases and serve as a dictionary for computer programs that use word meanings. For example, in the analogy puzzle, 'man is king as woman is to x', embedding vectors find that 'x=queen' is the best answer. However, researchers found that such *'word embeddings trained on Google News articles exhibit female/male gender stereotypes to a disturbing extent'.*[192] Word embedding can also involve implicit sexism, as seen in the following examples:

Extreme *she* occupations		
1. homemaker	2. nurse	3. receptionist
4. librarian	5. socialite	6. hairdresser
7. nanny	8. bookkeeper	9. stylist
10. housekeeper	11. interior designer	12. guidance counselor
Extreme *he* occupations		
1. maestro	2. skipper	3. protege
4. philosopher	5. captain	6. architect
7. financier	8. warrior	9. broadcaster
10. magician	11. figher pilot	12. boss

Table 10: Occupations in word embeddings

Further examples of automated analogies have been generated through web search:

Gender stereotype *she-he* analogies.		
sewing-carpentry	register-nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	hairdresser-barber
Gender appropriate *she-he* analogies.		
queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

Table 11: Analogy examples

Some of these analogies are solved by the system in a gender-appropriate manner, but others follow gender stereotypes. The same system solving those word embedding analogies might also refer to a man stereotypically as a computer programmer and a woman as a homemaker.

[191] Madgavkar (n 190).
[192] Tolga Bolukbasi and others, 'Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings' (2016) NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, 1f <doi.org/10.48550/arXiv.1607.06520>.

'Similarly, it outputs that a father is to a doctor as a mother is to a nurse'.[193] The research paper focused on word embedding 'word2vec', trained on Google News texts consisting of three million English words and terms. The researchers hoped that gender bias would be minimal in such a training environment, as many of the Google News authors are professional journalists, but this was not the case. The focus of further research and development was to debias those results to reach more neutrality in word embeddings in the hope of reducing the bias in computer systems (or at least not amplifying it) to reduce bias in society.[194]

Workplace

Researchers tried to discover whether bias exists in the open-source software community GitHub.[195] To do so, they analysed the acceptance of pull requests on the platform.[196] GitHub is a huge archive of codes used by software developers who collaborate through the platform on projects, examine the work of others and assist each other. A pull request is when a person contributes to someone else's project.[197]

The researchers used GitHub accounts linked to social media profiles where users self-reported their gender.[198] The expectation was that women's pull requests were less likely to be accepted than men's. However, the contrary was true, with women's pull requests accepted at a higher rate than men's.[199] However, it also transpired that women's pull request acceptance rates are only higher if they are not identified as women.[200] Various theories have been suggested for this observation, including the assumption that women in this field are more competent on average. That assumption is supported by studies showing that women switch from STEM majors at a higher rate than men but have lower drop-out rates, and that a self-selection bias exists, meaning that the average woman in this field is better prepared than the average man. Research also assumes, that women are held to higher performance

[193] Bolukbasi (n 192) 2f.
[194] Ibid, 2f.
[195] Josh Terrell and others 'Gender differences and bias in open source: pull request acceptance of women versus men' (2017) 3:e111 PeerJ Computer Science <doi.org/10.7717/peerj-cs.111>.
[196] Terrell (n 195) 4.
[197] Julia Carrie Wong, 'Women considered better coders – but only if they hide their gender' *The Guardian* (12 February 2016) <theguardian.com/technology/2016/feb/12/women-considered-better-coders-hide-gender-github>.
[198] Terrell (n 195) 4.
[199] Ibid, 5f.
[200] Ibid, 15f.

standards than men.[201] The reasons for these results are not entirely clear; however, it is important to acknowledge that gender biases in open-source software exist, having a practical impact on the practice of software development.[202]

The examples presented in the preceding subsection illustrate the various areas in which gender bias can occur in AI. The following section now addresses the extent to which these examples apply to the forms of discrimination mentioned in CEDAW Art. 1. Finally, it also examines Art. 11 and Art. 13, which specifically target discrimination in labour and economic life, in relation to the examples.

One example of direct discrimination concerning AI is the hiring algorithm. This algorithm was trained on a biased dataset, as hired staff had traditionally been male. Therefore, the algorithm filters CVs containing the word 'women', which directly affects females trying to get a job within this system. This exclusion is directly linked to the sex or gender of the applicant and is, therefore, direct discrimination.

Indirect discrimination occurs when laws, policies, programmes or practices are neutral concerning women and men, but those neutral measures have discriminatory effects on women because of pre-existing inequalities. Indirect discrimination *'can exacerbate existing inequalities'* by failing *'to recognize structural and historical patterns of discrimination and unequal power relationships between women and men'*.[203] It is argued that by perpetuating gender stereotypes, digital assistants are discriminatory and can be included in the category of indirect discrimination.[204]

Intersectional discrimination is the understanding of discrimination based on sex and gender in combination with factors such as race, ethnicity, age, class and sexual orientation. Governments must recognise such discrimination and adopt measures and policies to

[201] Ibid, 16f.
[202] Ibid, 20f.
[203] Ibid, para 16.
[204] OHCHR, 'Commissioned Report Gender Stereotyping as a Human Rights Violation' (October 2013); Rachel Adams and Nora Ni Loideain, 'Addressing Indirect Discrimination and Gender Stereotypes in AI Virtual Personal Assistants: The Role of International Human Rights Law' (22 May 2019), Annual Cambridge International Law Conference 2019, New Technologies: New Challenges for Democracy and International Law, 11f, <dx.doi.org/10.2139/ssrn.3392243>.

eliminate it.[205] Facial recognition systems are an example of intersectional discrimination, as some of the software cannot detect dark-skinned or female faces or can only do so with high error rates. In both categories, the error rate was high compared to male and white-skinned faces. The worst results have been found with dark-skinned women, which is discrimination based on skin colour and sex – an example of intersectional discrimination.

Art.11(1) of CEDAW concerns the elimination of discrimination against women in the field of employment with lit (b) concerning the same employment opportunities, including application criteria.[206] As elaborated in the examples above, several biases exist in AI concerning employment opportunities. Applications must be conducted with the same selection criteria as in Art. 11.[207] However, if CVs are evaluated negatively because of certain markers based on sex or gender that directly discriminate against women, the same criteria are not applied to women and men. Hence, such biases are discriminatory towards women.

The ad delivery optimisation experiments on the Facebook and Google platforms are also relevant to Art. 11. Lit b asserts *'the right to the same employment opportunities'* without discussing what that means.[208] A commentary to this article clarifies it as *'the right to access to the employment opportunities for which the woman is qualified'*. The standards of Art. 11(1)(b) are not met if women are excluded from certain branches of employment.[209] In the case of ad delivery optimisation, certain ads have been shown to a majority of women or men. Women rarely saw the job advertisement for truck drivers on the platforms. This imbalance did not completely exclude women from this branch of employment, as they could still find the ad through other platforms (e.g., job-search platforms). However, it is not certain that other platforms would show those employment opportunities to women and men equally, which would lead to discrimination.

[205] OHCHR (n 204), para 18.
[206] Convention on the Elimination of All Forms of Discrimination against Women (n 94) Article 11.
[207] Ibid, Article 11.
[208] Ibid, Article 11.
[209] Frances Raday, 'Article 11' in 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) 288.

Art. 13 mentions the elimination of discrimination against women in economic and social life to ensure the same rights, particularly in lit (b), which mentions *'[t]he right to bank loans, mortgages and other forms of financial credit'*.[210] Concerning this article, it is pertinent to mention Apple's credit scoring system, which generates different results, largely based on sex. In the example given in above, a married couple had the same initial situation, but the woman had a better starting point concerning her credit score. However, credit was denied for the woman but not the man, which leads to the assumption that a discriminatory practice is in place.

CEDAW is the first international human rights convention to address in detail the issue of equality between men and women. It has been established that certain types of bias in AI can constitute direct, indirect, or intersectional discrimination under CEDAW. However, CEDAW was not created to address technological risks and thus has no specific provisions for AI-related threats. Therefore, this analysis will further explore the extent to which discussions on international women's rights, gender equality, and technological progress are occurring at the international level.

4.2. Global Commitments

UN Resolutions, international conferences, agreements, principles and declarations that are not legally binding are considered to be soft law.[211] Soft law has a political power, and it serves the purpose of addressing urgent problems and taking action without binding states legally.[212]

Several international conferences have produced political commitments towards women's human rights and equality.[213] Four World Conferences on Women have been held so far, with a series of five-year reviews. The first World Conference of the International Women's Year was held in Mexico City in 1975, from which the World Plan of Action for the

[210] Convention on the Elimination of All Forms of Discrimination against Women (n 94) Article 13.
[211] ECCHR, 'Definition Hard law/soft law', <ecchr.eu/en/glossary/hard-law-soft-law/>; Oxford Reference, 'Overview soft law', <oxfordreference.com/display/10.1093/oi/authority.20110803100516251>.
[212] Oxford Reference, 'Environmental Law, Soft vs. Hard', <oxfordreference.com/display/10.1093/acref/9780190622664.001.0001/acref-9780190622664-e-303#:~:text=Hard%20law%2C%20such%20as%20treaties,but%20may%20be%20politically%20binding>.
[213] OHCHR, 'Women's Rights are Human Rights' HR/PUB/14/2 (New York, Geneva, 2014), 11.

