The Right to Equality and Non-discrimination in Artificial Intelligence: Legal Issues, Challenges, and Prospects

by

Sare Karacan

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for
the Degree of

Master of Laws

in

Public Law

KOÇ
ÜNİVERSİTESİ

July 18, 2023

The Right to Equality and Non-discrimination in Artificial Intelligence: Legal Issues, Challenges, and Prospects

Koç University

Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a master's thesis by

Sare Karacan

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Dr. Bertil Emrah Oder

_____

Assist. Prof. Yiğit Sayın

_____

Assist. Prof. Nafiye Yücedağ Keskin

Date: 18.07.2023

To my nearest and dearest ones:

my family

ABSTRACT

The Right to Equality and Non-discrimination in Artificial Intelligence: Legal Issues, Challenges, and Prospects

Sare Karacan

Master of Laws in Public Law

July 18, 2023

Artificial intelligence (AI) is inducing a transformation of the world and human lives. One element of this transformation is the impact of artificial intelligence on human rights, and one critical concern within human rights is its discriminatory impacts. The discriminatory impacts of artificial intelligence require researching the matter in the context of the right to equality and non-discrimination. The first chapter of this study handles the equality and discrimination related matters in artificial intelligence as a human rights law problem within the scope of the right to equality and non-discrimination by going over relevant definitions and notions with the aim to integrate its technical and ethical facets with the legal perspective. The second chapter examines how discrimination through artificial intelligence emanates by scrutinizing different areas of artificial intelligence use and examples of discrimination through artificial intelligence under the discrimination grounds established in human rights law. The second chapter later addresses discrimination through artificial intelligence under the existing international human rights law. It explores different potentials for positive and procedural obligations and aims to visualize discrimination through artificial intelligence by building on the discrimination test developed by the European Court of Human Rights (ECtHR). Looking at the existing international human rights law also provides a bridge to examining how specific regulations for artificial intelligence may address discrimination through artificial intelligence. The third chapter examines what the regulatory works on artificial intelligence promise for the right to equality and non-discrimination in the context of artificial intelligence by exploring both non-binding and binding documents, including those from the Council of Europe (CoE) and the European Union (EU), in light of the principles for artificial intelligence, including transparency and explainability, accountability, human control and oversight, and security and safety.

Keywords: artificial intelligence, the right to equality and non-discrimination, artificial intelligence fairness, algorithmic bias, human rights law, artificial intelligence regulation

ÖZETÇE

Yapay Zeka Bağlamında Eşitlik Hakkı ve Ayrımcılık Yasağı:

Hukuki Durumlar, Zorluklar ve Olasılıklar

Sare Karacan

Kamu Hukuku, Yüksek Lisans

18 Temmuz 2023

Yapay zekâ, dünya ve insan yaşamında bir dönüşümü tetiklemektedir. Bu dönüşümün bir unsuru, yapay zekanın insan hakları üzerindeki etkisidir ve insan hakları bağlamında yapay zekanın ayrımcı etkileri önem arz eden bir endişe kaynağıdır. Yapay zekanın ayrımcı etkileri, konunun eşitlik hakkı ve ayrımcılık yasağı bağlamında araştırılmasını gerekli kılmaktadır. Çalışmanın birinci bölümü, yapay zekada eşitlik ve ayrımcılıkla ilgili konuları, bir insan hakları hukuku sorunu olarak ele almak suretiyle eşitlik hakkı ve ayrımcılık yasağı kapsamında, ilgili kavramları irdeleyerek, meselenin teknik ve etik yönlerini hukuki bakış açısıyla bütünleştirmek amacıyla ele almaktadır. İkinci bölüm ise ilk olarak, yapay zekanın farklı kullanım alanlarını ve yapay zekâ dolayısıyla gerçekleşen ayrımcılık örneklerini insan hakları hukukunda belirlenen ayrımcılık nedenleri çerçevesinde irdeleyerek yapay zekâ dolayısıyla ayrımcılığın nasıl meydana gelebileceğini incelemektedir. Ayrıca ikinci bölüm, yapay zekâ dolayısıyla gerçekleşen ayrımcılığı, mevcut uluslararası insan hakları hukuku kapsamında, pozitif ve usule ilişkin yükümlülükler için farklı potansiyelleri araştırarak ve Avrupa İnsan Hakları Mahkemesi'ndeki ayrımcılık testini temel alarak tahayyül etmeyi amaçlamaktadır. Mevcut uluslararası insan hakları hukukuna bakmak, yapay zekaya özgü düzenlemelerin yapay zekâ yoluyla ayrımcılığı nasıl ele alabileceğini incelemek için de bir köprü sağlamaktadır. Üçüncü bölüm, yapay zekaya dair şeffaflık ve açıklanabilirlik, hesap verebilirlik, insan kontrolü ve gözetimi, güvenlik ve emniyet gibi çeşitli ilkeler ışığında hem bağlayıcı olmayan hem de bağlayıcı nitelikteki düzenleyici belgeleri inceleyerek, yapay zekâ bağlamında eşitlik hakkı ve ayrımcılık yasağının korunması için Avrupa Konseyi ve Avrupa Birliği'nin çalışmaları başta olmak üzere yapay zekâya yönelik düzenleyici belge etütlerinin neler vaat ettiğini incelemektedir.

Anahtar sözcükler: yapay zekâ, eşitlik hakkı ve ayrımcılık yasağı, yapay zekada hakkaniyet, algoritmik önyargı, insan hakları hukuku, yapay zeka düzenlemeleri

ACKNOWLEDGEMENTS

Writing a thesis is a journey that can only be accomplished with the support of many people. First and foremost, no words are enough to express my gratitude to Prof. Dr. Bertil Emrah Oder, who encouraged me from the day I asked her to be my advisor when I presented potential topics until the end of my study. She has made me believe in myself throughout the writing process. I am grateful for her guidance, feedback, support and for her contributions throughout this research.

Furthermore, I would like to express my deepest gratitude to Prof. Dr. H. Burak Gemalmaz for encouraging me to work hard and meticulously, endorsing my interest in researching artificial intelligence, and supporting my thesis process since day one.

I would also like to thank my defense committee members Assistant Professors Dr. Yiğit Sayın and Dr. Nafiye Yücedağ Keskin for reading and commenting on my thesis.

I am extremely grateful to Assistant Professor Dr. İ. Mert Ertan for giving attention to my thesis and feedback on my writing, which have led to important contributions, and for sharing wisdom on academia and life anytime I needed advice.

Thanks should also go to my research assistant colleagues. Zeynep Tanyeri had endured an immense workload in the last months of writing and listened to me when I needed a talk. Aybike Yılmaz has shared her ideas and academic insight and provided motivational support; writing would be more burdensome without those funny posts. Zeynep Günler has been a companion in doing many things, we had valuable conversations, and she has always motivated me for studying. Working and sharing the academic struggle with each of you has been a pleasure.

I would be remiss in not mentioning Esra Özcan, Executive Director at Koç Law School, for her generous help in arranging all of the processes during the LL.M. studies.

I must thank my dearest friends Hediye Erzurumlu for inspiring me with her discipline and passion for her work, and for our long study sessions and therapeutic video chats; Melike Arıkanlı, for always being there since our childhood years, and now in adulthood listening me for hours on academic life.

Lastly, my family, whose support I feel every second of my life; describing my gratitude for you is unattainable. Thanks to my mom for helping me stay sane for all of my life and particularly during this study, my dad for guiding me through life and being the source of my interest in technology, and my older brother Abdullah Enes Karacan for inspiring me to believe in myself and my thesis.

TABLE OF CONTENTS

ABBREVIATIONS

AI	Artificial Intelligence
AI HLEG	The High-Level Expert Group on Artificial Intelligence
AI-Ethics	Artificial Intelligence Ethics
AI-Lifecycle	Artificial Intelligence Lifecycle
AI-System	Artificial Intelligence System
CAHAI	Ad hoc Committee on Artificial Intelligence
CAI	Committee on Artificial Intelligence
CEDAW	Convention on the Elimination of All Forms of Discrimination Against Women
CEPEJ	European Commission for the Efficiency of Justice
CEPEJ Charter	European ethical Charter on the use of Artificial Intelligence in Judicial Systems and their Environment
CoE	Council of Europe
CoE Draft Convention on AI	[Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
CRC	Convention on the Rights of the Child
CRMW	International Convention on the Protection of the Rights of All Migrant Workers and Members of their Families
CRPD	Convention on the Rights of Persons with Disabilities
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
ESC	European Social Charter
EU	European Union
EU AI Act Proposal	Proposal for a Regulation of The European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

EU Fundamental Rights Charter	Charter of Fundamental Rights of the European Union
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
ICCPR	International Covenant on Civil and Political Rights
ICERD	International Convention on the Elimination of All Forms of Racial Discrimination
ICESCR	International Covenant on Economic, Social and Cultural Rights
IEEE	The Institute of Electrical and Electronics Engineers
MEPs	Members of the European Parliament
NLP	Natural Language Processing
OECD	Organisation for Economic Co-operation and Development
OECD AI Recommendation	OECD Recommendation of the Council on Artificial Intelligence
TFEU	Treaty on European Union and the Treaty on the Functioning of the European Union
Toronto AI Declaration	The Toronto Declaration Protecting the Right to Equality and Non-discrimination in Machine Learning Systems
U.S.	United States of America
UK	United Kingdom
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNESCO AI Recommendation	UNESCO Recommendation on the Ethics of Artificial Intelligence
WHO	World Health Organization

INTRODUCTION

The human rights impact of artificial intelligence constitutes a worrying matter.[1] The discriminatory aspect of artificial intelligence is among the human rights concerns for artificial intelligence, reflected in the legal realm within the right to equality and non-discrimination. Artificial intelligence may lead to various positive and negative impacts when considered within the scope of the right to equality and non-discrimination.[2] Artificial intelligence based *automated decision making* mechanisms may result in discrimination in areas including employment,[3] criminal justice,[4] healthcare[5], and credit score determination[6].

Discrimination in the context of artificial intelligence is mainly considered an ethical problem.[7] Ethical artificial intelligence is commonly discussed, but the ethical approach is not enough. At the point where there is an interference to human rights, the law comes into play.[8] Although it is clear that ethical evaluation is of indisputable importance, if these discussions remain only in the ethical realm, it will not be enough to prevent discrimination through algorithms and leave them unsanctioned.[9] The law needs to step in as the human rights impact of artificial intelligence extends to the body of rights set

---

[1] *See* Lorna McGregor, Daragh Murray & Vivian Ng, *International Human Rights Law as a Framework for Algorithmic Accountability*, 68 INT. COMP. LAW Q. 309, 310 (2019), https://www.cambridge.org/core/product/identifier/S0020589319000046/type/journal_article (last visited Apr 9, 2022).

[2] FILIPPO A. RASO ET AL., *Artificial Intelligence & Human Rights: Opportunities & Risks*, 20 (2018), https://papers.ssrn.com/abstract=3259344.

[3] Office of the United Nations High Commissioner for Human Rights, The Right to Privacy in the Digital Age, 16 9 (2021), https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session48/Documents/A_HRC_48_31_AdvanceEditedVersion.docx.

[4] Mathias Risse, *Human Rights and Artificial Intelligence: An Urgently Needed Agenda*, 41 HUM. RIGHTS Q. 1, 10 (2019), https://muse.jhu.edu/article/716358 (last visited Oct 26, 2021). ; Julia Angwin et al., *Machine Bias*, PROPUBLICA, May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[5] Risse, *supra* note 4 at 11.

[6] Tom van Nuenen et al., *Transparency for Whom? Assessing Discriminatory Artificial Intelligence*, 53 COMPUTER 36, 36 (2020).

[7] Brent Daniel Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, 3 BIG DATA SOC. 1 (2016), https://doi.org/10.1177/2053951716679679.

[8] ACCESS NOW & LİNDSEY ANDERSEN, *Human Rights in the Age of Artificial Intelligence*, 40 17 (2018), https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf; McGregor, Murray, and Ng, *supra* note 1 at 312; Christiaan van Veen & Corinne Cath, *Artificial Intelligence: What's Human Rights Got To Do With It?*, MEDİUM, May 2018, https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5 (last visited Jun 6, 2023).