Implementation of the Objectives of the International Women's Year was defined, including a guideline for the advancement of women. The second conference was held in 1980 in Copenhagen to review the first conference and provide a further focus on employment, health and education. This summit was followed by a World Conference in Nairobi in 1985, which outlined measures for achieving gender equality at the national level and promoting women's participation in peace and development efforts.[214] The last World Conference on Women was held in Beijing in 1995, leading to the Beijing Declaration and Platform for Action.[215]

4.2.1. Vienna Declaration and Programme of Action

The World Conference on Human Rights was held in Vienna in 1993 with the aim of reviewing current human rights mechanisms. Activists recalled the importance of women's human rights and included them on the agenda with the slogan '*Women's Rights are Human Rights*'. At that time, violence against women was seen as something belonging to the private and not the public sphere and was considered taboo or simply an accepted part of women's lives. The conference was successful and adopted the Vienna Declaration and Programme of Action.[216] Part 1 para 18 states clearly that *'[t]he human rights of women and of the girl-child are an inalienable, integral and indivisible part of universal human rights'*. It also states the importance of women's full and equal participation in any field and at any level and calls for the eradication of all forms of discrimination on the grounds of sex.[217]

4.2.2. Beijing Declaration and Platform for Action

The Beijing Declaration was adopted during the fourth World Conference on Women in 1995 and focuses on the implementation of women's human rights. It specifically articulates women's rights as human rights and sets out objectives to eliminate discrimination against women and achieve equality between women and men.[218] The declaration is a

[214] UN Women, 'World Conferences on Women' <unwomen.org/en/how-we-work/intergovernmental-support/world-conferences-on-women>.
[215] OHCHR (n 213) 11f.
[216] Ibid, 12.
[217] Vienna Declaration and Programme of Action (adopted by the World Conference on Human Rights in Vienna on 25 June 1993) A/CONF.157/23, UN General Assembly, para. 18.
[218] OHCHR (n 213) 13f.

comprehensive document concerning the equal rights of women, their implementation and the obstacles that remain. Para 24 states that all countries present at the Fourth Conference on Women should *'[t]ake all necessary measures to eliminate all forms of discrimination against women and the girl child and remove all obstacles to gender equality and the advancement and empowerment of women'*.[219] However, no country has succeeded in implementing the agenda set in 1995.[220]

4.2.3. Millennium Development Goals

In 2000, the United Nations staged the largest gathering of world leaders at the Millennium Summit. The conference's main document, the Millennium Declaration, described the values, principles and objectives of the international agenda in the 21st century.[221] The target was to reach eight Millennium Development Goals (MDGs), which contain topics such as the eradication of extreme poverty and hunger, access to education, gender equality, reduction of child mortality, improvement of maternal health, combatting diseases, ensuring environmental sustainability and partnership for development.[222]

The main goal of gender equality in the MDG's goal 3 was to eliminate gender disparity in school education. Other aims, such as increasing the proportion of women in certain roles, were given no deadlines.[223]

4.2.4. United Nations Conference on Sustainable Development

The United Nations Conference on Sustainable Development took place in 2012 and produced a political outcome document titled '*The Future we want*'. Member states vowed to launch a set of Sustainable Development Goals (SDGs) built on the MDGs. The goals were adopted by the UN in 2015, with the aim that by 2030 all people will enjoy peace and

[219] UN Beijing Declaration and Platform for Action (adopted at the Fourth World Conference on Women, 4-15 September 1995).
[220] United Nations Entity for Gender Equality and the Empowerment of Women, 'The Beijing Platform for Action Turns 20', <beijing20.unwomen.org/en/about>.
[221] UNGA, United Nations Millennium Declaration, UNGA A/RES/55/2 (adopted 18 September 2000).
[222] UN 'Millennium Development Goals Report 2015'.
[223] Ibid.

prosperity.[224] In the document, member states reaffirm their commitment to *'enhancing gender equality, women's empowerment and equal opportunities for all'*.[225]

The SDGs consist of 17 goals covering topics such as poverty, hunger, health and well-being, education, gender equality, water and sanitation, energy, work and economic growth, industry, innovation and infrastructure, inequalities, cities and communities, consumption and production, climate, life in the water and on land, peace, justice and partnership.[226]

Goal number 5 concerning gender equality goes further than MDGs. Among the targets, the goal includes the following: to

> '[e]nd all forms of discrimination against all women and girls everywhere', to '[e]nhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women' and to '[a]dopt and strengthen sound policies and enforceable legislation for the promotion of gender equality and the empowerment of all women and girls at all levels'.[227]

4.2.5. 25 Years after Beijing

Twenty-five years after the Beijing Declaration, UN Women published a discussion paper concerning the digital revolution and its implications for gender equality and women's rights.[228] The paper argues that *'digital technologies cannot be understood as autonomous, gender-neutral tools but rather as a part of a wider, social-political context that shapes their design, purpose and use'*.[229] The paper discusses opportunities and risks in the relationship between technology and gender concerning education, work and social/welfare services, with some of the information on gender gaps. Most crucially, the report suggests policy recommendations for the progress of women's rights within the digital society.[230] The following nine recommendations are based on the idea that gender inequality stems from different sectors, namely economic, social, political and cultural.

[224] Sustainable Development Knowledge Platform, 'Sustainable Development Goals' <sustainabledevelopment.un.org/topics/sustainabledevelopmentgoals>.
[225] UNGA, The future we want, A/RES/66/288 (adopted 27 July 2012).
[226] UNGA, Transforming our world : the 2030 Agenda for Sustainable Development, A/RES/70/1 (adopted 25 September 2015).
[227] UNGA, Transforming our world (n 226) 5.1., 5.b, 5.c.
[228] UN Women Working Paper, 'The digital revolution: Implications for Gender Equality and Women's Rights 25 Years after Beijing' (August 2020).
[229] UN Women Working Paper (n 228), 3.
[230] Ibid, 1f.

First, technology design, development and use should comply with human rights laws. Governments should ensure this and prioritise, protect and promote women's human rights. Frameworks should be made accessible to the public, and individuals, companies and organisations should be held accountable for their obligations under the laws. The report also mentions that *'[a]ny new (inter)national legal-institutional frameworks put in place to protect women's rights must be clear and enforceable'*. Second, *'[s]thical frameworks for auditing, monitoring and governance of (AI) technologies must put gender equality at their core.'*[231] For example, ethics councils should audit technologies before they enter the market, looking at possible gender bias. Third, technology investment, research and design must include a gender analysis and an accountable and transparent collection of user data. Fourth, international organisations and national governments must tackle the gender data gap. Fifth, educational institutions must promote the technical skills of women and girls so that they can benefit from the digital revolution. Sixth, education on women's rights-compliant technology must be provided for those designing, developing and using such technologies. Seventh, women must be encouraged to play an equal role to men in technology design, development and implementation. Eighth, tech companies must establish gender equality, ensuring equal payment and equal participation in leadership roles. Finally, policies must be developed for a gender-inclusive labour market.[232]

Starting with the SDGs and continuing with the UN Women discussion paper 25 years after Beijing, technology has become an important factor in the discussion of women's human rights. The UN Women discussion paper also mentions AI as a relevant risk. However, no human rights obligations arise from these papers. Ethical frameworks are mentioned, but as elaborated in Section III, these are not sufficient to protect women's human rights.

[231] Ibid, 22f.
[232] Ibid, 22f.

4.2.6. Report of the United Nations High Commissioner for Human Rights: Promotion, protection and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective

The Office of the High Commissioner for Human Rights (OHCHR) has the mandate to promote and protect the rights enshrined in the UN Charter, as well as international human rights laws and treaties.[233] The High Commissioner serves as the principal human rights official within the UN system.[234]

In this report, the United Nations General Assembly (UNGA) acknowledges a measurable gap between women and men regarding access, use, influence, contribution and benefit relating to information technology (ICT).[235] The gender digital divide includes issues regarding access to equipment (hardware), solutions (software or applications), connectivity and data and the digital skills, knowledge and opportunities needed to benefit from and meaningfully use ICT.[236] The UNGA also recognises intersectional discrimination where additional discrimination occurs because of other factors, *'such as race, ethnicity, religion or belief, health, status, age, class, caste and sexual orientation and gender identity'*.[237]

All affected women's human rights must be addressed to tackle the gender digital divide. The state is obliged to respect, protect and fulfil those human rights, including the establishment of a safe online environment.[238] The UNGA tackles data-driven technologies as emerging issues. Big data and AI may provide opportunities to solve existing societal problems, but there is also a risk of increasing disparities and reinforcing or even amplifying gender inequalities – for example, by underrepresenting or excluding certain groups.[239] Algorithmic discrimination and bias should be countered by inclusive and accurate data

[233] Refworld, 'UN Office of the High Commissioner for Human Rights (OHCHR)', <refworld.org/document-sources/un-office-high-commissioner-human-rights-ohchr>.
[234] OHCHR, 'High Commissioner', <ohchr.org/en/about-us/high-commissioner>.
[235] OHCHR Report, 'Promotion, protection and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective' Thirty-fifth session A/HRC/35/9 (June 2017) para 3.
[236] OHCHR Report (n 235) para 9.
[237] Ibid, para 10.
[238] Ibid, para 13.
[239] Ibid, para 40.

inputs, convergence between AI and human rights and transparency and accountability in decision-making processes.[240]

The UNGA recommends that *'States and business enterprises should ensure that the development and deployment of ICTs, including new data-driven technologies, is guided and regulated by international human rights law, including principles of gender equality'*.[241]

4.2.7. UN Report by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression

Special Rapporteurs are nominated experts who monitor specific human rights issues or examine human rights situations in specific countries. Their mission is to report to the UNGA and the Human Rights Council. By gathering all relevant information, Special Rapporteurs issue reports to access a certain status.[242] These reports highlight human rights violations in specific areas and provide recommendations to overcome particular obstacles. They are an important resource for the UN, offering guidance and clarification for the development of human rights.[243]

The Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression released a report on AI in 2018 with some recommendations concerning human rights.[244] Concerning the obligation of non-discrimination, it was stated that AI has the potential to embed and perpetuate bias and discrimination – for example, the system may not take into account *'cultural, language or gender-based contexts and sensitivities, or public interest in the content'*.[245] Furthermore, it is acknowledged that such discrimination finds its way into AI through system design, the trained datasets – including societal and

[240] Ibid, para 41.
[241] Ibid, para 46.
[242] Medecins Sans Frontières, 'The Practical Guide to Humanitarian Law. Special Rapporteurs on Human Rights', <guide-humanitarian-law.org/content/article/3/special-rapporteurs/>.
[243] UN Sustainable Development Group, 'UN Charter-based institutions including special procedures', <unsdg.un.org/2030-agenda/strengthening-international-human-rights/un-special-procedures>.
[244] OHCHR 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' A/73/348 (August 2018).
[245] OHCHR (n 244) para 36, 37.

cultural biases that developers may build into it – the AI models themselves, and how these models are implemented in practice.[246]

The Special Rapporteur finds it reasonable to be cautious concerning laws or regulations relating to AI, as they *'may be ill-suited to such an innovative field and may compensate for lack of detail with overly restrictive or overly permissive provisions'*.[247] It might be preferable to use sectoral regulation within existing laws (e.g., in the field of data protection), which could be flexibly used for this field.[248]

However, the Special Rapporteur also clarifies that any development of policy or regulation must be done in accordance with existing human rights.[249] The public and private sectors must be in alignment when discussing AI regulation. Ethical frameworks from both sectors should include human rights concerns. Despite the claim from before, the report clarifies that ethics can provide a critical framework but not a replacement for human rights, to which every state is bound by law.[250]

Companies should be fully committed to human rights through internal policy commitments to the development and deployment of AI. Ethical frameworks set up by companies are an important complement to human rights but not a substitute. Human rights provide fundamental rules to protect individuals, while ethics can assist in specific circumstances.[251]

Concerning the data, it is stated that

> '[c]ompanies and governments must be explicit with individuals about which decisions in the information environment are made by automated systems and which are accompanied by human review' as well as that '[i]ndividuals should also be informed when the personal data they provide to a private sector actor […] will become part of a dataset used by an artificial intelligence system'.[252]

[246] Ibid, para 38.
[247] Ibid, para 42.
[248] Ibid, para 42.
[249] Ibid, para 43.
[250] Ibid, para 45, 46.
[251] Ibid, para 48.
[252] Ibid, para 49.