[9] ACCESS NOW AND ANDERSEN, *supra* note 8 at 17.

out in human rights instruments.[10] It should be noted that examining discriminatory impacts as a human rights law matter does not imply a standing against technology or innovation.[11]

Authorities with legal power have begun working on artificial intelligence to catch up with the pace of the developments in the artificial intelligence field, with an approach to both considering ethical and legal aspects. The EU published an ethics guidelines document on artificial intelligence in 2019 and a regulation proposal in 2021 to introduce EU-wide legally binding rules on artificial intelligence.[12] The CoE also established the Ad hoc Committee on Artificial Intelligence (CAHAI) and later Committee on Artificial Intelligence (CAI), and made a call to prevent discrimination based on artificial intelligence, and in 2023, published a draft framework convention regarding the human rights impacts of artificial intelligence.[13] The United Nations High Commissioner for Human Rights has also published a report underlining the effects of artificial intelligence on human rights.[14] States are also developing artificial intelligence strategies to keep up with the developments in artificial intelligence.[15] The Republic of Türkiye also published an artificial intelligence strategy in August 2021.[16] Additionally, Türkiye's Judicial Reform Strategy of 2019 declares a prospect for using artificial intelligence in judicial activities.[17] It should be noted that national-level works have been left out of the scope of

[10] Risse, *supra* note 4 at 11.
[11] McGregor, Murray, and Ng, *supra* note 1 at 335.
[12] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 39 (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419; European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (last visited Aug 1, 2022).
[13] CHRISTOPHE LACROIX, *Report of the Committee on Equality and Non-Discrimination on Preventing Discrimination Caused by the Use of Artificial Intelligence*, 20 (2020), https://pace.coe.int/pdf/afacbf21374dd109246029ed6e5ccd4e9317509f2ef3131cf1db1c6d10b96bb8/doc.%2015151.pdf (last visited May 20, 2023); Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, (2023), https://rm.coe.int/cai-2023-01-revised-zero-draft-framework-convention-public/1680aa193f (last visited Feb 22, 2023).
[14] OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR HUMAN RIGHTS, *supra* note 3.
[15] *See* OECD.AI (2021), *Powered by EC/OECD (2021), Database of National AI Policies*, https://oecd.ai/en/dashboards/overview (last visited Jun 6, 2023); LAURA GALINDO, *An Overview of National AI Strategies and Policies*, 26 (2021), https://goingdigital.oecd.org/data/notes/No14_ToolkitNote_AIStrategies.pdf.
[16] Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, *Ulusal Yapay Zekâ Stratejisi 2021-2025*, (2021), https://cbddo.gov.tr/uyzs.
[17] T.C. Adalet Bakanlığı, *Yargı Reformu Stratejisi - 2019*, (2019), https://sgb.adalet.gov.tr/Resimler/SayfaDokuman/23122019162931YRS_TR.pdf.

this thesis as the problem is approached from the perspective of international human rights law, except for examples and explanations given when necessary.

The importance of looking at artificial intelligence from a human rights perspective beyond ethics is righteously emphasized in the literature.[18] Studies on artificial intelligence from a human rights law perspective are increasing at the time of writing of this thesis.[19] Even though progress is made in addressing the legal lacuna, it still continues at the time of this study. This thesis seeks to contribute to addressing the lacuna in the studies regarding artificial intelligence from human rights law perspective, focusing on the right to equality and non-discrimination. It builds on contributing to existing research on the matters in the artificial intelligence literature and the legal literature suggesting addressing artificial intelligence with human rights.

This study examines artificial intelligence within the right to equality and non-discrimination dimensions, handling the issue as a human rights law problem. It aims to

[18] ACCESS NOW AND ANDERSEN, *supra* note 8 at 3. ; GENEVIEVE SMITH & ISHITA RUSTAGI, *Mitigating Bias in Artificial Intelligence: An Equity Fluent Leadership Playbook*, 67 45 (2020), https://haas.berkeley.edu/wp-content/uploads/UCB_Playbook_R10_V2_spreads2.pdf.

[19] *See* McGregor, Murray, and Ng, *supra* note 1; Nathalie A. Smuha, *Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea*, 34 PHILOS. TECHNOL. 91 (2021), https://doi.org/10.1007/s13347-020-00403-w (last visited May 7, 2023); FREDERIK J. ZUIDERVEEN BORGESIUS, DISCRIMINATION, ARTIFICIAL INTELLIGENCE, AND ALGORITHMIC DECISION-MAKING (2018), https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73; BENJAMIN WAGNER, *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, 60 (2018), https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10 (last visited Aug 4, 2022); RASO ET AL., *supra* note 2; Some studies focus on the EU non-discrimination law perspective. E.g. Philipp Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law*, 55 COMMON MARK. LAW REV. (2018), https://kluwerlawonline.com/journalarticle/Common+Market+Law+Review/55.4/COLA2018095 (last visited Nov 9, 2021); Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why Fairness Cannot Be Automated: Bridging the Gap between EU Non-Discrimination Law and AI*, 41 COMPUT. LAW SECUR. REV. 105567 (2021), https://www.sciencedirect.com/science/article/pii/S0267364921000406; Raphaële Xenidis, *Tuning EU Equality Law to Algorithmic Discrimination: Three Pathways to Resilience*, 27 MAASTRICHT J. EUR. COMP. LAW 736 (2020), https://doi.org/10.1177/1023263X20982173 (last visited Apr 11, 2022); Raphaële Xenidis & Linda Senden, *EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination*, *in* GENERAL PRINCIPLES OF EU LAW AND THE EU DIGITAL ORDER 151 (Ulf Bernitz et al. eds., 2020), https://papers.ssrn.com/abstract=3529524 (last visited Jul 28, 2022); JANNEKE GERARDS & RAPHAËLE XENIDIS, *Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non Discrimination Law*, 192 (2021), https://data.europa.eu/doi/10.2838/544956 (last visited Aug 4, 2022); European Union Agency for Fundamental Rights, *#BigData: Discrimination in Data-Supported Decision Making*, (2018), https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making (last visited Aug 6, 2022); EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, BIAS IN ALGORITHMS: ARTIFICIAL INTELLIGENCE AND DISCRIMINATION. (2022), https://data.europa.eu/doi/10.2811/536044 (last visited May 21, 2023); EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, GETTING THE FUTURE RIGHT: ARTIFICIAL INTELLIGENCE AND FUNDAMENTAL RIGHTS (2020), https://data.europa.eu/doi/10.2811/58563 (last visited May 21, 2023).

develop the current legal understanding of *discrimination through artificial intelligence*[20] by determining the effects of artificial intelligence on the right to equality and non-discrimination, to examine how the existing framework of the right to equality and non-discrimination can be interpreted in discrimination through artificial intelligence, and to make suggestions on the promises of regulations on artificial intelligence for the right to equality and non-discrimination.

In order to accomplish its purpose, this study poses several questions with the intent of finding answers and several questions for calling attention. These questions include:

- How could concerns for human rights in artificial intelligence be connected with human rights law?
- How to integrate similar wordings with legal connotations in the artificial intelligence field to human rights law when examining discrimination through artificial intelligence?
- What could human rights law learn from the artificial intelligence ethics (AI-ethics) discussions?
- What could human rights law provide with its existing legal framework to discrimination through artificial intelligence?
- Do existing human rights law mechanisms suffice to address discrimination through artificial intelligence?
- How could the human rights law mechanisms be adapted and improved to address better for discrimination through artificial intelligence?
- What are some examples of discrimination through artificial intelligence?
- How could a categorization of discrimination through artificial intelligence be made by the discrimination grounds in the framework of the right to equality and non-discrimination?
- Are there distinct positive and procedural obligations on states regarding addressing the right to equality and non-discrimination in the context of artificial intelligence?

[20] The reason for choosing the wording *"discrimination through artificial intelligence"* is to express that discrimination occurs due to artificial intelligence without referring to artificial intelligence as "the discriminator" and to encompass the different forms of discrimination in the context of artificial intelligence.

- Are new and specific regulations required to address the right to equality and non-discrimination in artificial intelligence?
- What do the forthcoming regulatory frameworks on artificial intelligence offer for the right to equality and non-discrimination?
- Are the forthcoming regulations fit to address discrimination through artificial intelligence in the context of the right to equality and non-discrimination?

The scope of examination of human rights impacts of artificial intelligence in this study is limited to the right to equality and non-discrimination. Even though the scope of the thesis is limited to the right to equality and non-discrimination, it should be kept in mind that this right encompasses a broad scope, and its violation may also indicate a violation of another human right. Due to the nature of the subject matter, the legal perspectives this study benefits from are not narrowed down to a specific legal body. Such an approach is needed as this study not only attempts to examine the issue as it is but also aims to project a future building on how discrimination through artificial intelligence can be addressed now by making theoretical discussions.

Firstly, this study handles the issue as a legal problem from the human rights law perspective by connecting fairness, bias, and discrimination through artificial intelligence with the right to equality and non-discrimination. Secondly, this study categorizes different possibilities of discrimination through artificial intelligence under the discrimination grounds recognized in law and adapt the existing human rights law so as to explore its capabilities of providing for discrimination through artificial intelligence. Thirdly, this study probes the future regulation of the right to equality and non-discrimination with artificial intelligence specific regulations.

In order to give a clearer understanding of the problem, Chapter 1 identifies the scope of the study by scouring various relevant terms. Although considerable attention is given to the discussions about the human rights impacts of artificial intelligence, more is needed to address them as a matter of human rights law. Therefore, the first chapter aims to construct the issue legally. In doing so, the first chapter examines equality and discrimination in the legal sense after exploring the essential artificial intelligence related terms. Mentioning different terms and subjects in an overview is required to clarify and build the thesis.

When constructing the problem as a legal matter, different understandings of the right to equality and non-discrimination in various human rights documents are mentioned. The reason for mentioning different legal traditions thereunder is to connect the understanding of equality and discrimination to the legal landscape, to depict what equality, fairness, and discrimination may offer, that is, to bring a broader understanding of the notions in the legal sense concerning artificial intelligence.

After laying out the problem as a human rights law matter, Chapter 2 delves into exploring how artificial intelligence may impact the right to equality and non-discrimination. The second chapter mentions different areas of artificial intelligence use to picture how it is utilized today. The second chapter then moves on to exploring discrimination through artificial intelligence under the discrimination grounds widely recognized under international human rights law. Finally, how existing international human rights law handles the right to equality and non-discrimination is discussed and explored to probe what it may offer to address discrimination through artificial intelligence. The aim thereof is to adapt and visualize the existing human rights law's perspective with artificial intelligence. In doing so, the ECtHR has been chosen, as it is amongst the most prominent human rights bodies, and the lack of its perspective in studies is more apparent.

Chapter 3 aims to tackle the regulation of artificial intelligence with a view to the right to equality and non-discrimination. While non-binding and binding regulations specific to artificial intelligence are discussed, human rights impacts are mainly addressed in the non-binding documents. Various non-binding documents and the two most prominent forthcoming legally binding documents from the EU and the CoE are examined in this study, concentrating on their implications for the right to equality and non-discrimination. Similar national regulatory works are excluded from the scope of the study. Since the study aims to contribute to discussions on regulating artificial intelligence, investigating different regulatory works from different authorities is convenient. The non-binding documents help provide a theoretical basis for the third chapter, mainly through principles for artificial intelligence. The forthcoming legally binding documents are traced through their provisions relating to the right to equality and non-discrimination.

Research methods for this study demanded an interdisciplinary approach at times. As the subject matter emanates from the technological domain, exploring not only legal documents and legal research but also technical works in the artificial intelligence field, newspapers, and research in ethics were required to conduct the study. Examining research beyond the legal sphere seeks to support a jurist's perspective on what the law might say to them and what the law might learn from them. Resonating with that, obtaining foundational knowledge and demonstrating a clear understanding of the artificial intelligence basics was necessary to mold coherent debates within the study.

Given that the subject matter is relatively new and still expanding, there have been particular difficulties in research regarding possible methodological limitations. The subject matter requires constantly keeping up with the latest developments in the artificial intelligence domain where at present, each day, a piece of news is issued, and also, in the legal domain, where the works have gained weight since the start of this study.

Chapter 1:
A LEGAL PROBLEM: ARTIFICIAL INTELLIGENCE AND THE RIGHT TO EQUALITY AND NON-DISCRIMINATION

Once thought to be a typical science fiction movie subject in the past, enough time has gone by since artificial intelligence became a subject that the world follows the developments about it in the daily news. Today, it is easier to speak of areas that do not use artificial intelligence than to list the areas that adopt it. Artificial intelligence has conquered the world and impacts many aspects of human life.[21] Nevertheless, instead of artificial intelligence taking over the world, the present-day adoption of artificial intelligence may lead to the emergence of tyrannical practices from humans to humans in different ways.[22]

Artificial intelligence is not merely a technology produced and used by engineers, software developers, or data scientists; on the contrary, it is adopted in both everyday life and business life within a far-reaching range, extending to areas such as security, marketing, sales, customer service, human resources, finance, transportation, healthcare and medicine, telecommunication, energy, retail.[23] Psychology, ethics, mathematics, economics, statistics, politics, and law are a few science fields that artificial intelligence uses as a resource and vice versa.[24] The use of artificial intelligence in many of these areas amounts to an intricate matter of subject.

---

[21] Frank Pasquale & Danielle Keats Citron, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. LAW REV. 1, 2 (2014); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 3 (1st edition ed. 2016); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, *in* CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY 1, 1 (2018); PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 286 (2015).

[22] CEM SAY, 50 SORUDA YAPAY ZEKA 155 (7th ed. 2018).

[23] IBM, *IBM Global AI Adoption Index 2022*, 10 (2022), https://www.ibm.com/downloads/cas/GVAGA3JP; MICHAEL CHUI & SANKALP MALHOTRA, *Notes from the AI Frontier: AI Adoption Advances, but Foundational Barriers Remain*, 11 3 (2018), https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain (last visited Jul 26, 2022).

[24] Michael Chui et al., *Notes from the AI Frontier: Insights from Hundreds of Use Cases*, MCKINSEY GLOB. INST. 32, 1 (2018), https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20ai%20frontier%20applications%20and%20value%20of%20deep%20learning/notes-from-the-ai-frontier-insights-from-hundreds-of-use-cases-discussion-paper.ashx; PETER NORVIG & STUART RUSSELL, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH, GLOBAL EDITION 20 (4th edition ed. 2021).

A reader of this study is using and/or being affected by the use of artificial intelligence, possibly even without knowing. Some everyday use of artificial intelligence includes logging into one's email account through dictating the website link on the search bar using speech recognition and seeing a new email dropped into the 'spam' folder because of the spam filtering system.[25] The user may open that email and realize that it is not spam, and may click a button that is marked 'it is not spam'. By doing this, the user is guiding the email software that the emails sent from that address are not spam so that future emails from the same address would not fall into the spam folder.[26]

Beyond speech recognition, spam filtering, content moderation, game playing, and robotic devices, examples of artificial intelligence usages include facial recognition and surveillance, medical diagnosis for diseases, judicial processes including policing and deducing criminal sentences, criminal justice, employment for hiring or managing employees, social security, finance for deciding on who is eligible for a credit loan or calculating interest rates, insurance fees, and even more.[27] Anything relating to humans may bear a legal facet. Humans develop and use artificial intelligence, meaning that artificial intelligence will have impacts that could raise legal issues, especially human rights issues that needs to be tackled by enforcement mechanisms under the human rights law.[28] Algorithms and AI-systems are commonly embraced for supporting or informing decision-making, which may have human rights implications.[29] One such concern of human rights issues in artificial intelligence is the right to equality and non-discrimination.

---

[25] Frederik J. Zuiderveen Borgesius, *Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence*, 24 INT. J. HUM. RIGHTS 1572, 1572 (2020), https://doi.org/10.1080/13642987.2020.1743976.

[26] Google Developers, *Framing: Key ML Terminology | Machine Learning*, MACHINE LEARNING FOUNDATIONAL COURSES CRASH COURSE (2022), https://developers.google.com/machine-learning/crash-course/framing/ml-terminology (last visited Aug 10, 2022). *See* the machine learning title under the heading 1.1.5 Ambit of Artificial Intelligence, page 11 for more explanation.

[27] NORVIG AND RUSSELL, *supra* note 24 at 46–48; Zuiderveen Borgesius, *supra* note 25 at 1572; EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, *Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement*, 36 2 (2020), https://data.europa.eu/doi/10.2811/524628 (last visited Aug 6, 2022); Angwin et al., *supra* note 4; DOMINGOS, *supra* note 21 at xiv; FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 4–5 (2015), https://www.degruyter.com/document/doi/10.4159/harvard.9780674736061/html (last visited May 10, 2022); Martin Ebers, *Liability for Artificial Intelligence and EU Consumer Law*, 12 J. INTELLECT. PROP. INF. TECHNOL. ELECTRON. COMMER. LAW 204, 206 (2021), https://heinonline.org/HOL/Page?handle=hein.journals/jipitec12&id=211&div=&collection=; ZUIDERVEEN BORGESIUS, *supra* note 19 at 14–18; McGregor, Murray, and Ng, *supra* note 1 at 317.

[28] Smuha, *supra* note 19 at 98.

[29] McGregor, Murray, and Ng, *supra* note 1 at 317.

The imminent effects of artificial intelligence on the right to equality is a subject that cannot be overlooked. This study aims to examine such issues, identify the problems that may arise from using artificial intelligence methods, how the use of artificial intelligence may concern the right to equality and non-discrimination, and explore the present arguments in academia and the efforts to regulate artificial intelligence.

Efforts to govern artificial intelligence have resulted in inquiring about human rights law.[30] This study deals with the problem of equality and discrimination in artificial intelligence from the perspective of human rights law. Problems of equality and discrimination due to artificial intelligence are legal problems besides presenting social and ethical problems. There are concerns for human rights protection to be abstract and broad when examining artificial intelligence impacts; however, when approached within the human rights law documents that define norms and mechanisms that provide established interpretation, such concerns would be mainly addressed.[31]

For this reason, and since the study is a thesis in the field of law, this chapter aims to deliver the right to equality and non-discrimination as a legal problem in the context of artificial intelligence. To define this problem with its different elements, it is necessary to deal with the basic concepts that make up the problem's elements and reveal the connection between them. In this direction, the discussions on the legal definition of artificial intelligence will be discussed based on the technical descriptions. As will be mentioned in the following sections, although the concepts such as equality, fairness, and discrimination and the importance of these concepts and the solutions to the problems related to them are frequently mentioned in the field of artificial intelligence, these concepts are not used in a way that corresponds to their legal counterparts, and although the concepts used are the same, their concrete equivalents and solutions may be different in legal and non-legal areas.[32] However, when a problem about the concepts transpires in the field of law, these concepts will be approached from a legal perspective. Therefore, it is necessary to reconstruct these concepts based on what they say and express in law.[33] Thus, artificial intelligence is a subject that needs an interdisciplinary perspective to be

[30] Smuha, *supra* note 19 at 94.
[31] *Id.* at 99.
[32] *See* Alice Xiang & Inioluwa Deborah Raji, *On the Legal Compatibility of Fairness Definitions*, 1, 1 (2019), http://arxiv.org/abs/1912.00761; Smuha, *supra* note 19 at 100; McGregor, Murray, and Ng, *supra* note 1 at 324; GERARDS AND XENIDIS, *supra* note 19 at 47.
[33] Xiang and Raji, *supra* note 32 at 1; *See* GERARDS AND XENİDİS, *supra* note 19.

transferred in the field of law. Through such an approach, lawyers will be able to understand the concepts used in the field of artificial intelligence in their context, and at the same time, artificial intelligence researchers and artificial intelligence actors in different disciplines will ultimately be able to interpret legal concepts keeping human rights law in their mind.[34] Addressing discrimination through artificial intelligence within human rights may help manage the prospect of ethical relativism.[35] Discrimination through artificial intelligence should be addressed together by legal and artificial intelligence communities so that they can guide each other in resolving the matter and identifying where the existing human rights law is ambiguous and insufficient.[36] Introducing some related concepts, especially the types of artificial intelligence and the notion of bias, and transferring them to the law field is equally essential in legally addressing the issue. In this section, after the framework of artificial intelligence and related concepts is presented, concepts such as bias, fairness, equality, and discrimination will be examined from a legal point of view. Then the connection of these concepts with artificial intelligence will be made. For this reason, the concepts explained here do not aim to repeat what is known but to demonstrate the subject as a human rights law problem.

1.1 Artificial Intelligence

The term artificial intelligence was born in 1955, when the father of the word, John McCarthy, expressed it for the first time with his colleagues in a project proposal that sought *"...to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves"*.[37] Artificial intelligence was originally about *"...making a machine behave in ways that would be called intelligent if a human were so behaving"*.[38] Even though human intelligence is the basis for its concept, artificial intelligence does not necessarily mimic how human intelligence works.[39]

---

[34] See Xiang and Raji, *supra* note 32 at 1; Smuha, *supra* note 19 at 100–101; McGregor, Murray, and Ng, *supra* note 1 at 323; GERARDS AND XENIDIS, *supra* note 19 at 47.
[35] Smuha, *supra* note 19 at 93.
[36] Wachter, Mittelstadt, and Russell, *supra* note 19 at 9, 68; Smuha, *supra* note 19 at 100.
[37] John McCarthy et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 2 (1955), http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf; NORVIG AND RUSSELL, *supra* note 24 at 36.
[38] McCarthy et al., *supra* note 37 at 11.
[39]John McCarthy, *What Is Artificial Intelligence?*, (2007), http://www-formal.stanford.edu/jmc/whatisai.pdf.

Directed toward making *"effective and safe"* machines that thrive under circumstances that are new and unfamiliar, artificial intelligence uses different techniques to this end.[40] However, a few questions follow up on this purpose. What is required for artificial intelligence to be effective and safe? What do effectiveness and safety mean in the context of artificial intelligence? Who will consider artificial intelligence effective and safe, and when will it be an effective and safe machine? These questions are essential because, for example, when an artificial intelligence based application profiles its users, there is no doubt that this will affect several human rights. Again, safety could be understood as technical safety, for example, when the artificial intelligence in question is a robotic vacuum cleaner. Instead, safety will require compliance with human rights regarding an artificial intelligence tool such as a facial recognition application. If the 'safeness' authority is the developer company in such a case, it may still consider the artificial intelligence based application safe although it is not in terms of human rights.[41] Such implications of artificial intelligence are relevant to discrimination through artificial intelligence when examining a discrimination claim. For example, the supposed effectiveness of an AI-system may be relevant to the discrimination test explored in the second chapter. Likewise, safety is regarded as a principle for artificial intelligence, as explained in the third chapter.

Before the birth of artificial intelligence as a term, there were already works in that field. *"Can machines think?"* was the question Alan Turing had asked and denounced for being 'meaningless' in his article, where he laid out the infamous Turing test.[42] The Turing test is essentially a test formulated to measure a machine's capability to convince a human that it is a human, not a computer.[43] The Turing Test works as such: there is a computer, a human, and an interrogator.[44] The computer tries to convince the interrogator that it is a human by claiming human attributes and tries to convince the human that it is not a computer.[45] The other human tries to help the interrogator by leading the interrogator

[40] NORVIG AND RUSSELL, *supra* note 24 at 19.
[41] *See* Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI*, 41 COMPUT. LAW SECUR. REV. 105567, 3 (2021).
[42] Alan M. Turing, *Computing Machinery and Intelligence*, LIX MIND 433, 433 (1950), https://doi.org/10.1093/mind/LIX.236.433 (last visited Apr 9, 2022).
[43] *Id.*
[44] *Id.* at 433–434; NORVIG AND RUSSELL, *supra* note 24 at 1035.
[45] Turing, *supra* note 42 at 433–434; NORVIG AND RUSSELL, *supra* note 24 at 1035.

with clues.[46] If the interrogator fails to distinguish between the computer and the human, labeling the computer as human, it means that the computer has passed the Turing Test.[47] Since it was put in, many trials of the Turing Test have been made.[48] Nevertheless, the Turing test still had not been successfully passed by any computer.[49] Today, the artificial intelligence field is much broader and evolved into encompassing different meanings.[50]

1.1.1 Algorithms

The story of artificial intelligence had begun long before even Turing's works and roots back to algorithms. The word algorithm is named by the founder of algebra, al-Khwarizmi, making him a pioneer in the field of algorithms.[51] An algorithm is a chain of steps for concluding or unraveling a problem.[52] Algorithms are the basis of computer programming, but algorithms do not only exist in the world of ones and zeroes.[53] This suggestion has two meanings. In the first sense, it means that computer programs have not stayed in computers but transformed the social life and caught the eye of social scientists to examine this transformation and its impacts.[54] In the second sense, it means that individuals use algorithms in everyday life.