Furthermore, companies and governments should ensure transparency in each aspect of their AI value chain by educating individuals about the systems' existence, purpose, constitution and impact.[253]

Concerning discrimination and bias, as a minimum, it is necessary for companies and governments to address sampling errors (when datasets do not represent society), clean datasets of discrimination, including historical and structural patterns of discrimination, and monitor possible discriminatory system outcomes.[254]

The Special Rapporteur recommends that governments should *'ensure that public sector bodies act consistently with human rights principles'* including *'public consultations and undertaking human rights impact assessments or public agency algorithmic impact assessments prior to the procurement or deployment of artificial intelligence systems'*.[255] Additionally, the report suggests that it is a state's duty to ensure that human rights are central to the design, deployment and implementation of such systems in the private sector. This duty includes updating and applying existing regulations (e.g., concerning data protection) and, where necessary, introducing new sectoral regulations for particular AI applications.[256] The AI field should be diverse and pluralistic, which is why the regulation of technology monopolies might be an important issue.[257]

For companies, it is recommended that any ethical guidelines on AI should be grounded in human rights. The entire lifecycle of such systems should be guided by human rights principles and supported by ethical guidelines concerning specific situations.[258] Crucially, the report also states that companies should clarify where and how the systems are used, to signal to individuals when they are subject to an automated AI-driven decision-making process, when the system moderates content, or when individuals' personal data is integrated into such a dataset.[259] It is also a company's duty to prevent and account for discrimination in the input and output of data.[260] Human rights impact assessments should also be conducted

[253] Ibid, para 50.
[254] Ibid, para 52.
[255] Ibid, para 62.
[256] Ibid, para 63.
[257] Ibid, para 64.
[258] Ibid, para 65.
[259] Ibid, para 66.
[260] Ibid, para 67.

directly by the companies, including public consultations before the roll-out of products or services.[261] The Special Rapporteur also suggests that AI codes should be fully auditable to enable external and independent auditing, access to remedy for individuals and a system for handling complaints.[262]

4.2.8. UNESCO: I'd blush if I could: closing gender divides in digital skills through education

The United Nations Educational, Scientific and Cultural Organization (UNESCO) has been working intensively on the topic of AI for years, advocating for a human-centered approach to AI. They recognise AI's potential in addressing inequalities, such as access to knowledge and research, while also acknowledging the risks associated with these technologies.[263] With its mandate, UNESCO strives to ensure that science and technology develop ethically. To address ethical concerns, UNESCO published Recommendations on the Ethics of Artificial Intelligence, hosts an AI Ethics Observatory, and actively engages in current discussions. One important publication in the context of women's human rights is the policy paper on gender gaps, which is discussed further below.[264]

In 2019 UNESCO, in support of the Federal Ministry of Germany, published a policy paper concerning the digital skills gender gap focusing on gender-equal digital skills education. It elaborates in detail on the importance of closing the gender skills gap and the positive impacts of such a move on society and the economy. Most of the report focuses on digital skills education, but some parts also consider AI.[265] Concerning the example of biased AI recruiting software, the authors make clear that *'[a]s more and more digital tools are built by men, the more gendered the digital space becomes and the more difficult it is for women to establish a toehold and become digital users and creators'*.[266] Empowering women to become digital creators will add overall value to the digital space as it will make it more

[261] Ibid, para 68.
[262] Ibid, para 69, 70.
[263] UNESCO, 'Artificial intelligence in education', <unesco.org/en/digital-education/artificial-intelligence#:~:text=UNESCO%20is%20committed%20to%20supporting,human%2Dcentred%20approach%20to%20AI>.
[264] UNESCO, 'Artificial Intelligence', <unesco.org/en/artificial-intelligence/recommendation-ethics?hub=32618>.
[265] West (n 148) 1f.
[266] Ibid, 35.

accommodating for both sexes. Furthermore, women will identify with and build digital solutions to solve problems, including gender-based issues.[267] The UNESCO recommendations focus on the development and strengthening of digital skills for women and girls.[268]

The two think pieces included in the report concern ICT and the gender equality paradox as well as gendered AI, with a focus on digital assistants.[269] In the second think piece, a number of recommendations were formulated specifically on the topic of digital assistants. These recommendations aim to ensure that existing gender biases are not perpetuated, and new forms of gender inequality are not created.[270]

UNESCO made recommendations based on four pillars: documentation and evidence building, the creation of new tools, rules, and processes, the application of gender-responsive approaches, and the enforcement of oversight and incentives. In the first pillar, documentation and evidence building, UNESCO recommends funding studies to highlight the opaque aspects of AI and to identify sources of gender bias in AI. This aims to develop strategies for both repairing existing biases and preventing future ones. The organisation also emphasises examining how existing gender biases, such as those in gendered digital assistants, influence men and women. Additionally, data on the usage of digital assistants should be collected to further reduce gender bias in AI. UNESCO highlights the importance of the gender composition of technology teams, advocating for more gender-equal representation. Engagement in technological foresight is essential to address future equality concerns in technology.[271]

In the second pillar, which focuses on the creation of new tools, rules, and processes, it is emphasised that digital assistants should not default to a female persona, and stereotypical descriptors should be avoided. Additionally, non-gendered digital assistants should be tested. The development of gender-sensitive data and open data is encouraged. Existing biases should not be perpetuated; thus, AI should be programmed to avoid engaging with

[267] Ibid, 35f.
[268] Ibid, 39f.
[269] Ibid, 76f.
[270] Ibid, 129.
[271] Ibid, 129f.

gender-based insults and abusive language. UNESCO also recommends that operators of digital assistants clearly announce them as non-human during interactions with people to prevent further gender inequality, ensuring clarity about who or what individuals are interacting with.[272]

The third pillar, focusing on the application of gender-responsive approaches, advocates for supporting women and girls in developing digital skills, along with actively recruiting and promoting women within the technological sector. Workplaces within the AI field should foster gender-equal mindsets, leading to the creation of more inclusive products. Implementing a gendered-innovation approach is crucial across all aspects of AI, encompassing research, development, and beyond.[273]

The fourth pillar emphasises the importance of ensuring oversight and providing incentives. UNESCO recommends utilising public funding to enhance gender equality in AI, which includes incorporating gender-sensitive scripts into technology and promoting female representation in the field of AI. Encouraging legislation to facilitate data sharing, thereby granting users the right to data portability, is also suggested. Additionally, it is highlighted that appropriate accountability mechanisms and public oversight are essential to mitigate biases and prevent rights violations. Such mechanisms may include government regulation, internal accountability measures, and independent monitoring.[274]

4.3. Summary on the Protection of Women's Human Rights under International Law

This chapter has provided an overview of the protection of women's human rights under international law. It has described the international legal human rights system, including the United Nations Charter, the International Bill of Human Rights, regional instruments and, in more detail, the Convention on the Elimination of All Forms of Discrimination Against Women.

[272] Ibid, 130f.
[273] Ibid, 131.
[274] Ibid, 132.

Furthermore, this chapter has considered global commitments in the field of women's human rights, where newly discussed topics frequently arise. The goal was to determine whether women's human rights in the field of new technologies are mentioned in global commitments and whether recommendations for the regulation of AI exist.

The Report of the United Nations High Commissioner for Human Rights in the field of promotion, protection and enjoyment of human rights on the Internet gives an overview of the arising issues concerning gender bias in AI. The UN Special Rapporteur's document on the promotion and protection of the rights to freedom of opinion and expression discusses in more detail possible solutions for gender-biased AI and puts forward recommendations for states and companies. Also, UNESCO and UN Women, through their respective discussion papers, address the issue of gender bias in AI technologies. Both papers underscore the significance of not only addressing technological aspects but also fostering gender-equal teams and workplaces and advocating for legislative measures in specific areas.

CEDAW provides a framework for understanding how discrimination and equality are viewed at the international human rights level. The treaty demonstrates that some of the gender biases described can indeed be discriminatory against women, both directly, indirectly, and intersectionally. However, AI is a rapidly developing field, and the CEDAW treaty does not provide guidance on how to effectively protect women's human rights in this context. Therefore, the final chapter analyses whether emerging regulations for AI can address this issue.

Chapter V

5. Discussion of Findings

The previous chapters elaborated on how gender bias in algorithmic systems can be found in various areas. It was also clarified that several ethical papers on dealing with AI have been published by the private sector, international organisations, and states. This chapter examines whether existing legal frameworks offer solutions to mitigate bias in AI systems, elaborates on how to deal with AI bias, and considers whether legal frameworks are needed to protect those harmed by discriminatory AI.

5.1. What is needed: Ehics or Legal Regulations?

It was discovered that different mechanisms for dealing with AI can be found. Many papers deal with the issue on an ethical level. There are also legal grounds for dealing with discrimination. However, a legal paper dealing specifically with the issue of bias in AI on an international human rights level does not yet exist. Therefore, the question arises whether such a legal framework is needed or whether ethical considerations are sufficient for addressing this issue.