The term algorithm has different definitions, but all mention its components: the entering data as the 'input', a set of rules, and the final product as the 'output'.[55] Indeed, technical definitions explain the algorithms within the boundaries of mathematics.[56] When used in computer programs, algorithms automatically follow the chain of steps and attain a solution, removing the need for human oversight and algorithms becoming the

[46] Turing, *supra* note 42 at 434.
[47] NORVIG AND RUSSELL, *supra* note 24 at 1035.
[48] *Id.*
[49] *Id.*
[50] See regarding the journey of artificial intelligence *Id.* at 1–36.
[51] *Id.* at 27.
[52] ETHEM ALPAYDIN, MACHINE LEARNING: THE NEW AI x (Revised and updated edition ed. 2021).
[53] ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING 1 (4 ed. 2020).
[54] *Id.*; CARSTEN ORWAT, RISKS OF DISCRIMINATION THROUGH THE USE OF ALGORITHMS 11 (2020), https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/EN/publikationen/Studie_en_Diskriminierungsrisiken_durch_Verwendung_von_Algorithmen.pdf;jsessionid=E305EAE5E78553015D76E4A5951368B1.intranet241?__blob=publicationFile&v=2.
[55] ALPAYDIN, *supra* note 53 at 1; Dunja Mijatović, Council of Europe Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*, 24 (2019), https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 (last visited Jul 27, 2022).
[56] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 24.

primary or partial decision maker.[57] In the second case, humans benefit from algorithms to make the decision themselves.[58] Algorithms are not suitable for all problems, meaning that merely having the inputs is not enough to reach the output because how to convert the first one to the second is not always known.[59] Data collections provide information to learn from and find the appropriate algorithm for the problem.[60] "Learning algorithms" help to sort out massive data sets.[61] That is related to artificial intelligence, more specifically, machine learning and big data, which are discussed below.

1.1.2 Defining Artificial Intelligence

One cannot make a uniform definition of artificial intelligence.[62] A reason for that is the difference in approaches to deciding upon the purpose of artificial intelligence.[63] A second reason is a difference in interpretation of the matter; some regard intelligence as what is on the outside -the conduct, whereas some regard it as the intellect on the inside.[64] For example, the meaning of 'intelligence' in the Turing test was "acting humanly".[65] Another meaning for 'intelligence' in the context of artificial intelligence is the capability to "do the right thing" under unknown conditions.[66] A third reason is confusion from overusing the term and inconsistency on what to enclose when defining artificial intelligence.[67] Artificial intelligence is a vast domain of study and comes in different forms.[68] Therefore, artificial intelligence is a common term for various technologies and mechanisms.[69] Criticisms toward the overuse of the term are based on the view that since the term artificial intelligence makes reference to human intelligence and thus the human brain and its capabilities, it is not correct to name machine learning or automated processes as artificial intelligence because these do not function as the human brain or

[57] SIMON CHESTERMAN, WE, THE ROBOTS?: REGULATING ARTIFICIAL INTELLIGENCE AND THE LIMITS OF THE LAW 53 (2021); ZUIDERVEEN BORGESIUS, *supra* note 19 at 11.
[58] ZUIDERVEEN BORGESIUS, *supra* note 19 at 11.
[59] ALPAYDIN, *supra* note 52 at 17.
[60] *Id.*
[61] NORVIG AND RUSSELL, *supra* note 24 at 44.
[62] *Id.* at 19; Ebers, *supra* note 27 at 205.
[63] NORVIG AND RUSSELL, *supra* note 24 at 19.
[64] *Id.*
[65] *Id.* at 20.
[66] *Id.* at 22.
[67] Ebers, *supra* note 27 at 205.
[68] NORVIG AND RUSSELL, *supra* note 24 at 19.
[69] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 5; Ebers, *supra* note 27 at 205.

make 'intelligent' actions.[70] According to this view, using the term artificial intelligence could be confusing or misguiding.[71] Another view suggests using the term automated decision making systems is the correct wording.[72] Some mention algorithmic decision-making systems.[73] Some confine the issue with the term machine learning.[74] Nevertheless, some still prefer using the term artificial intelligence.[75]

There are some other related definitions to be made. The steps of research, design, data collection, experimenting, application, and administration constitute the artificial intelligence's lifecycle.[76] Artificial intelligence helps automate things. Automation refers to technology building and implementing for observing and controlling the production and delivery of goods and services.[77] It should be noted that while artificial intelligence systems can also be automated, automated systems represent a broader scope than the first one.[78] Artificial intelligence systems may involve prediction-based decision-making, requiring methods and techniques similar to human intelligence.[79] In contrast, automated systems may refer to automatically conducting predictable routine tasks, including beyond artificial intelligence.[80] Automation is widely adopted in robotics; for example, automated tools are used in factories to connect certain parts of a machine.[81] The advantage of automation is aiding the workforce to leave time-consuming and sometimes unsafe labor to automated systems.[82] It should be noted that there are also some worries that the workforce is going to be affected worse than being affected in a good way because

---

[70] Kathy Pretz, *Stop Calling Everything AI, Machine-Learning Pioneer Says*, IEEE SPECTRUM, Mar. 2021, https://spectrum.ieee.org/stop-calling-everything-ai-machinelearning-pioneer-says; ALGORITHMWATCH, *Automating Society: Taking Stock of Automated Decision-Making in the EU*, 148 9 (2019), https://algorithmwatch.org/de/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf (last visited Jul 25, 2022).
[71] ALGORITHMWATCH, *supra* note 70 at 9.
[72] *Id.*
[73] ORWAT, *supra* note 54.
[74] World Economic Forum Global Future Council on Human Rights 2016-18, *How to Prevent Discriminatory Outcomes in Machine Learning*, (2018), https://www.weforum.org/whitepapers/how-to-prevent-discriminatory-outcomes-in-machine-learning/ (last visited Feb 10, 2022).
[75] ZUIDERVEEN BORGESIUS, *supra* note 19 at 14.
[76] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 36.
[77] International Society of Automation, *What Is Automation? - ISA*, ISA.ORG (2022), https://www.isa.org/about-isa/what-is-automation (last visited Jul 25, 2022).
[78] Michael Gaynor, *Automation and AI Sound Similar, but May Have Vastly Different Impacts on the Future of Work*, BROOKINGS, https://www.brookings.edu/articles/automation-and-artificial-intelligence-sound-similar-but-may-have-vastly-different-impacts-on-the-future-of-work/.
[79] *Id.*
[80] *Id.*
[81] *Id.*
[82] NORVIG AND RUSSELL, *supra* note 24 at 1037.

of artificial intelligence.[83] Automation also comes with the risk of amplifying the inequalities in society as it could particularly worsen the income inequalities.[84] Automated decision making, or decision support systems use automatic decision making models to decide on things by substituting a person or an entity.[85] Algorithmic decision making systems, either partially or completely automated, focus on analyzing massive quantities of data to make connections or provide relevant data to make decisions.[86] In the second case, where there is fully automated systems, human involvement may be eliminated entirely.[87] Algorithmic systems are not necessarily qualified as 'artificial intelligence' as they may or may not be based on artificial intelligence technologies.[88] These systems may only use regular computations and not the techniques of artificial intelligence and work as regular computer programs but could also utilize methods of artificial intelligence.[89]

How the CoE and the EU define artificial intelligence is also of particular importance in establishing the context of the main subject of the study, as these stand out as two organizations dealing with artificial intelligence from the legal perspective. [90]

The CoE's approach to definition of artificial intelligence is seen in the Recommendation of the CoE Commissioner for Human Rights on artificial intelligence, which states:

---

[83] Spyros Makridakis, *The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms*, 90 FUTURES 46, 57 (2017), https://www.sciencedirect.com/science/article/pii/S0016328717300046 (last visited Aug 21, 2022).
[84] NORVIG AND RUSSELL, *supra* note 24 at 1038.
[85] ALGORITHMWATCH, *supra* note 70 at 9.
[86] EUROPEAN PARLIAMENT DIRECTORATE GENERAL FOR PARLIAMENTARY RESEARCH SERVICES, CLAUDE CASTELLUCCIA & DANIEL LE MÉTAYER, UNDERSTANDING ALGORITHMIC DECISION-MAKING: OPPORTUNITIES AND CHALLENGES 1 (2019), https://data.europa.eu/doi/10.2861/536131 (last visited Jul 25, 2022).
[87] *Id.*
[88] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 24; ALGORITHMWATCH, *supra* note 70 at 9.
[89] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 24.
[90] The studies conducted by the CAHAI assigned by the Council of Europe also refer to the definitions made by the EU HLEG AI (High-level Expert Group on Artificial Intelligence), the working group under the EU assigned for artificial intelligence studies. This suggests that both organizations follow each other's work, which is notable as the need for common or at least concordant definitions in legal approaches to artificial intelligence is apparent. See Catelijne Muller & The CAHAI Secretariat, *The Impact of AI on Human Rights, Democracy and the Rule of Law*, *in* TOWARDS REGULATION OF AI SYSTEMS: GLOBAL PERSPECTIVES ON THE DEVELOPMENT OF A LEGAL FRAMEWORK ON ARTIFICIAL INTELLIGENCE SYSTEMS BASED ON THE COUNCIL OF EUROPE'S STANDARDS ON HUMAN RIGHTS, DEMOCRACY AND THE RULE OF LAW 21, 21–23 (2020), https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a.

> "AI is used as an umbrella term to refer generally to a set of sciences, theories and techniques dedicated to improving the ability of machines to do things requiring intelligence."[91]

Moreover, the Recommendation of the CoE Commissioner for Human Rights on artificial intelligence includes a definition for artificial intelligence systems (AI-systems):

> "An AI system is a machine-based system that makes recommendations, predictions or decisions for a given set of objectives. It does so by: (i) utilising machine and/or human-based inputs to perceive real and/or virtual environments; (ii) abstracting such perceptions into models manually or automatically; and (iii) deriving outcomes from these models, whether by human or automated means, in the form of recommendations, predictions or decisions."[92]

Automated decision making is defined by CoE Human Rights Commissioner as:

> "A process of making a decision by automated means. It usually involves the use of automated reasoning to aid or replace a decision-making process that would otherwise be performed by humans. It does not necessarily involve the use of AI but will generally involve the collection and processing of data".[93]

A precise definition for artificial intelligence systems is made under Article 2(a) of the "Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law" (CoE Draft Convention on AI) publicized in 2023.[94] According to the definition hereof, the AI-systems may use computational methods derived from statistics or mathematics. The AI-system may either carry out functions associated with or require human intelligence or assist or replace human decision-making. The definition provides non-exhaustive examples of functions, including recognition, planning, prediction, and classification. It appears that CoE does

[91] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 5.
[92] *Id.*
[93] *Id.* at 24.
[94] Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, *supra* note 13 at Article 2 Article 2 of the CoE Draft Convention on AI defines AI systems as *"(...) any algorithmic system or a combination of such systems that, as defined herein and in the domestic law of each Party, uses computational methods derived from statistics or other mathematical techniques to carry out functions that are commonly associated with, or would otherwise require, human intelligence and that either assists or replaces the judgment of human decision-makers in carrying out those functions. Such functions include, but are not limited to, prediction, planning, classification, pattern recognition, organisation, perception, speech/sound/image recognition, text/sound/image generation, language translation, communication, learning, representation, and problem-solving".*

not look at whether the methods used fall into machine learning, deep learning, or any other artificial intelligence subfield.[95]

Article 2(b) of the CoE Draft Convention on AI defines the AI-system's lifecycle as *"all phases of existence"* from its design to its *"decommissioning"*. The subsequent paragraphs of the article introduce three different parties associated with artificial intelligence: the provider, the user, and the subject.