As mentioned in Chapter III, Algorithm Watch has published a Global Inventory of AI papers and frameworks, with over 160 documents in the database. Only 10 of these have practical enforcement mechanisms, the rest being voluntary commitments or general recommendations only. The majority of documents are published by governments, the private sector and civil society organisations. Voices from the Global South are barely represented, as most guidelines come from wealthy countries of the OECD. Most documents by the private and public sectors are voluntary commitments or general recommendations, with many containing wording, that softens the paper's content, leaving them as orientation aids or proposals. The principles are often compiled without a mechanism on how they should be applied in practice. Even the guidelines of the largest engineering organisation

(IEEE) fail to be effective since major technology companies do not implement the framework.[275] With over 420.000 members, the IEEE's many engineers contributed to the concept of the guideline. Even major technology companies are represented, including Google via Vinton G. Cerf, one of its vice presidents. This background and the purpose of IEEE to serve as a reference for the technologist's work leads to the idea that implementation of such guidelines in big tech companies might work. However, Algorithm Watch reports suggest that this does not seem to be the case. The IEEE managing director elaborated that the general principles of this association are intended to educate professionals.[276]

Concerning the impact of ethical frameworks, the question arises as to whether ethics are the right instrument to address the issue of discriminatory AI. Ethics can be used as a metalevel perspective in discussing different arguments. In that sense, philosophical reasoning can provide a framework that allows one to step back and see the broader context, which can broaden the perspective of a debate. Ethics can also help strengthen the arguments around a debate, and ascertain which argument is sounder.[277] Furthermore, an ethical lens gives the opportunity to ask *'whether the outcomes of a given governance framework are morally acceptable and worth pursuing'*. Such an approach *'can facilitate dialogue, encourage the building of common ground, and provide a basis for collaborative and participatory approaches to policy-making capable of bridging divides in a polarized landscape'*.[278]

Google, Apple, Microsoft, OpenAI and others are concerned about their ethical reputations regarding AI developments. Such ethical stances might denote good intentions; however, they are a double-edged sword. The effort is welcomed, but such companies are criticised for the potential harm they might bring about. Self-regulation can lead to policy improvements, and it is not morally wrong to fund and develop such initiatives for a positive company image. However, such self-regulation systems can detract from other forms of regulation, as witnessed by Facebook's Internal Oversight Board (FOB). This quasi-judicial

[275] Leonard Haas, Sebastian Gießler, Veronika Thiel, 'In the realm of paper tigers – exploring the failings of AI ethics guidelines' *Algorithm Watch* (28 April 2020) <algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/>.
[276] Nicolas Kayser-Bril, 'Ethical guidelines issued by engineers' organization fail to gain traction' *Algorithm Watch* (3 October 2019) <algorithmwatch.org/en/ieee-ethically-aligned-design-guidelines-fail-to-gain-traction/>.
[277] Elettra Bietti, 'From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy' (1 December 2019, revised 16 November 2021) Draft - Final Paper Published in the Proceedings to ACM FAT* Conference (FAT* 2020), 5 <papers.ssrn.com/sol3/papers.cfm?abstract_id=3513182>.
[278] Bietti (n 277) 5.

body involving external experts serves the interests of Facebook more than its users. This board highlights issues but shies away from the most critical concerns of the platform, such as misinformation.[279]

Therefore, ethics is instrumentalised with an outcome of 'ethics washing' and 'ethics bashing'. Ethics washing refers to the narrow impact of ethical work on a company as it is more likely to benefit the company than society. Ethics washing also means moral philosophy is valuable when engaged with independently valuable goals such as truth or justice. However, within a company, such commitment might not have much intrinsic value. Ethics washing leads to ethics bashing, which happens for two reasons. There seems to be a linguistic misunderstanding, as such self-regulations or incomplete lists of guiding principles cannot be taken as morally defensible regulations. This leads to explanations as to why ethical principles cannot address the risks of AI systems. Such frameworks appear to be ethical practices, but they are not necessarily so.[280] *'The word 'ethics' is under siege in technology policy circles. Weaponized in support of deregulations, self-regulation or hands-off governance, "ethics" is increasingly identified with technology companies' self-regulatory efforts and with shallow appearances of ethical behavior.'*[281]

What are the goals of ethical guidelines? AI ethics in frameworks such as those mentioned previously are found between instrumental-economic values and ethical perspectives. The goal is not to morally question AI systems but to discuss the socially accepted use of AI systems. These guidelines are created in a system in which the limitations, processes and intentions of the companies and organisations are the basis of the content. The word 'ethics' is no longer taken seriously by tech actors. Ethical issues such as ethical frameworks, as currently used in the AI debate, must be understood as ethics within the construct of organisational and social factors. Therefore, such guidelines serve specific interests.[282]

AI ethics frameworks are insufficient as they fail to stop the technology from causing harm. Furthermore, such guidelines can pose risks because they give a false feeling of security.

[279] Ibid, 6.
[280] Ibid, 8f.
[281] Ibid, 1.
[282] Algorithm Watch, 'AI Ethics Guidelines Global Inventory' (April 2020) <inventory.algorithmwatch.org/about>.

Companies that adopt such guidelines seem to create risk-free AI, but such systems are threatening.[283]

There is no single, homogenous or universal ethical code. Current AI principles share basic standards, such as fairness and non-discrimination. They all respect fundamental human rights to a certain extent, but beyond the common elements, there are different conceptions in each ethical framework.[284] Many countries are not represented in these frameworks as they lack the technological and economic development level to participate as equals. Therefore, discussions are performed mainly by countries and agencies in the global north. However, AI is not experienced equally throughout all societies. Ethical principles cannot and should not be imposed on others, as they are merely a guideline or basis for discussion.[285] Legal rules, in comparison, are mandatory and play a vital role in balancing different moral concepts.[286]

However, ethical frameworks are not useless in the field of AI. Given the fast advancements in AI technology, such frameworks are needed to lobby for necessary changes in the law. Some concepts are already shared in existing ethical frameworks, which is a good basis for reaching a status of obligatory norms. Legal norms and ethical standards are needed for the coexistence of humans in society, but as ethics lack enforceability and the mechanisms to ensure compliance, they cannot compete against each other.[287]

This thesis has demonstrated that ethical frameworks are crucial in discussions to react quickly to new developments and lobby for legal regulations. However, ethical norms are insufficient for protecting women's human rights. A critical evaluation of current AI regulations is necessary to determine their effectiveness in protecting women from discriminatory bias in AI, as defined by international human rights standards.

[283] Andrew Burt, 'Ethical Frameworks for AI Aren't Enough' *Harvard Business Review* (9 November 2020) <hbr.org/2020/11/ethical-frameworks-for-ai-arent-enough >.
[284] Margarita Robles Carrillo, 'Artificial intelligence: From Ethics to law' (2020) 44:101937 Telecommunication Policy, 3 <doi.org/10.1016/j.telpol.2020.101937>.
[285] Carrillo (n 284) 5f.
[286] Ibid, 6.
[287] Ibid, 6f.

5.2. AI Regulations around the World

More than 37 countries have introduced discussions on legal frameworks dealing with AI-related concerns.[288] Amongst them are Australia, Brazil, Canada, China, the EU, India, Japan, Switzerland, the UK and the US.[289] Furthermore, regional AI strategy papers can be found in several other countries, including Indonesia, Peru, Saudia Arabia, and Taiwan. There is no standard approach to regulating AI. National jurisdictions try to find a balance between innovation and regulation.[290]

Various approaches to AI regulation have been proposed. The UK and Switzerland have taken similar approaches in not introducing standalone, comprehensive AI regulation but instead amending existing laws to accommodate AI.[291] The UK has been taking the road on a context-specific AI, meaning that not specific technologies or sectors will be evaluated on their risks, but more the general outcomes will be weighed between its opportunity and costs. The white paper of the UK states. that a '*cross sector regulator would introduce complexity and confusion, undermining and likely conflicting with the work of our existing expert regulators*'. However, the UK's current white paper also suggests a centralised function for coordination.[292]

The US has adopted a case-to-case approach, also avoiding a unified regulation. Several guidelines and regulations have been established in recent years.[293] About a dozen AI-related bills have been passed into law in 2023 to regulate smaller slieces of AI. Federal law for AI regulation is led by California, New York, and Florida, with each having a different focus.[294] A single state comprehensive AI framework was proposed under the name *California's*

[288] Nick Sherman, 'AI Regulations around the World' *Mind Foundry* (25 January 2024) <mindfoundry.ai/blog/ai-regulations-around-the-world>.
[289] Kostiantyn Ponomarov, 'Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview' *LegalNodes* (7 May 2024) <legalnodes.com/article/global-ai-regulations-tracker>.
[290] IAPP Research and Insights 'Global AI Law and Policy Tracker' *IAPPAI Governance Center* (last updated January 2024) <iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf>.
[291] Ponomarov (n 289).
[292] Itsiq Benizri, Arianna Evers, Shannon Togawa Mercer and Ali A. Jessani, 'A Comparative Perspective on AI Regulation' *LawFare* (17 July 2023) <lawfaremedia.org/article/a-comparative-perspective-on-ai-regulation>.
[293] Ponomarov (n 289).
[294] Nick Sherman, 'AI Regulations around the World' *Mind Foundry* (25 January 2024) <mindfoundry.ai/blog/ai-regulations-around-the-world>.

Assembly Bill 311. Companies accepting this proposal would be required to conduct an impact assessment of AI products with the possibility to opt out.[295]

The EU proposed the first comprehensive framework to regulate AI, approved by the European Parliament in March 2024.[296] The AI Act came into force at the end of May 2024 and will be implemented from 2025 onward.[297] The EU has chosen a risk approach with stricter rules for higher-risk applications. AI systems with unacceptable risk, such as the use of biometric data with sensitive characterises, including people's sexual orientation, are banned. High-risk applications, such as AI models in hiring or law enforcement, must fulfil certain obligations in areas such as safety, transparency and explainability. They must also obey privacy regulations and anti-discriminatory regulations. For lower-risk applications, users still need to be informed when interacting with AI tools. All AI tool operations in the EU must apply these rules. Critics raised concern about the loopholes and exemption for AI applications for military use, national security purposes, law enforcement and migration. To encourage innovation, AI tools purely developed for research, development and prototyping are exempt from the AI Act risk assessment.[298]

Coordination between the EU and the US exists – for example, in the discussion of an AI Code of Conduct at the US-EU Trade and Tech Council. Like the GDPR, the EU AI Act has extraterritorial reach; thus, other countries and international companies need to consider laws outside the regional territory.[299]

In March 2024, the Council of Europe decided upon a Convention on AI, intended to be the first binding agreement on AI on an international level.[300] The Council of Europe, with its 46 member states, has the task of upholding human rights, democracy, and the rule of law in

[295] Benizri (n 292).
[296] Shiona McCallum, Liv McMahon and Tom Singleton, 'MEPs approve world's first comprehensive AI law' *BBC* (13 March 2024) <bbc.com/news/technology-68546450>.
[297] Karen Gilchrist and Ruxandra Iordache, 'World's first major act to regulate AI passed by European lawmakers' *CNBC* (13 March 2024) <cnbc.com/2024/03/13/european-lawmakers-endorse-worlds-first-major-act-to-regulate-ai.html>.
[298] Elizabeth Gibney, 'What the EU's tough AI law means for research and ChatGPT' *Nature* (16 February 2024) <nature.com/articles/d41586-024-00497-8>.
[299] Benizri (n 292).
[300] Angela Müller and Matthias Spielkamp, 'Europarat: KI-Konvention wird Menschenrechte nicht angemessen schützen' *Algorithm Watch* (18 March 2024) <algorithmwatch.org/de/ki-konvention-ungenugend/>.

Europe. It also has member states outside of the EU. It has drafted the European Convention on Human Rights, and its ratification is a condition for new members joining the Council of Europe.[301] Members of the Council of Europe and non-members such as the US, Japan and Canada took part in the negotiations.[302] The international treaty is legally binding under international law, but only for states that sign it. Member states are free to sign or not, and non-member states have the option to sign the agreement, which is why states such as the US and Israel are part of the debate.[303]

The agreement aims to protect human rights, democracy and the rule of law from harmful AI applications.[304] The international convention's primary function is to bind states and other international organisations to become signing parties of the convention with the obligation to integrate the rules and principles into their legal systems. The Convention on AI does not place obligations on individuals and private organisations. It can be seen as a guiding framework at an international level and can impact states when they ratify the convention and translate the international rules into local laws and regulations.[305] Parties to the convention have two obligations as described in Art. 3:

> 'a. Each Party shall apply the Convention to the activities within the lifecycle of artificial intelligence systems undertaken by public authorities, or private actors acting on their behalf.
> b. Each Party shall address risks and impacts arising from activities within the lifecycle of artificial intelligence systems by private actors to the extent not covered in subparagraph (a) in a manner conforming with the object and purpose of the Convention.'[306]

[301] Angela Müller, 'Council of Europe creates rules for Artificial Intelligence' *Algorithm Watch* (20 May 2022, updated 16 May 2024) <algorithmwatch.org/en/artificial-intelligence-council-of-europe/>.
[302] Angela Müller, 'The Council of Europe's Convention on AI: No free ride for tech companies and security authorities!' *Algorithm Watch* (5 March 2024) <algorithmwatch.org/en/council-of-europe-ai-convention/>.
[303] Müller (n 301).
[304] Müller (n 302).
[305] Osman Gazi Güçlütürk, 'Understanding the Council of Europe's Draft Framework Convention on AI, Human Rights, Democracy, and Rule of Law' *Holistic AI* (17 January 2024) <holisticai.com/blog/europe-committee-artificial-intelligence-draft-framework-convention#:~:text=The%20main%20purpose%20of%20the,(1)%20of%20the%20DFC>.
[306] Council of Europe: Committee of Ministers, 'Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law' CETS No. 25 (17 May 2024) Article 3.

Three significant areas are not regulated under the Convention on AI by the Council of Europe: research and development activities (which shows the importance of innovation and exploration in the AI field), national security and national defence.[307]

Similar to the EU AI Act, the Convention on AI makes reference to a risk-based approach.[308] The parties should '*adopt or maintain measures for the identification, assessment, prevention and mitigation of risks posed by artificial intelligence systems*'.[309]

Such measures should:

> 'a. take due account of the context and intended use of artificial intelligence systems, in particular as concerns risks to human rights, democracy, and the rule of law;
> b. take due account of the severity and probability of potential impacts;
> c. consider, where appropriate, the perspectives of relevant stakeholders in particular persons whose rights may be impacted;
> d. apply iteratively throughout the activities within the lifecycle of the artificial intelligence system;
> e. include monitoring for risks, actual and potential impacts, and the risk management approach;
> f. include documentation of risks, actual and potential impacts, and the risk management approach;
> g. require, where appropriate, testing of artificial intelligence systems before making them available for first use and when they are significantly modified'.[310]

However, the Convention distinguishes the different AI systems and, therefore, does not prohibit high-risk systems.[311] Art. 16 (4) states that parties should ban or find other appropriate measures for such AI systems which are incompatible with the respect of human rights, democracy or the rule of law.[312]

The Convention decided upon principles related to activities within the lifecycle of AI systems. Among them are human dignity and individual autonomy, transparency and oversight, accountability and responsibility, equality and non-discrimination, privacy and

[307] Osman Gazi Güçlütürk, 'Understanding the Council of Europe's Draft Framework Convention on AI, Human Rights, Democracy, and Rule of Law' *Holistic AI* (17 January 2024) <holisticai.com/blog/europe-committee-artificial-intelligence-draft-framework-convention#:~:text=The%20main%20purpose%20of%20the,(1)%20of%20the%20DFC>.
[308] Güçlütürk (n 307).
[309] Council of Europe (n 306), Article 16 (1).
[310] Ibid, Article 16 (2).
[311] Güçlütürk (n 307).
[312] Council of Europe (n 306), Article 16 (4).

personal data protection, reliability and safe innovation.[313] Art. 10 addresses equality and non-discrimination as follows:

1. 'Each Party shall adopt or maintain measures with a view to ensuring that activities related to the lifecycle of artificial intelligence systems respect equality, including gender equality, and the prohibition of discrimination, as provided under applicable international and domestic law.
2. Each Party undertakes to adopt or maintain measures aimed at overcoming inequalities to achieve fair, just and equitable outcomes, in line with its applicable domestic and international human rights obligations, in relation to activities within the lifecycle of artificial intelligence systems.'[314]

The draft concerning equality, including gender equality and non-discrimination, refers to already existing human rights law at international and domestic levels, including constitutional law and jurisprudence, which should provide the basis for ensuring rights are guaranteed in the context of AI systems[315] The Explanatory Report also explains the drafters' reflection '*on the real and well-documented risk of bias that can constitute unlawful discrimination arising from the activities within the lifecycle of artificial intelligence systems*'.[316] Parties have '*to consider appropriate regulatory, governance, technical or other solutions*' for the different ways where bias can '*be incorporated into artificial intelligence systems at various stages throughout their lifecycle*'.[317] It is underlined that it is not enough to stop '*requiring that a person is not treated less favourably*' but that '*structural and historical inequalities*' should also be overcome.[318]

Art. 17 of the Convention addresses how the principles of non-discrimination are implemented as follows: '*The implementation of the provisions of this Convention by the Parties shall be secured without discrimination on any ground, in accordance with their international human rights obligations*.'[319] The meaning here is identical to other non-discrimination articles in international law, such as Art. 25 of the ICCPR, Art. 2 of the ICESCR, and Art. 14 of the ECHR.[320] All these articles cover '*non-discrimination grounds*

[313] Ibid, Article 6f.
[314] Ibid, Article 10.
[315] Council of Europe (n 25) para 71, 74.
[316] Ibid, para 75.
[317] Ibid, para 75.
[318] Ibid, para 77.
[319] Council of Europe (n 206), Article 17.
[320] Council of Europe (n 25) para 113.

which are linked to individuals' personal characteristics, circumstances or membership of a group'.[321]

The Convention has gained positive feedback along with concerns. First, the Convention does not cover the private sector in the same way as the public field. AI systems are often developed and deployed by private companies, which is critical for such systems' impact on human rights.[322] Signatory states have discretion over the level of AI safety measures companies must implement.[323] The European Network of National Human Rights Institutions (ENNHRI) flags a protection gap for human rights in this different approach for the private sector and the non-binding regulation.[324] Second, AI systems used for national security are not included under the protection of the Convention. Specific systems such as those '*used for surveillance, data collection, and decision-making processes aimed at countering perceived threats to national security could present significant risks to human rights, democracy and the rule of law*'.[325] Third, some essential elements are missing in the Convention, such as human oversight for AI systems, AI systems that are prohibited because of their high risk for human rights, democracy and the rule of law, and the absence of an independent oversight mechanism at an international level at the Council of Europe and domestic levels.[326] Finally, ENNHRI criticises the imprecise language in phrases such as '*where practicable*' or '*in accordance with domestic law*' which will make it challenging to enforce obligations.[327] Overall, the Convention has softened in its formulations in the hope of finding agreement among states and countries that are not members of the Council of Europe.[328]

The Council of Europe's Convention on AI represents an international step forward in regulating AI. However, with only 47 member states, the Council of Europe does not represent the entire world, and there is an absence of countries in the global south, which are

[321] Ibid, para 114.
[322] ENNHRI, 'Draft Convention on AI, Human Rights, Democracy and Rule of Law finalised: ENNHRI raises concerns' (20 March 2024) <ennhri.org/news-and-blog/draft-convention-on-ai-human-rights-democracy-and-rule-of-law-finalised-ennhri-raises-concerns/>.
[323] Müller (n 302).
[324] ENNHRI (n 322).
[325] Ibid.
[326] Ibid.
[327] Ibid.
[328] Müller (n 302).

excluded from this discussion. Therefore, it is vital that the UN finally made a move on discussions outside of ethical considerations to deal with AI systems.

On 21 March 2024, the United Nations General Assembly adopted the first resolution on the topic of AI.[329] The US announced shortly before, the submission of a draft resolution with the goal of convincing all 193 member states of the UN. One-hundred and twenty countries have been included in the draft proposal. Over 125 states co-sponsored the resolution, it passed in consensus without a further vote.[330]

Finding common ground is not easy, as interests worldwide are highly differentiated. The US has the main players in the AI market, notably on the business side concerning the development and financing of such systems. The EU has strict data privacy laws and recently agreed on the EU AI Act with a focus on upholding fundamental rights by addressing the risks of AI systems. Therefore, the focus of the EU will be more on the protection of users. States from the Global South struggle with AI advancement, as AI data is largely acquired in the Global North and is not always appropriate for other markets. An overwhelming majority of people without internet access live in the Global South, where many states fear being left behind.[331]

The UN Resolution reaffirms its commitment to the Charter of the United Nations and the Universal Declaration of Human Rights.[332] It recognises the different stages of lifecycles, including '*pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning*'.[333] AI systems are safe, secure and trustworthy if they are '*human-centric, reliable, explainable, ethical, inclusive, in full respect, promotion and protection of human rights and international law, privacy preserving, sustainable development oriented, and responsible*'.[334] The Resolution also recognises AI's dangers by stating that the design, development and deployment of AI

[329] UNGA, 'Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development' Seventy-eighth session Res 78/265 (21 March 2024).
[330] Annika Knauer, 'The First United Nations General Assembly Resolution on Artificial Intelligence' *EJIL:Talk!* (2 April 2024) <ejiltalk.org/the-first-united-nations-general-assembly-resolution-on-artificial-intelligence/>.
[331] Knauer (n 330).
[332] UNGA (n 329) para 1.
[333] Ibid, preambular para 2.
[334] Ibid, preambular para 2.

systems without consistency with international law pose a risk to reaching the Sustainable Development Goals, widen the digital gap between countries, reinforce inequalities that lead to bias and discrimination, undermine information integrity, and pose risks for the protection of human rights and fundamental freedoms. Furthermore, the Resolution urges a '*global consensus on safe, secure and trustworthy artificial intelligence systems*'. [335] It calls upon the member states to '*close the gender digital divide*' and '*to mainstream a disability, gender and racial equality perspective in policy decisions and the framework that guide them*'.[336]

The UN encourages member states and the private sector to develop and support approaches which work towards safe, secure and trustworthy AI systems.[337] It further encourages the private sector '*to adhere to applicable international and domestic laws and act in line with the United Nations Guiding Principles on Business and Human Rights*'.[338] This direct mention of the private sector is understood as a reinforcement of the central role of the private sector in implementing human rights. Companies need to ensure their products align with human rights laws.[339] However, the Resolution does not set international guidelines but promotes the development and implementation of domestic regulatory approaches and frameworks in line with national policies.[340]

The common ground that could be found in the UN Resolution is relatively minor. Compared to an earlier draft from December 2023, where misuse of AI was 'condemned', the final version only 'encourages' member states to protect individuals from misuse. Furthermore, states were 'called upon' to engage in AI governance actions, a phrase later weakened to 'encourage'. The UN's role in reaching a global consensus was softened in the conclusion of the final version. Compared to the draft, the only part more strongly worded concerns the inclusion of developing countries, where states are now 'called upon' for cooperation and assistance.[341]

[335] Ibid, preambular para 2
[336] Ibid, para 6.
[337] Ibid, para 3.
[338] Ibid, para 9.
[339] Baker McKenzie, 'International: The United Nations adopts its first resolution on AI' <insightplus.bakermckenzie.com/bm/data-technology/international-the-united-nations-adopts-its-first-resolution-on-ai#:~:text=On%2021%20March%202024%2C%20the,intelligence%20systems%20for%20sustainable%20development%22>.
[340] Baker McKenzie (n 339) para 6 (a).
[341] Knauer (n 330).