The EU defined artificial intelligence in various documents. The first Communication of the EU on artificial intelligence dated 2018 defines artificial intelligence as:

> "…systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."[96]

The EU had brought together an independent expert group on artificial intelligence (AI HLEG) who later suggested the definition above be replaced with their definition. The definition made by AI HLEG starts by putting a definition for AI-systems by describing how these systems work towards a goal.[97] According to AI HLEG's definition, AI-systems can be software or hardware that humans design and may be given a goal that the systems need to work toward by perceiving their environment and interpreting and processing the data, which can be structured or unstructured.[98] The second part of AI HLEG's definition explains artificial intelligence as a discipline and describes its different techniques consisting of machine learning, machine reasoning, and robotics.[99]

Subsequently, the EU published its proposal for a regulation on artificial intelligence (EU AI Act Proposal) in 2021.[100] The Explanatory Memorandum for the EU AI Act Proposal mentions that the EU's definition of artificial intelligence will be technology-

---

[95] See heading 1.1.5, titled "Ambit of Artificial Intelligence," for definitions of artificial intelligence subfields.
[96] This work is written in American English, but if British English is used in the quotations, it is left in its original form when quoting. ARTIFICIAL INTELLIGENCE FOR EUROPE, (2018), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN (last visited Aug 9, 2022).
[97] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *A Definition of AI: Main Capabilities and Scientific Disciplines*, 7 (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651 (last visited Aug 6, 2022).
[98] *Id.*
[99] *Id.* at 6.
[100] European Commission, *supra* note 12.

neutral, which gives room to the field to evolve and is supposed to be "*future-proof"* so that the regulation would have longevity.[101]

In the Explanatory Memorandum, the EU has described artificial intelligence as:

> "Artificial Intelligence (AI) is a fast evolving family of technologies that can bring a wide array of economic and societal benefits across the entire spectrum of industries and social activities. By improving prediction, optimising operations and resource allocation, and personalising service delivery, the use of artificial intelligence can support socially and environmentally beneficial outcomes and provide key competitive advantages to companies and the European economy.*"*[102]

This resembles more of an explanation of the matter rather than a definition. Even the heading of the recital is titled "Reasons for and Objectives of the Proposal" and not "definition". Nonetheless, article 3 of the EU AI Act Proposal is reserved for definitions, where a definition for AI-systems is made. Article 3(1) of the EU AI Act Proposal defines an AI-system as:

> "...software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.*"*[103]

Annex I of the EU AI Act Proposal is titled "Artificial Intelligence Techniques and Approaches" and explains the techniques and approaches for artificial intelligence.[104] Examples of software AI-systems include voice assistants, search engines, face recognition systems whereas examples of hardware systems include autonomous cars and drones.[105]

---

[101] *Id.* at Explanatory Memorandum Recital 5.2.1.
[102] *Id.* at Explanatory Memorandum Recital 1.1.
[103] *Id.* at article 3.
[104] European Commission, *Annexes to the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, Annex I (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (last visited Aug 1, 2022) Annex I states: *"(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.*"
[105] EUROPEAN COMMISSION, *supra* note 96.

The EU AI Act Proposal's definition of artificial intelligence has been criticized for being circular.[106] The criticisms indicate that the definition could be more precise on whether the computational techniques do not necessarily fall under the definition of artificial intelligence and whether the EU AI Act Proposal applies to these, even in cases of simple automation.[107] In any case, impacts on human rights do not necessarily change according to the techniques used.[108] Though some suggestions or criticism could be made on these definitions to improve them before they are adopted, finding a polished definition of artificial intelligence is beyond the scope of this study.

The definitions of artificial intelligence in the EU AI Act Proposal and the Recommendation of the CoE Draft Convention on AI are similar in their approaches. The EU AI Act Proposal defines AI-systems through techniques and approaches and mentions the capability to *generate outputs* under the purposes defined by humans. Likewise, the CoE Draft Convention on AI has preferred to approach it by referring to the capability to "*carry out functions"* using computational methods or other mathematical techniques. It should be noted that the techniques and approaches in the EU AI Act Proposal are limited to those listed in Annex I. However, the techniques and methods thereunder are flexible and open to amendments according to Article 4 of the EU AI Act Proposal. There is no such limit in the CoE Draft Convention on AI.

Interpretation of artificial intelligence by the EU and the CoE seem to include other techniques than machine learning. While recognizing the intricacies of the scope of artificial intelligence, this study chooses to proceed with artificial intelligence as a general term but refers to other terms when required.[109] Even though it is true that artificial intelligence brings human thinking into minds, it is also true that artificial intelligence holds different understandings, and that understanding is one of them.[110] Lack of consensus on the choice of wording may result in lacunae or back loops in law enforcement. It is true that the term artificial intelligence is nearly used as part of pop

---

[106] Nathalie A. Smuha et al., *How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act*, 1, 14 (2021), https://papers.ssrn.com/abstract=3899991 (last visited Aug 6, 2022).
[107] *Id.*
[108] *Id.*
[109] *See* FREDERİK ZUİDERVEEN BORGESİUS, DİSCRİMİNATİON, ARTİFİCİAL INTELLİGENCE, AND ALGORİTHMİC DECİSİON-MAKİNG 14 (2018), https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.
[110] Pretz, *supra* note 70; Contra ALGORITHMWATCH, *supra* note 70 at 9.

culture, referring to various specialties.[111] Nevertheless, this study approaches artificial intelligence not from the technically strict point of view but rather from the everyday use of the word as it has become the mainstream phrase incorporating different existing applications of artificial intelligence.[112] The main reason for that is to explain artificial intelligence more efficiently and because focusing on strict definitions is not the most proper solution for a legal approach that requires an exhaustive viewpoint. When dealing with discrimination through artificial intelligence from the legal perspective, the essential problem is the function and effect of the systems, rather than the technical ways of how artificial intelligence works.[113] Individuals use or encounter these technologies daily without knowing whether it is artificial intelligence in the exact technical meaning. Whether artificial intelligence makes direct decisions or is used as part of a system, it is called artificial intelligence, referring to its non-human feature. Indeed, artificial intelligence as machine learning is different from general artificial intelligence and calling each automated system as artificial intelligence may not be correct. Nevertheless, making rigid preferences about which sub-branch or exactly which form of artificial intelligence is not needed when dealing with questions about equality and discrimination. Suppose artificial intelligence technology evolves to another point; an approach that includes different forms of that technology to address comparable and parallel problems correctly is more adequate. Likewise, as stated above, it is seen that the legal authorities that will decide on a normative definition are directed towards keeping the definition comprehensive. Both of the mentioned definitions from legal authorities aim to describe artificial intelligence in a way that has a broad scope of application and is open-ended and suitable for the long term. Analyzing the definitions made by such legal authorities is essential because, when implemented, they will become benchmarks expected from the practitioners in the field.

1.1.3 Big Data's Relation to Artificial Intelligence

From all the different definitions of artificial intelligence, it is clear how data constitutes a critical component. As will be explained, data is closely associated with concerns of equality and non-discrimination that might evolve from artificial intelligence.

[111] Ebers, *supra* note 27 at 205.
[112] *Id.*
[113] CHRISTOPHE LACROIX, *supra* note 13 at 7.

Therefore, further explanations of data should be made. One of such data-related notions is big data. Big data refers to *"very large data-sets"* that have grown in the age of the internet.[114] Big data includes enormous amounts of texts from various languages, graphics in the form of photos and videos, audio, data from online outlets such as social media, user activity, tracking data, cookies from websites, and others.[115]

Big data has allowed the algorithms to utilize the huge amount of data.[116] Pasquale refers to a *"black-box effect"* resulting from big data and observes that removing this effect is burdensome.[117] Pasquale made an analogy to the black box in terms of recording information and not knowing the limits and consequences of the use of information, and this analogy is frequently used in the fields of artificial intelligence and big data.[118]

With black-box, transparency -and opacity in this case- discussions come along.[119] Transparency is critical when regulating artificial intelligence systems, which the third chapter of the study will discuss. Understanding big data and its relation to artificial intelligence is essential with regard to recognizing the impact of data on artificial intelligence resulting in discrimination.

1.1.4 Personal Data's Relation to Artificial Intelligence

Personal data and data protection, which has been a subject frequently studied and popularized in the last years, involves any and all information on an identified or identifiable individual and may emerge as another relevant legal notion for artificial intelligence.[120] Personal data is considered sensitive data when the information concerns

[114] NORVIG AND RUSSELL, *supra* note 24 at 44.
[115] *Id.*
[116] *Id.*
[117] PASQUALE, *supra* note 27 at 3.
[118] *Id.*
[119] *Id.* at 2.
[120] Although definitions vary, the most accepted definition for personal data today is in the EU's General Data Protection Regulation (GDPR). Another similar definition is found in the CoE's Convention no. 108., which was the first regulation in the personal data field. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), OJ L 119 4.5.2016 article 4.1 (2016), https://eur-lex.europa.eu/eli/reg/2016/679/oj (last visited Aug 10, 2022); Council of Europe, *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, ETS 108 article 2.1 (1981).

the race, ethnicity, religion, beliefs, sexual orientation, the health of an identifiable person.[121]

Although it could concern personal data when data feeding the artificial intelligence is attributable to an identified or identifiable person, it may not always be the case.[122] Therefore, there is no need for data to belong to an identified or identifiable person for artificial intelligence to result in outcomes concerning the right to equality and non-discrimination.[123]

Some authors consider data protection law as a solution for discrimination through artificial intelligence, using it to support the non-discrimination law framework in human rights.[124] Connecting the EU General Data Protection Regulation (GDPR) to non-discrimination law and incorporating it with algorithmic fairness tools are suggested accordingly to address the discriminatory impact of artificial intelligence.[125] At the very least, some scholars think that personal data protection law may help get ahead of artificial intelligence's discriminatory effects.[126] Significantly, the GDPR's Article 22, titled "Automated Individual Decision-Making, Including Profiling,"[127] is about automated decision-making systems and prohibits the usage of fully automated decision making systems that makes decisions based on personal data.[128]

---

[121] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 20.
[122] Zuiderveen Borgesius, *supra* note 25 at 1581.
[123] *Id.*
[124] Hacker, *supra* note 19 at 1145; Zuiderveen Borgesius, *supra* note 25 at 1582; ZUIDERVEEN BORGESIUS, *supra* note 19 at 31; GERARDS AND XENIDIS, *supra* note 19 at 48–50.
[125] Hacker, *supra* note 19 at 1146.
[126] Zuiderveen Borgesius, *supra* note 25 at 1582.
[127] Article 22 of the GDPR reads as: *"1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. 2. Paragraph 1 shall not apply if the decision: (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent. 3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision. 4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place."*
[128] ZUIDERVEEN BORGESIUS, *supra* note 19 at 40; Minesh Tanna & William Dunning, *Bias and Discrimination*, *in* ARTIFICIAL INTELLIGENCE: LAW AND REGULATION , 431 (Charles Kerrigan ed., 2022), https://www.elgaronline.com/view/edcoll/9781800371712/9781800371712.00035.xml.

As said, the GDPR is considered a tool for addressing discrimination through artificial intelligence.[129] The crossover of discrimination through artificial intelligence and data protection law is when Article 22 of the GDPR is violated; in such a case, enforceable mechanisms offered by data protection law are indeed ready for use.[130]

Article 22(3) calls for safeguarding measures such as the right to contest the decision if automated decision making is based on a circumstance falling under subparagraph (a), being necessary for entering a contract, and (c), explicit consent of data subject.[131] Under the circumstances of Article 22(3) GDPR, using personal data requires implementing protection safeguards relevant to the rights of data subjects to make the data processing more "fair."[132] Authors even argue that bias diagnosing and minimizing techniques should be essential; if they are lacking, there should be a violation of Article 22(3) GDPR.[133] Yet, the challenges include that Article 22 only applies to fully automated decision making systems, whereas the forms of automated decision making systems extend beyond that; therefore, more is needed to resort to Article 22 of the GPDR for handling decisions made by using automated decision making systems.[134]

Article 22(4) of the GDPR asserts that if sensitive data is in question, the exceptions recognized under Article 22(2) do not apply unless the data subject gives explicit consent for processing or the processing is necessary for a public interest, on the condition that appropriate safeguarding measures are taken for the rights, freedoms and legitimate interests of the data subject.[135] However, even if there is consent for automated decision making, the consent should not end in discrimination because the consent does not cover the latter.[136] Another problem with processing sensitive data is sometimes inferring sensitive data from non-sensitive data, for example, social media likes or posts.[137] The non-sensitive data could also be a "proxy" for sensitive data, such as the birth of place

[129] GERARDS AND XENIDIS, *supra* note 19 at 110.
[130] Hacker, *supra* note 19 at 1177.
[131] EUROPEAN PARLIAMENT DIRECTORATE GENERAL FOR PARLIAMENTARY RESEARCH SERVICES, THE IMPACT OF THE GENERAL DATA PROTECTION REGULATION ON ARTIFICIAL INTELLIGENCE 61 (2020), https://data.europa.eu/doi/10.2861/293.
[132] Hacker, *supra* note 19 at 1177.
[133] *Id.*
[134] GERARDS AND XENIDIS, *supra* note 19 at 110.
[135] EUROPEAN PARLIAMENT DIRECTORATE GENERAL FOR PARLIAMENTARY RESEARCH SERVICES, *supra* note 131 at 62.
[136] *Id.* at 62.
[137] *Id.*

being a proxy for ethnicity; even though not directly revealing the sensitive data, both cases could lead to discrimination.[138]

Furthermore, processing data considered sensitive or special categories of data are allowed under strict conditions under Article 9 of the GDPR.[139] Discrimination could occur if there is a use of sensitive data in AI systems, as some of this type of data coincide with discrimination grounds such as race or religion, and the rest of this data could potentially be considered as "other discrimination grounds."[140] However, sometimes problems could emerge due to not processing sensitive data, as it would make proving discrimination more difficult.[141]

On the other hand, The Consultative Committee of the CoE Convention no. 108 has published "Guidelines on Artificial Intelligence and Data Protection," which focuses explicitly on the data protection aspect of human rights impacts concerning artificial intelligence, encompassing provisions for different artificial intelligence actors such as developers, service providers, legislators, and policy makers.[142] Personal data also becomes relevant when automated systems are used for profiling, which could result in discrimination.[143] Just as big data, so too personal data is pertinent for transparency and explainability discussions on artificial intelligence.[144] For the explained reasons, personal data will be mentioned in this study when it is relevant and necessary but left out from

---

[138] *Id.*
[139] Zuiderveen Borgesius, *supra* note 25 at 1581 Special categories of data are listed under Article 9(1) of the GDPR and include "*racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation."* .
[140] *Id.* at 1581; Michael Veale & Reuben Binns, *Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data*, 4 BIG DATA SOC. 1, 4 (2017), https://doi.org/10.1177/2053951717743530 (last visited May 10, 2023).
[141] Zuiderveen Borgesius, *supra* note 25 at 1581; Olivier De Schutter & Julie Ringelheim, *Ethnic Profiling: A Rising Challenge for European Human Rights Law*, 71 MOD. LAW REV. 358, 376 (2008), http://www.jstor.org/stable/25151207 (last visited Aug 15, 2023); *See also* Veale and Binns, *supra* note 140 at 4.
[142] Council of Europe Directorate General of Human Rights and Rule of Law, *Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108) - Guidelines on Artificial Intelligence and Data Protection*, (2019), https://rm.coe.int/human-rights-and-business-recommendation-cm-rec-2016-3-of-the-committe/16806f2032 (last visited Aug 4, 2022).
[143] ZUIDERVEEN BORGESIUS, *supra* note 19 at 40; Tanna and Dunning, *supra* note 128 at 431; Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), *supra* note 120 at Recital 71.
[144] Zuiderveen Borgesius, *supra* note 25 at 1580; Tanna and Dunning, *supra* note 128 at 430.

being the main scope of the study as it would shift its perspective to personal data protection and related rights, which calls for an entire respective study.[145]

1.1.5 Ambit of Artificial Intelligence

Artificial intelligence comprises a diverse set of scientific branches, theoretical approaches, and methods in its ambit.[146] In order to better understand artificial intelligence and the scope of the study, it will be helpful to examine the subfields of artificial intelligence. These subfields differ in their particularities, applying different methods.[147] While it is not the entirety, the subfields of artificial intelligence include machine learning, artificial neural networks, natural language processing, speech recognition, robotics, and computer vision.[148]

Machine Learning

As mentioned above, EU AI Act Proposal lists machine learning as an artificial intelligence approach under Annex I. CoE Commissioner for Human Rights defines machine learning as:

> "A field of AI made up of a set of techniques and algorithms that can be used to "train" a machine to automatically recognise patterns in a set of data. By recognising patterns in data, these machines can derive models that explain the data and/or predict future data. In summary, it is a machine that can learn without being explicitly programmed to perform the task." [149]

Machine learning is what usually comes to mind when artificial intelligence is put into words.[150] Although artificial intelligence needs machine learning in general, it is a field of artificial intelligence and not an equivalent for it in all cases.[151]

---

[145] Some authors suggest AI fairness and discrimination related matters should be addressed by data protection law and with GDPR. See e.g. Hacker, *supra* note 19 at 1170–1183.
[146] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 20.
[147] NORVIG AND RUSSELL, *supra* note 24 at 19; John A Bullinaria, *IAI : The Roots, Goals and Sub-Fields of AI*, 10 (2005), https://www.cs.bham.ac.uk/~jxb/IAI/w2.pdf.
[148] Bullinaria, *supra* note 147 at 10.
[149] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 24.
[150] NORVIG AND RUSSELL, *supra* note 24 at 19; ALPAYDIN, *supra* note 52 at 18; Ebers, *supra* note 27 at 205; Zuiderveen Borgesius, *supra* note 25 at 1574; Michael Jordan, *Artificial Intelligence — The Revolution Hasn't Happened Yet*, MEDIUM (Apr. 30, 2018), https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7 (last visited Aug 12, 2022); Author Zachary C. Lipton, *From AI to ML to AI: On Swirling Nomenclature & Slurried Thought*, APPROXIMATELY CORRECT (Jun. 5, 2018), https://www.approximatelycorrect.com/2018/06/05/ai-ml-ai-swirling-nomenclature-slurried-thought/ (last visited Aug 12, 2022).
[151] ALPAYDIN, *supra* note 52 at 18; NORVIG AND RUSSELL, *supra* note 24 at 19; ZUIDERVEEN BORGESIUS, *supra* note 19 at 13.

Machine learning comprises learning techniques for computers.[152] Machine learning works not by having the inputs given by the programmers but by examining the examples and data and making inferences from them, resulting in the building or advancing software.[153] Machine learning requires data and machines cannot 'learn' without data because it is the leading component in machine learning.[154] Data becomes the decisive authority of what is going to be conducted.[155] With the booming of the internet, which was later followed by big data, the last two decades carried machine learning to a better place due to massive resource data.[156] Data becomes valuable through data mining. Data mining helps to pull out patterns of tacit information that are not directly deducible.[157] Such information could be found online, in 'data repositories', 'data streams', or similar 'storerooms.'[158]

Machine learning enables conducting things where it is not possible to give straightforward instructions in programming by providing exemplary conduct, thus it presents an opportunity to achieve things when the desired goal, such as recalling a face from a photograph or being able to walk in a busy street without bumping into somebody, is unfit to be represented through standard algorithms.[159] How does a machine learn to do such things? Alpaydın explains that machine learning begins with a "general model" that may perform differently with different parameters.[160] Modifying these parameters to accord with the training model constitutes the learning process.[161] Training makes the general model conform to particular tasks.[162] Thus, firstly, the algorithms are not as they are used in regular computer programs for machine learning, yet they exist as such

[152] Thomas G Dietterich, *Machine Learning*, ENCYCLOPEDIA OF COGNITIVE SCIENCE (Lynn Nadel ed., 2006), https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470018860.s00036 DOI: 10.1002/0470018860.s00036.
[153] Thomas G Dietterich, *Machine Learning*, ENCYCLOPEDIA OF COMPUTER SCIENCE (Edwin D. Reilly, Anthony Ralston, & David Hemmendinger eds., 4 ed. 2003).
[154] NORVIG AND RUSSELL, *supra* note 24 at 723; ALPAYDIN, *supra* note 52 at 12.
[155] ALPAYDIN, *supra* note 52 at 12.
[156] *Id.* at xiii; ZUIDERVEEN BORGESIUS, *supra* note 19 at 13.
[157] JIAWEI HAN, JIAN PEI & HANGHANG TONG, DATA MINING: CONCEPTS AND TECHNIQUES xxi (2022); ZUIDERVEEN BORGESIUS, *supra* note 19 at 14.
[158] HAN, PEI, AND TONG, *supra* note 157 at xxi; ZUIDERVEEN BORGESIUS, *supra* note 19 at 14.
[159] Dietterich, *supra* note 153; ALPAYDIN, *supra* note 53 at 1.
[160] ALPAYDIN, *supra* note 53 at 1.
[161] *Id.*
[162] *Id.*

models.[163] Secondly, learning programs take the place of computer programmers in machine learning.[164]

Generalization is an essential machine learning capacity.[165] However, generalizations are not always accurate.[166] Appropriateness of the model for the problem, the amount of data, and the optimization precision determine the correctness of the generalization.[167] Generalization may not be safe when the outcome is used in a system involving humans, especially if the generalization is incorrect. This is one occurrence of how generalization could lead to problematic outcomes that could hinder the right to equality and non-discrimination.