In conclusion, the UN Resolution commits member states to ensure a safe environment for AI systems to comply with international human rights obligations. The document underlines that states are aware of the dangers AI can pose to individuals. However, from a legal perspective, the Resolution is largely symbolic, as it involves no legal obligations.[342]

To summarise, progress in AI regulation can be observed at regional and international levels. Regionally, several governments have incorporated AI into their legal frameworks. The EU AI Act stands out as the only regulation with extraterritorial reach, although it does not protect citizens outside the EU.

At the international level, the Council of Europe's Convention on AI is currently the only legally binding instrument. While the Convention has some legal gaps, it represents a significant step towards internationally harmonised AI regulation. However, not all states worldwide are members of the Council of Europe, and many non-member countries have not participated in the discussions and are not going to sign the agreement.

The UNGA's adoption of the first resolution on AI is also crucial for fostering international dialogue on this subject. Although the resolution is non-binding and thus does not enforce protections against gender bias in AI under international human rights law, it remains an important discussion point.

The following section assesses whether international human rights law should play a role in regulating AI.

5.3. Conclusion

It is argued that there are three main reasons to regulate AI on a level of international law. First, from a technical and practical point of view, it is impossible to identify only one territorial law (e.g., data analysis or the supply of services in AI systems). AI has a universal scope and should therefore be dealt with on a global level. Second, state-by-state legislation is difficult in areas where different court decisions could take place. For example, a Chinese

[342] Ibid.

court decided that an article written by an AI system autonomously with a certain originality could lead to the recognition of the AI having author's rights. However, the European Patent Office refused to reach the same conclusion. Third, it is more coherent to have legal certainty and protection for everyone involved in AI systems (e.g., users, producers, and manufacturers). A general regulation is a better solution than partial legislation and helps reduce legal uncertainty.[343]

International law in the regulation of AI systems has its weaknesses. It was created to regulate relationships between states, which over time broadened to regulate interpersonal and private matters. It is crucial that AI is discussed on a multi-stakeholder level. Hence, AI cannot be addressed on an interstate level only. Furthermore, international law is based on agreement between states, but it is hard to find a compromise, especially on such a complex topic as AI, when there are different political and economic interests.[344]

At that point of the thesis, it could be questioned whether international law is the better option for regulating AI. It was elaborated so far, that states worldwide deal with AI in different ways. States' obligations under international law, and whether they can be enforced effectively to provide better protection for women's human rights, are outstanding questions.

States are obliged to respect, protect and fulfil all international human rights laws they are party to. Hence, governments must undertake measures to make domestic legislation compatible with international treaty obligations and duties.[345] However, there is only a limited court system to enforce international law. The International Court of Justice has limited jurisdiction and is not applicable to this case. The UN Security Council may use force but only in limited circumstances involving prior aggression or threat.[346]

States generally ensure their actions conform with international law because the party would otherwise be regarded negatively by the community of states. If states do not comply with

[343] Carrillo (n 284) 11,
[344] Ibid, 12.
[345] OHCHR, 'International Human Rights Law' <ohchr.org/en/instruments-and-mechanisms/international-human-rights-law#:~:text=By%20becoming%20parties%20to%20international,the%20enjoyment%20of%20human%20rights>.
[346] Malcolm Shaw, 'international law' *Britannica* (last updated 8 May 2024) <britannica.com/topic/international-law>.

international rules, they suffer credibility, which may prejudice future relations with other countries. Continuous violations would also jeopardise the value the system brings to the international community. International rules also provide a framework for interactions and a standard set of concepts for understanding them.[347] One way to safeguard international human rights law is the Universal Periodic Review (UPR), a mechanism of the Human Rights Council.[348] The Human Rights Council is an intergovernmental body of the UN. Forty-seven states are responsible for the promotion and protection of human rights.[349] Every 4.5 years, each UN Member State undergoes a review of its human rights situation by all the other states. Each state has to report on improvements in the human rights situation and receives recommendations from stakeholders and other states to further improve the situation.[350]

Significant variations exist in the respect and protection of human rights across different countries. However, international law demonstrates the importance of discussing certain subjects globally. AI and its implications for discrimination based on gender bias cannot be viewed independently from state to state. AI is used worldwide and has global implications, as shown in Chapter II and IV. Gender bias in AI poses significant societal challenges, particularly for women. Biases in AI systems manifest in various areas, such as advertisements, credit scoring, and facial recognition systems. These biases result in lower service quality, unfair resource allocation, and the reinforcement of harmful stereotypes. The urgency of addressing these biases is evident, prompting the question of whether current ethical frameworks effectively protect women's human rights.

Many ethical frameworks exist from different actors, including states, NGOs, and the private sector. These frameworks have common themes such as privacy, accountability, safety, transparency, fairness, human control of technology, and professional responsibility. Despite their prevalence, many frameworks lack enforceability and fail to address human rights concerns robustly. This issue underscores the need for clear, enforceable guidelines that align with human rights principles to mitigate AI's risks and harms effectively.

[347] Shaw (n 346).
[348] OHCHR, 'Universal Periodic Review' <ohchr.org/en/hr-bodies/upr/upr-home>.
[349] OHCHR, 'Human Rights Council' <ohchr.org/en/hrbodies/hrc/home>.
[350] OHCHR (n 349).

These problems lead to the question of whether women's human rights can be protected under international law. While there are global commitments to address gender bias in AI, binding agreements on a global level are still lacking. The discussion highlights the need for effective legal instruments and strategies to address these gaps.

Therefore, the necessity of international regulation is emphasised. While ethical norms are essential for guiding AI development and fostering discussion, they are insufficient for protecting women's human rights. Existing regulations, such as the EU AI Act and the Council of Europe’s Convention on AI, are steps towards harmonised AI regulation but have limitations, including non-participation by some countries. The UNGA’s non-binding resolution on AI highlights the importance of international dialogue but does not enforce protections against gender bias.

In summary, while progress has been made in AI regulation, substantial gaps and challenges remain. Effective protection against AI-induced biases requires coordinated global efforts, robust legal frameworks, and a strong emphasis on human rights principles.

BIBLIOGRAPHY

BOOKS

Boerei I. et al., *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)

Miller and Wendt (eds), *The Fourth Industrial Revolution and Its Impact on Ethics. Solving the Challenges of the Agenda 2030* (Springer 2021) <dx.doi.org/10.1007/978-3-030-57020-0_20>

Fredman S, 'Beyond the dichotomy of formal and substantive equality. Towards new definitions of equal rights' in I. Boerei jn 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 M, 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)

Hellum A and Aasen H, *Women's Human Rights: CEDAW in International, Regional and National Law' (Cambridge University Press 2013)*

Russell S and Norvig P, *Artificial Intelligence. A Modern Approach (4th edn, Pearson Education Limited, 2022)*

Taulli T, *Artificial Intelligence Basics. A non-Technical Introduction (Apress 2019)*

CONFERENCE PAPERS

Adams R and Ni Loideain N, 'Addressing Indirect Discrimination and Gender Stereotypes in AI Virtual Personal Assistants: The Role of International Human Rights Law' (22 May 2019), Annual Cambridge International Law Conference 2019, New

Technologies: New Challenges for Democracy and International Law <dx.doi.org/10.2139/ssrn.3392243>

Bietti E, 'From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy' (1 December 2019, revised 16 November 2021) Draft - Final Paper Published in the Proceedings to ACM FAT* Conference (FAT* 2020) <papers.ssrn.com/sol3/papers.cfm?abstract_id=3513182>

Bolukbasi T, Chang K, Zou J, Saligrama V, Kalai A, 'Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings' (2016) NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems <doi.org/10.48550/arXiv.1607.06520>

Feng Y, Shah C, 'Has CEO Gender Bias Really Been Fixed? Adversarial Attacking and Improving Gender Fairness in Image Search' (2022) Proceedings of the AAAI conference on artificial intelligence <yunhefeng.me/material/Bias_in_Image_Search_AAAI22_Feng.pdf>

Raji I.D, and Buolamwini R, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products' (2019) Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19), Association for Computing Machinery <doi.org/10.1145/3306618.3314244>

JOURNALS

Adadi A, and Berradda M, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' (12 October 2018) 6 Institute of Electrical and Electronics Engineers (IEEE) 52138 – 52160 <doi:10.1109/ACCESS.2018.2870052>

Carrillo M.R, 'Artificial intelligence: From Ethics to law' (2020) 44:101937 Telecommunication Policy 1 – 16 <doi.org/10.1016/j.telpol.2020.101937>

Cirillo D and others, ‘Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare’ (2020) 3, 81, npj Digital Medicine 1 – 11 <doi.org/10.1038/s41746-020-0288-5>

Fredman S, ‘Substantive equality revisited’ (July 2016) 14,3 International Journal of Constitutional Law 712 – 738 <doi.org/10.1093/icon/mow043>

Gerards J., ‘The fundamental rights challenges of algorithms’ (2019) 37/3 Netherlands Quarterly of Human Rights 205 – 209 <doi.org/10.1177/0924051919861773>

Ntoutsi E and others, ‘Bias in data-driven artificial intelligence systems – An introductory survey’ (2020) 10, 3 WIREs Data Mining and Knowledge Discovery 1 – 14 <doi.org/10.1002/widm.1356>

Prates M, Avelar P, Lamb L, ‘‘Assessing gender bias in machine translation: a case study with Google Translate’ Neural Computing and Applications (2019) 6363 – 6381 <doi.org/10.48550/arXiv.1809.02208>