On a closer look, machine learning has three main learning techniques consisted of supervised learning, unsupervised learning, reinforcement learning. Firstly, there is supervised learning where learning is done by studying pairings of inputs and outputs.[168] In supervised learning, outputs act as expositions for the inputs, and these outputs are called labels.[169] For example, inputs could be pictures of books labeled with the word 'book'; when the machine comes upon a book picture after studying the labels, it can now guess the label, which is the 'book.'[170] The supervisor should not introduce inaccurate limits and bias into the learning process.[171]

Secondly, in unsupervised learning, the machine is not given labels as in the case of supervised learning.[172] Instead, it often uses clustering as a method, which operates by identifying input collections, and then being able to cluster them when encountered.[173] For example, the machine could cluster book pictures after being fed many book pictures as input.[174] There is only input data whereas the output is undefined, and there is no supervisor in unsupervised learning.[175] Unlike in supervised learning, no supervisor depicts the labels and classes for the training.[176] Unsupervised learning monitors what

[163] *Id.*
[164] ALPAYDIN, *supra* note 52 at xiii.
[165] *Id.* at 51.
[166] *Id.*
[167] *Id.*
[168] NORVIG AND RUSSELL, *supra* note 24 at 671.
[169] *Id.*
[170] *Id.*
[171] ALPAYDIN, *supra* note 52 at 150.
[172] NORVIG AND RUSSELL, *supra* note 24 at 671.
[173] *Id.*
[174] *Id.*
[175] ALPAYDIN, *supra* note 52 at 143.
[176] *Id.* at 150.

happens under typical situations by detecting regularities in the inputs—comparing the frequency of patterns.[177] In this case, patterns and regularities with insignificant changes in various fields are anticipated.[178] The primary purpose of unsupervised learning is to determine the data structure by abstracting the characteristics and what is typical of the data.[179] The data becomes ready for labeling for different purposes after finding what is typical for the data.[180]

Thirdly, reinforcement learning means learning from taking actions and then being met with awards or penalties and determining future actions according to the previous award or penalty.[181] Game-playing machines such as online chess engines often get better at playing the game through reinforcement learning.[182] An agent interacting with the environment is the decision-maker in reinforcement learning.[183]

The agent makes various actions, some of which are met with rewards and some not.[184] Different action sets may be required for an assignment, and the agent learns the best possible actions to reach the reward; for instance, the reward is winning the game in chess, so the agent would learn which actions would lead toward winning the game.[185] Reinforcement learning differs from supervised learning in the lack of supervision; there is 'criticism' rather than 'supervision.'[186] Whether there is a reward is not known before the conduct of actions.[187] Reinforcement learning helps to interpret the actions on their usefulness to an end goal.[188]

Artificial Neural Networks

Another artificial intelligence approach is artificial neural networks. Human neural systems start functioning with neurons, which get electrochemical signals from their dendrites and send these from one axon to another.[189] Artificial neural networks take

[177] *Id.* at 143.
[178] *Id.* at 144.
[179] *Id.* at 150.
[180] *Id.*
[181] NORVIG AND RUSSELL, *supra* note 24 at 671.
[182] *Id.*
[183] ALPAYDIN, *supra* note 52 at 161.
[184] *Id.*
[185] *Id.* at 162.
[186] *Id.*
[187] *Id.* at 162–163.
[188] *Id.* at 163.
[189] Roy de Kleijn, *Artificial Intelligence Versus Biological Intelligence: A Historical Overview*, *in* LAW AND ARTIFICIAL INTELLIGENCE: REGULATING AI AND APPLYING AI IN LEGAL PRACTICE 29, 34 (Bart Custers & Eduard Fosch-Villaronga eds., 2022), https://doi.org/10.1007/978-94-6265-523-2_2.

inspiration from real neurons and work by transferring activations between different units.[190] These units are composed of different layers; an input layer, an output layer, and more layers when needed.[191]

Neurons collect the different activations sent through synapses and measure their synaptic weights.[192] If the activation weighs more than the threshold value, the neuron is fired, the activation is sent to other neurons, and the output matches the activation.[193] Learning algorithms in artificial neural networks aim to calculate the connection weights of the neurons.[194]

Deep Learning

A different artificial intelligence approach, deep learning, is the utilization of basic layers of adjustable computing elements for machine learning.[195] Even though deep learning roots back to the 1970s, it is an area that has been booming recently.[196] Deep learning is used for various tasks, including speech recognition, machine translation, visual object recognition, and medical diagnosis.[197] It should be noted that *robust* hardware systems are required for deep learning.[198]

In deep learning, hidden layers combine the values in the last layer and learn more complex input functions.[199] The need for intervention from outside is minor in deep learning.[200] Feature adjustments do not require to be done by hand.[201] Big data enables the learning algorithm to analyze what is required by itself.[202]

[190] *Id.* at 34–35.
[191] *Id.* at 35.
[192] ALPAYDIN, *supra* note 52 at 107.
[193] *Id.*
[194] *Id.* at 109.
[195] NORVIG AND RUSSELL, *supra* note 24 at 44.
[196] *Id.*
[197] *Id.* at 44–45.
[198] See the heading 3.4.2 on "robustness" *Id.* at 45.
[199] ALPAYDIN, *supra* note 52 at 127.
[200] *Id.* at 129.
[201] *Id.*
[202] *Id.*

1.2 Bias and Fairness in Artificial Intelligence

1.2.1 Bias in Artificial Intelligence

Addressing discrimination through artificial intelligence requires recognizing technical elements, data sources, which the study aimed to present until this point, and biases that initially come from humans but have additional connotations in artificial intelligence.[203] Bias emanates from human cognition, meaning that where there are humans, there is bias.[204] Considering artificial intelligence techniques observe examples sourced from humans or aim to be similar to human thinking; bias is inevitably reflected in artificial intelligence systems or in the decisions that are made by those systems.[205] Artificial intelligence may result in spreading or aggravating human bias.[206] As will be explained, algorithmic bias also connects to legal and philosophical discussions on addressing social inequalities.[207]

Data bias has different meanings and different types.[208] Bias may result from data, programmers, and the algorithm itself.[209] As the machine learns by feeding from data, the outcome of a task will mirror any biases in the data: "garbage in, garbage out" is a catchphrase to describe the influence of data on machine learning, because inadequacies in data are one of the leading causes of problematic outcomes as machines train on data.[210] Similar sayings exist for bias and discrimination. Bias based on data is called "bias in,

[203] Anna Lauren Hoffmann, *Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse*, 22 INF. COMMUN. SOC. 900, 904 (2019), https://doi.org/10.1080/1369118X.2019.1573912.
[204] Tanna and Dunning, *supra* note 128 at 423.
[205] *Id.* at 426; See for an overview of human to data bias types Ninareh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning*, 54 ACM COMPUT. SURV. 115:1, 8 (2021), https://doi.org/10.1145/3457607 (last visited May 30, 2022).
[206] Hacker, *supra* note 19 at 1144; Toon Calders & Indrė Žliobaitė, *Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY: DATA MINING AND PROFILING IN LARGE DATABASES 43, 43 (Bart Custers et al. eds., 2013), https://doi.org/10.1007/978-3-642-30487-3_3 (last visited Jun 24, 2022); Indrė Žliobaitė, *Measuring Discrimination in Algorithmic Decision Making*, 31 DATA MIN. KNOWL. DISCOV. 1060, 2 (of the preprint edition.) (2017), http://link.springer.com/10.1007/s10618-017-0506-1.
[207] Hacker, *supra* note 19 at 1146.
[208] Mehrabi et al., *supra* note 205 at 5.
[209] Tanna and Dunning, *supra* note 128 at 425–428.
[210] Daniel Schönberger, *Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications*, 27 INT. J. LAW INF. TECHNOL. 171, 176 (2019), https://doi.org/10.1093/ijlit/eaz004 (last visited Jul 27, 2022).

bias out"; and discrimination based on biased data is called "discrimination in, discrimination out".[211]

Bias has two meanings: first, it could indicate prejudice, which could be directed at individuals, groups, or things; this bias may influence user interaction, system design, data gathering, and interpretation, and second, bias could derive as a result due to a system error at the sampling or reporting stages.[212]

Bias could lead to discrimination during the data mining process.[213] Beyond biased training data, discrimination could arise if humans use techniques intending to hide discriminatory wills and, in cases such as when the target variable and class labels are not appropriately defined, selecting features.[214] Furthermore, if legally interpretable metrics for measuring biases are developed, these could present significant evidence for proving discrimination through artificial intelligence.[215] There exist numerous types of biases, and understanding them is central for grasping how biases can later lead to discrimination in artificial intelligence.[216] Algorithmic bias as a bias category signifies the bias emanating from algorithmic functions, methods, and application of models instead of bias deriving from input data.[217] The term algorithmic bias is also used as an overall term when referring to different types of biases.

It should be noted that there are more types and categories of bias concerning artificial intelligence beyond the ones given hereunder. A bias categorization is made by Mehrabi et al. as bias from "data to algorithms," "algorithm to user," and "user to data."[218] For example, measurement bias, representation bias, and aggregation bias fall under bias from "data to algorithm,"; evaluation bias and algorithmic bias under "algorithm to user," social bias, and historical bias under "user to data."[219] Similarly, different types of biases

---

[211] Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE LAW J. 2218, 2224 (2019), https://papers.ssrn.com/abstract=3257004 (last visited Feb 11, 2022); Tanna and Dunning, *supra* note 128 at 425.
[212] Google Developers, *Bias (Ethics/Fairness)*, MACHINE LEARNING GLOSSARY (2022), https://developers.google.com/machine-learning/glossary#bias-ethicsfairness (last visited Aug 12, 2022).
[213] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. LAW REV. 671, 677 (2016).
[214] *Id.* at 677–694; ZUIDERVEEN BORGESIUS, *supra* note 19 at 15.
[215] Xiang and Raji, *supra* note 32 at 1 See headings 1.3.2 and 2.2.1 on how to prove discrimination.
[216] *See also* IBM, *Everyday Ethics for Artificial Intelligence*, 52 34–35 (2022), https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf (last visited May 23, 2023).
[217] Mehrabi et al., *supra* note 205; Ricardo Baeza-Yates, *Bias on the Web*, 61 COMMUN. ACM 54, 58 (2018), https://dl.acm.org/doi/10.1145/3209581 (last visited Aug 15, 2023).
[218] Mehrabi et al., *supra* note 205 at 4–9.
[219] *Id.*

are put into categories by Baeza-Yates under statistical, cultural, and cognitive biases, which could stem, for example, from algorithms, data, and users.[220] Concrete examples of discrimination through artificial intelligence stemming from the types of biases described hereunder may be found in the second chapter of this study.

Automation Bias

A unique type of bias is automation bias. A computer is more trusted than humans when making decisions; when a computer or AI-system makes the decisions, people trust it more than they do themselves because they consider computers objective and do not act upon bias, unlike persons.[221] That is because decisions made by automated systems are considered to be calculable, consistent and less arbitrary compared to human decisions.[222] Automatic decisions are treated as neutral actions.[223] People trust the analytic power of automated systems more than theirs.[224] Automated systems are so trusted that the perception of objectivity is considered a feature to prefer such systems over human-made decisions when in reality, decisions made by these systems could be unobjective either.[225] This is referred as automation bias.

Automation bias could lead to detrimental effects if, for example, humans do not act upon even if they suspect an outcome to be biased or abandon inspecting automated systems for such matters with the trust in the objectivity of the systems.[226] Accordingly, automation bias raises questions of human complacency and accountability.[227] As Chesterman notes, automation bias should not be a reason for getting rid of responsibility.[228] The third chapter of this study will examine accountability and

[220] Baeza-Yates, *supra* note 217 at 58.
[221] Tanna and Dunning, *supra* note 128 at 422.
[222] CHESTERMAN, *supra* note 57 at 69.
[223] Selena Silva & Martin Kenney, *Algorithms, Platforms, and Ethnic Bias: An Integrative Essay*, 55 PHYLON 1960- 9, 22 (2018), https://www.jstor.org/stable/26545017 (last visited Jun 7, 2022).
[224] CHESTERMAN, *supra* note 57 at 68; Raja Parasuraman & Dietrich H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, 52 HUM. FACTORS 381, 392 (2010), https://doi.org/10.1177/0018720810376055 (last visited Jun 6, 2022); The plane crash incident of the Turkish Airlines plane in 2009 has been attributed to automation bias. See Robert J. De Boer, Wijnand Heems & Karel Hurts, *The Duration of Automation Bias in a Realistic Setting*, 24 INT. J. AVIAT. PSYCHOL. 287 (2014), http://www.tandfonline.com/doi/abs/10.1080/10508414.2014.949205 (last visited Jun 12, 2023).
[225] Tanna and Dunning, *supra* note 128 at 422.
[226] Parasuraman and Manzey, *supra* note 224 at 391–392.
[227] CHESTERMAN, *supra* note 57 at 68–69.
[228] *Id.* at 69.

responsibility in more detail. If automation bias becomes ingrained in decision-makers, it may lead to data fundamentalism, that is, the belief that data reflects objective truth.[229]

Historical Bias

Bias may be rooted from history.[230] Historical biases that people are reflected in daily lives, which also applies to data. This conveys that even if data collection does not have bias *per se,* and sampling is done accurately, data can still be biased.[231] It is a work that requires much effort to eliminate the existing historical bias.[232]

Data mining could lead to worsening of historical bias.[233] Data mining also has the potential to lead people into fixing the problematic biases, as eliminating the historical bias is going to require an active intervention.[234]

Other types of bias are found in automated systems, differing from automation data, which is a bias that humans hold about automated systems.