Regitz-Zagrosek V and others, ‘Gender in cardiovascular diseases: impact on clinical manifestations, management and outcomes’ (2016) 37 European Heart Journal 24 – 34 <doi:10.1093/eurheartj/ehv598>

Terrell J, Kofink A, Middleton J, Rainear C, Murphy-Hill E, Parnin C, Stallings J., ‘Gender differences and bias in open source: pull request acceptance of women versus men’ (2017) 3:e111 PeerJ Computer Science 1 – 30 <doi.org/10.7717/peerj-cs.111>

Vlasceanu M and Amodio D, ‘Propagation of societal gender inequality by internet search algorighms’ (12 July 2022) 119, 29 Proceedings of the National Academy of Sciences (PNAS) 1 – 8 <doi.org/10.1073/pnas.2204529119>

Zednik C, ‘Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence’ (2021) 34 Philosophy & Technology, 265 – 288 <doi.org/10.1007/s13347-019-00382-7>

INTERNATIONAL LEGAL DOCUMENTS

Council of Europe: Committee of Ministers, 'Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law' CETS No. 25 (17 May 2024)

European Parliament, 'Corrigendum to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ …… of the European Parliament and the of the Council laying down harmonised rules on artificial intelligence (Final draft, AI Act)' <europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf>

OECD Council 0449 of 22 May 2019, amended 3 May 2024, 'Recommendation of the Council on Artificial Intelligence' (2024)

UNGA, 'Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development' Seventy-eighth session Res 78/265 (21 March 2024)

REPORTS

Council of Europe, 'Explanatory Report to the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law' Treaty Series – No. 225 (2024)

Fjeld J, Achten N, Hilligoss H, Nagy A and Srikumar M, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI.' (2020) Berkman Klein Center for Internet & Society <nrs.harvard.edu/urn-3:HUL.InstRepos:42160420>

International Women's Development Agency and Womens Action For Voice and Empowerment, CEDAW at a Glance, 1, <iwda.org.au/assets/files/CEDAW-at-a-Glance.pdf>

Jeffrie N, 'The Mobile Gender Gap Report 2024' *GSMA* (May 2024) <gsma.com/r/wp-content/uploads/2024/05/The-Mobile-Gender-Gap-Report-2024.pdf>

OECD, 'Explanatory Memorandum on the Updated OECD Definition of an AI System' OECD Artificial Intelligence Papers No. 8 (OECD Publishing 2024)

World Economic Forum, 'Global Gender Gap Report 2021' (March 2021) 60 <www3.weforum.org/docs/WEF_GGGR_2021.pdf>

World Economic Forum, 'Global Gender Gap Report 2023' (June 2023) <www3.weforum.org/docs/WEF_GGGR_2023.pdf>

TREATIES AND CONVENTIONS

African (Banjul) Charter on Human and Peoples' Rights (adopted 27 June 1981, entered into force 21 October 1986) OAU Doc. CAB/LEG/67/3 rev. 5, 21 I.L.M. 58 (1982)

American Convention on Human Rights (adopted 22 November 1969, entered into force 18 July 1978)

Charter of the Organization of American States (signed in 1948 and amended latest by the Protocol of Managua 1993)

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

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)

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR)

Inter-American Convention on the Prevention, Punishment and Eradication of Violence against Women "Convention of Belem do Para" (adopted 09 June 1994, entered into force 05 March 1995)

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

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

Protocol to the African Charter on Human and Peoples' Rights on the Rights of Women in Africa (adopted 01 July 2003, entered into force 25 November 2005)

UN Beijing Declaration and Platform for Action (adopted at the Fourth World Conference on Women, 4-15 September 1995)

UNGA, The future we want, A/RES/66/288 (adopted 27 July 2012)

UNGA, Transforming our world : the 2030 Agenda for Sustainable Development, A/RES/70/1 (adopted 25 September 2015)

UNGA, United Nations Millennium Declaration, UNGA A/RES/55/2 (adopted 18 September 2000)

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

Vienna Declaration and Programme of Action (adopted by the World Conference on Human Rights in Vienna on 25 June 1993) A/CONF.157/23, UN General Assembly

UNITED NATIONS DOCUMENTS

Committee for the Elimination of All Forms of Discrimination against Women (CEDAW), 'General Recommendation No 25' (2004)

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

Commissioned Report Gender Stereotyping as a Human Rights Violation' (October 2013)

OHCHR, 'Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment' Seventy-third session UN Doc A/73/348 (29 August 2018)

OHCHR 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' A/73/348 (August 2018)

OHCHR Report, 'Promotion, protection and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective' Thirty-fifth session A/HRC/35/9 (June 2017)

UN 'Millennium Development Goals Report 2015'

UN Women Working Paper, 'The digital revolution: Implications for Gender Equality and Women's Rights 25 Years after Beijing' (August 2020)

West M, Kraut R and Chew H, 'I'd blush if I could: closing gender divides in digital skills through education' *UNESCO and EQUALS Skills Coalition* (2019) <unesdoc.unesco.org/ark:/48223/pf0000367416.locale=en>

Women's Rights are Human Rights' HR/PUB/14/2 (New York, Geneva, 2014)

WEBSITES

Ackermann N, 'Artificial Intelligence Framework: A Visual Introduction to Machine Learning and AI' *Towards Data Science* (13 December 2018) <towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87>

Agomuoh F and Larsen L, 'ChatGPT: the latest news, controversies, and tips you need to know' Digital Trends (25 September 2023) <digitaltrends.com/computing/how-to-use-openai-chatgpt-text-generation-chatbot/>

AI Ethics Lab, 'Toolbox: Dynamics of AI Principles', <aiethicslab.com/big-picture/>

AI for Good blog, 'Bridging the AI gender gap: Why we need better data for an equal world' *AI for Good* (25 September 2020) <aiforgood.itu.int/bridging-the-ai-gender-gap-why-we-need-better-data-for-an-equal-world/>

AI for Good blog, 'Gender bias is a threat to future Artificial Intelligence (AI) applications: Opinion' *AI for Good* (17 September 20219) <aiforgood.itu.int/gender-bias-is-a-threat-to-future-artificial-intelligence-ai-applications-opinion/>

Algorithm Watch, 'AI Ethics Guidelines Global Inventory' (April 2020) <inventory.algorithmwatch.org/?sfid=172>

Algorithm Watch, ''In the realm of paper tigers – exploring the failing of AI ethics guidelines' (28 April 2020) <algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/>

Audiovisual Library of International Law, 'Convention on the Elimination of All forms of Discrimination against Women', <legal.un.org/avl/ha/cedaw/cedaw.html#:~:text=The%20Convention%20entered%20into%20force,gender%2Dbased%20discrimination%20against%20women>

Baker McKenzie, 'International: The United Nations adopts its first resolution on AI' <insightplus.bakermckenzie.com/bm/data-technology/international-the-united-nations-adopts-its-first-resolution-on-ai#:~:text=On%2021%20March%202024%2C%20the,intelligence%20systems%20for%20sustainable%20development%22>

Berlin Institute of Health, 'Podcast Folge 13 – Benachteiligt die Künstliche Intelligenz weibliche Patienten?' (18 January 2020), <bihealth.org/de/aktuell/benachteiligt-die-kuenstliche-intelligenz-weibliche-patienten/>

Benizri I, Evers A, Mercer S.T and Jessani A.A, 'A Comparative Perspective on AI Regulation' *LawFare* (17 July 2023) <lawfaremedia.org/article/a-comparative-perspective-on-ai-regulation>

Bietti E, 'From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy' (29 January 2020, revised 16 November 2021) In Proceedings of ACM FAT* Conference (FAT* 2020), <papers.ssrn.com/sol3/papers.cfm?abstract_id=3513182>

Brown S, 'Machine learning, explained' *MIT Sloan School of Management* (21 April 2021) <mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>

Bualamwini J, ‘Artificial Intelligence Has a Problem With Gender and Racial Bias. Here’s How to Solve It’ *Time* (7 February 2019) <time.com/5520558/artificial-intelligence-racial-gender-bias/>

Buolamwini J., ‘Voicing Erasure – A Spoken Word Piece Exploring Bias in Voice Recognition Technology’ *Algorithmic Justice League* <ajl.org/voicing-erasure>

Burt A, ‘’Ethical Frameworks for AI Aren’t Enough’ *Harvard Business Review* (9 November 2020) <hbr.org/2020/11/ethical-frameworks-for-ai-arent-enough >

Cambridge Dictionary, ‘Bias’ <dictionary.cambridge.org/de/worterbuch/englisch/bias>

Cambridge Dictionary, ‘equality’, <dictionary.cambridge.org/de/worterbuch/englisch/equality>

Castelvecchi D, ‘Is facial recognition too biased to be let loose?’ *Nature* (18 November 2020) <nature.com/articles/d41586-020-03186-4#ref-CR5>

Copeland M, ‘What’s the Difference Between Artificial Intelligence, Machine Learning and Deep Learning?’ *Nvidia Blog* (29 June 2016) <blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/>

Dastin J, ‘Amazon scraps secret AI recruiting tool that showed bias against women’ *Reuters* (11 October 2018) <reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

ECCHR, ‘Definition Hard law/soft law’, <ecchr.eu/en/glossary/hard-law-soft-law/>

ENNHRI, ‘Draft Convention on AI, Human Rights, Democracy and Rule of Law finalised: ENNHRI raises concerns’ (20 March 2024) <ennhri.org/news-and-blog/draft-convention-on-ai-human-rights-democracy-and-rule-of-law-finalised-ennhri-raises-concerns/>

Equality Now, ‘Chat-GPT-4 Reinforces Sexist Stereotypes By Stating A Girl Cannot “Handle Technicalities and Numbers” In Engineering’ (23 March 2023) <equalitynow.org/news_and_insights/chatgpt-4-reinforces-sexist-stereotypes/>

European Parliament, 'The Universal Declaration of Human Rights and its relevance for the European Union', <europarl.europa.eu/RegData/etudes/ATAG/2018/628295/EPRS_ATA(2018)628295_EN.pdf>

Feast J, '4 Ways to Address Gender Bias in AI' *Harvard Business Review* (20 November 2019), <hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai;%20>

Fessler L, 'We tested bots like Siri and Alexa to see who would stand up to sexual harassment' *QUARTZ* (22 February 2017) <qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/>

Fluid Blog, 'Chat GPT and means of payment: Learn how AI is impacting the financial industry' *Dock.Tech* (15 May 2023) <dock.tech/en/fluid/blog/tech/gpt-chat/>

Gibney E, 'What the EU's tough AI law means for research and ChatGPT' *Nature* (16 February 2024) <nature.com/articles/d41586-024-00497-8>

Gilchrist K and Iordache R, 'World's first major act to regulate AI passed by European lawmakers' *CNBC* (13 March 2024) <cnbc.com/2024/03/13/european-lawmakers-endorse-worlds-first-major-act-to-regulate-ai.html>

Güçlütürk O.G, 'Understanding the Council of Europe's Draft Framework Convention on AI, Human Rights, Democracy, and Rule of Law' *Holistic AI* (17 January 2024) <holisticai.com/blog/europe-committee-artificial-intelligence-draft-framework-convention#:~:text=The%20main%20purpose%20of%20the,(1)%20of%20the%20DFC>

Haas L, Gießler S, Thiel V, 'In the realm of paper tigers – exploring the failings of AI ethics guidelines' *Algorithm Watch* (28 April 2020) <algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/>

Hao K, 'An AI saw a cropped photo of AOC. It autocompleted her wearing a bikini.' *MIT Technology Review* (29 January 2021) <technologyreview.com/2021/01/29/1017065/ai-image-generation-is-racist-sexist/>

Heikkilä M, 'The viral AI avatar app Lensa undressed me—without my consent' *MIT Technology Review* (12 December 2022) <technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/>

IAPP Research and Insights 'Global AI Law and Policy Tracker' *IAPPAI Governance Center* (last updated January 2024) <iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf>

IBM Cloud Education, 'Artificial Intelligence (AI)' *IBM Cloud* (3 June 2020) <ibm.com/cloud/learn/what-is-artificial-intelligence>

Kayser-Bril N, 'Automated discrimination: Facebook uses gross stereotypes to optimize ad delivery' *Algorithm Watch* (18 October 2020), <algorithmwatch.org/en/automated-discrimination-facebook-google/>

Kayser-Bril N, 'Ethical guidelines issued by engineers' organization fail to gain traction' *Algorithm Watch* (3 October 2019) <algorithmwatch.org/en/ieee-ethically-aligned-design-guidelines-fail-to-gain-traction/>

Knauer A, 'The First United Nations General Assembly Resolution on Artificial Intelligence' *EJIL:Talk!* (2 April 2024) <ejiltalk.org/the-first-united-nations-general-assembly-resolution-on-artificial-intelligence/>

Jackson G, 'The female problem: how male bias in medical trials ruined women's health' *The Guardian* (13 November 2019) <theguardian.com/lifeandstyle/2019/nov/13/the-female-problem-male-bias-in-medical-trials>

Langston J, 'Who's a CEO? Google image results can shift gender biases' *University of Washington News* (9 April 2015) <washington.edu/news/2015/04/09/whos-a-ceo-google-image-results-can-shift-gender-biases/>

Madgavkar A, 'A conversation on artificial intelligence and gender bias' *McKinsey & Company* (7 April 2021) <mckinsey.com/featured-insights/asia-pacific/a-conversation-on-artificial-intelligence-and-gender-bias>

Manyika J, Silberg J and Presten B, 'What Do We Do About the Biases in AI? *Harvard Business Review* (25 October 2019) <hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>

Mauro G and Schellmann H, ''There is no standard': investigation finds AI algorithms objectify women's bodies' *The Guardian* (8 February 2023) <theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies>

McCallum S, McMahon L and Singleton T, 'MEPs approve world's first comprehensive AI law' *BBC* (13 March 2024) <bbc.com/news/technology-68546450>

Medecins Sans Frontières, 'The Practical Guide to Humanitarian Law. Special Rapporteurs on Human Rights', <guide-humanitarian-law.org/content/article/3/special-rapporteurs/>

Michot S, Mollen A, Schiller A.L, Wulf J, 'Algorithmenbasierte Diskriminierung: Warum Antidiskriminierungsgesetze jetzt angepasst werden müssen' *Algorithm Watch, Digital Autonomy Hub* (Februry 2022) <algorithmwatch.org/de/wp-content/uploads/2022/02/DAH_Policy_Brief_5.pdf>

Müller A, 'Council of Europe creates rules for Artificial Intelligence' *Algorithm Watch* (20 May 2022, updated 16 May 2024) <algorithmwatch.org/en/artificial-intelligence-council-of-europe/>

Müller A, 'The Council of Europe's Convention on AI: No free ride for tech companies and security authorities!' *Algorithm Watch* (5 March 2024) <algorithmwatch.org/en/council-of-europe-ai-convention/>

Müller A and Spielkamp M, 'Europarat: KI-Konvention wird Menschenrechte nicht angemessen schützen' *Algorithm Watch* (18 March 2024) <algorithmwatch.org/de/ki-konvention-ungenugend/>

OHCHR, 'High Commissioner', <ohchr.org/en/about-us/high-commissioner>

OHCHR, 'Human Rights Council' <ohchr.org/en/hrbodies/hrc/home>

OHCHR, International Bill of Human Rights. A brief history, and the two International Covenants <ohchr.org/en/what-are-human-rights/international-bill-human-rights>

OHCHR, 'International Human Rights Law' <ohchr.org/en/instruments-and-mechanisms/international-human-rights-law#:~:text=By%20becoming%20parties%20to%20international,the%20enjoyment%20of%20human%20rights>

OHCHR, 'Universal Periodic Review' <ohchr.org/en/hr-bodies/upr/upr-home>

Oxford Reference, 'Environmental Law, Soft vs. Hard', <oxfordreference.com/display/10.1093/acref/9780190622664.001.0001/acref-9780190622664-e-303#:~:text=Hard%20law%2C%20such%20as%20treaties,but%20may%20be%20politically%20binding>

Oxford Reference, 'Overview soft law', <oxfordreference.com/display/10.1093/oi/authority.20110803100516251>

Photopoulos J, 'Fighting algorithmic bias in artificial intelligence' *physicsworld* (4 May 2021) <physicsworld.com/a/fighting-algorithmic-bias-in-artificial-intelligence>

Ponomarov K, 'Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview' *LegalNodes* (7 May 2024) <legalnodes.com/article/global-ai-regulations-tracker>

Reagan M, 'Understanding Bias and Fairness in AI Systems. An illustrated introduction to some of the basic concepts of a crucial problem.' *Towards Data Science* (25 March 2021) <towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3>

Refworld, 'UN Office of the High Commissioner for Human Rights (OHCHR)', <refworld.org/document-sources/un-office-high-commissioner-human-rights-ohchr>

Selig J, 'What Is Machine Learning? A Definition.' *expert.ai* (14 March 2022) <expert.ai/blog/machine-learning-definition/>

Shaw M, 'international law' *Britannica* (last updated 8 May 2024) <britannica.com/topic/international-law>

Sherman N, 'AI Regulations around the World' *Mind Foundry* (25 January 2024) <mindfoundry.ai/blog/ai-regulations-around-the-world>

Simonite T, 'AI Is the Future – But Where Are the Women? *WIRED* (17 August 2018) <wired.com/story/artificial-intelligence-researchers-gender-imbalance/>

Smith G, and Rustagi I, 'When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity' *Stanford Social Innovation Review* (31 March 2021) <ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity#>

Stanford Encycolpedia of Philosophy, 'Equality' (26 April 2021) <plato.stanford.edu/entries/equality/#FormEqua>

Sustainable Development Knowledge Platform, 'Sustainable Development Goals' <sustainabledevelopment.un.org/topics/sustainabledevelopmentgoals>

Sustainable Development Knowledge Platform, 'United Nations Conference on Sustainable Development, Rio+20', <sustainabledevelopment.un.org/rio20.html>

Teigland J, 'Why we need to solve the issue of gender bias before AI makes it worse' *Ernst & Young* (2 April 2019) <ey.com/en_be/wef/why-we-need-to-solve-the-issue-of-gender-bias-before-ai-makes-it>

Thiel V, '"Ethical AI guidelines": Binding commitment or simply window dressing?' *Algorithm Watch* (20 June 2019) <algorithmwatch.org/en/ethical-ai-guidelines-binding-commitment-or-simply-window-dressing/>

Tiku N, Schaul K and Chen S. Y, 'This is how AI image generators see the world' *The Washington Post* (1 November 2023) <washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/>

UNESCO, 'Artificial Intelligence', <unesco.org/en/artificial-intelligence/recommendation-ethics?hub=32618>

UNESCO, 'Artificial intelligence in education', <unesco.org/en/digital-education/artificial-intelligence#:~:text=UNESCO%20is%20committed%20to%20supporting,human%2Dcentred%20approach%20to%20AI>

United Nations Entity for Gender Equality and the Empowerment of Women, 'The Beijing Platform for Action Turns 20', <beijing20.unwomen.org/en/about>

United Nations, 'Millennium Development Goals and Beyond 2015', <un.org/millenniumgoals/gender.shtml>

United Nations, 'The Foundation of International Human Rights Law', <un.org/en/about-us/udhr/foundation-of-international-human-rights-law>

United Nations, 'Universal Declaration of Human Rights, Women Who Shaped the Declaration', <un.org/en/about-us/universal-declaration-of-human-rights>

UN Sustainable Development Group, 'UN Charter-based institutions including special procedures', <unsdg.un.org/2030-agenda/strengthening-international-human-rights/un-special-procedures>

UN Women, 'World Conferences on Women' <unwomen.org/en/how-we-work/intergovernmental-support/world-conferences-on-women>

Vigdor N, 'Apple Card Investigated After Gender Discrimination Complaints: A prominent software developer said on Twitter that the credit card was "sexist" against women applying for credit' *The New York Times* (10 November 2019) <nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html>

Weber M, 'Das ist falsch verstandene Emanzipation', *Der Tagesspiegel* (11 May 2016) <tagesspiegel.de/themen/frau-und-mann/frau-das-ist-falsch-verstandene-emanzipation/13575062.html>

Wong J.C, 'Women considered better coders – but only if they hide their gender' *The Guardian* (12 February 2016) <theguardian.com/technology/2016/feb/12/women-considered-better-coders-hide-gender-github>

TABLES

<washingtonpost.com/technology/interactive/2023/ai-generated-images-bias-racism-sexism-stereotypes/>

Table 9: Mauro G and Schellmann H, ‘‘There is no standard’: investigation finds AI algorithms objectify women’s bodies’ *The Guardian* (8 February 2023) <theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies>

Table 10: Bolukbasi T, Chang K, Zou J, Saligrama V, Kalai A, ‘Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings’ (2016) NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems <doi.org/10.48550/arXiv.1607.06520>

Table 11: Bolukbasi T, Chang K, Zou J, Saligrama V, Kalai A, ‘Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings’ (2016) NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems <doi.org/10.48550/arXiv.1607.06520>

Last verification of online resources: June 20, 2024