Social Bias

Social bias in the context of artificial intelligence refers to bias coming from other individuals affecting one's decisions and reasoning.[235] A more general understanding of social bias suggests making decisions concerning a group or an individual based on certain perceptions regarding their group, such as ethnicity, gender, or race.[236] Social biases could also turn into societal biases when they are widespread.[237] Social and societal biases are prevalent causes for discrimination through artificial intelligence, as the socially biased data spreads bias in algorithms or AI systems, deepening the existing bias.[238]

---

[229] Tanna and Dunning, *supra* note 128 at 428.
[230] Harini Suresh & John V. Guttag, *A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle*, *in* EQUİTY AND ACCESS İN ALGORİTHMS, MECHANİSMS, AND OPTİMİZATİON 1, 4 (2021), http://arxiv.org/abs/1901.10002 (last visited Oct 4, 2022); Tanna and Dunning, *supra* note 128 at 425.
[231] Suresh and Guttag, *supra* note 230 at 4.
[232] Barocas and Selbst, *supra* note 213 at 672; Tanna and Dunning, *supra* note 128 at 425.
[233] Barocas and Selbst, *supra* note 213 at 674.
[234] *Id.* at 676; Tanna and Dunning, *supra* note 128 at 425.
[235] Mehrabi et al., *supra* note 205 at 8; Baeza-Yates, *supra* note 217 at 60.
[236] Bertrand K. Hassani, *Societal Bias Reinforcement through Machine Learning: A Credit Scoring Perspective*, 1 AI ETHICS 239, 239 (2021), https://doi.org/10.1007/s43681-020-00026-z (last visited Aug 15, 2023).
[237] *Id.*
[238] *Id.*

Learning Bias

Bias not only comes from data but also occurs during learning processes which could augment the bias.[239] There is a learning bias if modeling choices increase the performance differences for different samples.[240]

Measurement Bias

Feature and label selection, gathering, and processing stages for a prediction problem could result in measurement bias. [241] This type of bias arises in case a proxy facilitating measurement is inadequate to represent what is to be calculated.[242] Overly simplified data may overlook relevant factors for different assessments when there are many different factors.[243]

Representation Bias

Representation bias appears when data fails to sufficiently represent a part of a population.[244] Representation bias may emerge in three different ways. First, the population targeted and the population that the data is to be used may not overlap.[245] Secondly, the target population may have subgroups which their representation may lack in the data.[246] Thirdly, the procedure used for sampling may be limited during the sampling of the target population.[247]

Evaluation Bias

Evaluation bias happens when the data taken as the reference point fails to represent a group.[248] That could lead to only the sub-categories of the reference data being correctly represented.[249] Comparison of the learning models quantitatively carries the risk of evaluation bias if it is done for generalizing the quality of the model.[250] Performance

[239] Sara Hooker, *Moving beyond "Algorithmic Bias Is a Data Problem,"* 2 PATTERNS N. Y. N 100241, 1 (2021).
[240] Suresh and Guttag, *supra* note 230 at 5; Hooker, *supra* note 239 at 1.
[241] Suresh and Guttag, *supra* note 230 at 5; See Mehrabi et al., *supra* note 205 at 4.
[242] Suresh and Guttag, *supra* note 230 at 5.
[243] Barocas and Selbst, *supra* note 213 at 688; Suresh and Guttag, *supra* note 230 at 5; ZUIDERVEEN BORGESIUS, *supra* note 19 at 20.
[244] Suresh and Guttag, *supra* note 230 at 4; See Mehrabi et al., *supra* note 205 at 5.
[245] Suresh and Guttag, *supra* note 230 at 4.
[246] *Id.*
[247] *Id.*
[248] *Id.* at 6.
[249] *Id.*
[250] *Id.*

documenting metrics may again deepen evaluation bias in cases such as only one type of metric is chosen for measurement.[251]

Aggregation Bias

Aggregation bias arises when a single type of data model is used when different models should have been used for different groups or examples.[252] The model assumes that inputs and labels have the exact input-label mapping, whereas the variables have different meanings for different data subsets.[253] The result may not be ideal for either of the groups if the used dataset has a variable with different meanings for different groups, as there is a misconstrued consistency of input-label mappings.[254]

Deployment Bias

Deployment bias occurs when the problem is inconsistent with the way the model is used, even though the problem requires a different model.[255] Deployment bias may arise where a system is built and regarded as fully autonomous; in fact, it is not and involves different human decision-makers.[256] When human decision-makers interpret the results before using them, there could be deployment bias if they rely on the data as truth, acting with automation bias.[257]

1.2.2 Fairness in Artificial Intelligence

Fairness is understood and addressed in various ways in the artificial intelligence field.[258] Artificial intelligence is a subject that many researchers study extensively, not only its technical dimensions but also its social impacts. In academic discussions regarding artificial intelligence and its detrimental effects on humans, many studies focus on the notion of fairness.[259] These studies mostly approach fairness from the ethics

[251] *Id.*
[252] *Id.*; See Mehrabi et al., *supra* note 205 at 5.
[253] Suresh and Guttag, *supra* note 230 at 5.
[254] *Id.*
[255] *Id.* at 6.
[256] *Id.*
[257] *Id.*
[258] Mehrabi et al., *supra* note 205 Authors make detailed examination of different understandings and approaches to fairness.
[259] *See, e.g.,* Sahil Verma & Julia Rubin, *Fairness definitions explained*, *in* PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON SOFTWARE FAIRNESS 1 (2018), https://doi.org/10.1145/3194770.3194776 (last visited Aug 3, 2022); Philipp Hacker, *Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law*, 55 COMMON MARK. LAW REV. (2018), https://kluwerlawonline.com/journalarticle/Common+Market+Law+Review/55.4/COLA2018095 (last visited Nov 9, 2021); Allison Woodruff et al., *A Qualitative Exploration of Perceptions of Algorithmic*

perspective, interpreting the notion as respecting to various ethical requirements.[260] The practitioners in the artificial intelligence field also aim to address the fairness problem. However, trying to "convert" different fairness theories in machine learning is difficult.[261] Some of the prominent players in the field of artificial intelligence have put toolkits into action to help artificial intelligence developers mitigate fairness issues.[262] These toolkits are introduced to control 'unwanted' bias and handle related bias issues.[263] These tools allow working on datasets to achieve better-quality data by controlling data in ways including corruption, sensitiveness, gaps, and attribution balance.[264] Technical fairness procedures may help legal practitioners to examine the occurrence of discrimination.[265] However, the notion of fairness requires more clarification to achieve genuine and efficient fairness. Furthermore, as explained in Chapter 3, fairness comes up as a principle in documents regulating artificial intelligence. The notion of fairness lays at the center of discussions on equality and non-discrimination in artificial intelligence.

Fairness, *in* PROCEEDINGS OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 1 (2018), https://dl.acm.org/doi/10.1145/3173574.3174230 (last visited Jun 19, 2022); Alice Xiang & Inioluwa Deborah Raji, *On the Legal Compatibility of Fairness Definitions*, (2019), http://arxiv.org/abs/1912.00761 (last visited Jun 3, 2022); Reuben Binns, *On the Apparent Conflict Between Individual and Group Fairness*, (2019), http://arxiv.org/abs/1912.06883 (last visited Aug 21, 2022); Daniel Greene, Anna Lauren Hoffmann & Luke Stark, *Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning* (2019), http://scholarspace.manoa.hawaii.edu/handle/10125/59651 (last visited Oct 26, 2021); Andreas Kaplan & Michael Haenlein, *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, 62 BUS. HORIZ. 15 (2019); Songül Tolan, *Fair and unbiased algorithmic decision making: current state and future challenges*, ARXIV PREPR. ARXIV190104730 (2019); Ninareh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning*, 54 ACM COMPUT. SURV. 115:1 (2021); Sandra Wachter, Brent Mittelstadt & Chris Russell, *Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI*, 41 COMPUT. LAW SECUR. REV. 105567 (2021); Tao Zhang et al., *Fairness in Semi-Supervised Learning: Unlabeled Data Help to Reduce Discrimination*, 34 IEEE TRANS. KNOWL. DATA ENG. 1763 (2022); Unfairness By Algorithm: Distilling the Harms of Automated Decision-Making - Future of Privacy Forum, HTTPS://FPF.ORG/, https://fpf.org/blog/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/ (last visited Nov 9, 2021); Kathleen A. Creel & Deborah Hellman, *The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems*, CAN. J. PHILOS. (2022).

[260] GERARDS AND XENIDIS, *supra* note 19 at 48.

[261] Reuben Binns, *Fairness in Machine Learning: Lessons from Political Philosophy*, *in* PROCEEDINGS OF THE 1ST CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY 149, 9 (2018), https://proceedings.mlr.press/v81/binns18a.html (last visited May 10, 2023).

[262] Kush R. Varshney, *Introducing AI Fairness 360, A Step Towards Trusted AI - IBM Research*, IBM RESEARCH BLOG (Sep. 19, 2018), https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/ (last visited Aug 13, 2022); The PAIR Team Google, *Know Your Data*, https://knowyourdata.withgoogle.com (last visited Jul 26, 2022).

[263] Varshney, *supra* note 262; The PAIR Team Google, *supra* note 262.

[264] Kush R. Varshney, *Introducing AI Fairness 360, A Step Towards Trusted AI - IBM Research*, IBM RESEARCH BLOG (2018), https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/ (last visited Jul 25, 2022); The PAIR Team Google, *Know Your Data Documentation*, https://knowyourdata.withgoogle.com/docs/ (last visited Jul 26, 2022).

[265] Wachter, Mittelstadt, and Russell, *supra* note 19 at 48.

What is understood by fairness is essential.[266] Does fairness mean ethical fairness, and if so, according to which understanding of ethics? Even if the fairness criteria are deemed to be assured in machine learning provide fairness in a preferred ethical meaning, to examine whether such criteria will provide fairness in law is indispensable. Fairness has undoubtedly an ethical connotation, even in the legal context. Fairness is a notion that has different definitions and interpretations, which are primarily subjective and vague.[267] Defining fairness is a difficult task for law as well, as it is not easy find objective criteria for the term in positive law. Yet, the law can be considered as codification of fairness defining its broad concept.[268] The most precise explanations come from international human rights law to better understand fairness: the right to equality and non-discrimination.[269] In human rights law, fairness assessment is conducted contextually by making judicial interpretations when defining the relevant concepts.[270] The gap between the fairness evaluation methods in the artificial intelligence field and the legal field should be addressed to compensate for potential inconsistencies of discrimination in the legal sense.[271] Wachter et al. underline that "contextual equality" in the legal sense of fairness cannot entirely be automated as it may lack an understanding of social, political, and other relevant contextual factors.[272] Instead of aiming to achieve fairness only with technical means, establishing evidentiary material for assessing fairness by jurists could better analyze the right to equality and non-discrimination.[273]

When mentioning fairness, another question is whether it is equivalent to justice. Rawls constructed his theory of fairness on the notions of liberty and equality of opportunity to bring fairness and justice together.[274] Put simply; liberty should be for everybody partaking or being impacted by a practice, according to Rawls.[275] Second, everybody should have the possibility to be in a position.[276] 'Inequality' is only acceptable

[266] Emre Kazim et al., *Automation and Fairness*, *in* ARTIFICIAL INTELLIGENCE: LAW AND REGULATION 492, 493 (2022).
[267] Tanna and Dunning, *supra* note 128 at 434; GERARDS AND XENIDIS, *supra* note 19 at 47.
[268] Kazim et al., *supra* note 266 at 496.
[269] McGregor, Murray, and Ng, *supra* note 1 at 326.
[270] Wachter, Mittelstadt, and Russell, *supra* note 19 at 45.
[271] *Id.* at 4.
[272] *Id.* at 64, 66.
[273] *Id.* at 55.
[274]John Rawls, *Justice as Fairness*, 67 PHILOS. REV. 164, 164–165 (1958), https://www.jstor.org/stable/2182612 (last visited Oct 25, 2022).
[275] *Id.* at 165.
[276] *Id.*

for Rawls if it serves everybody's advantage.[277] Since the principle of equality and non-discrimination is regulated by law, fairness must coincide with the regulations in law. Who will determine that an AI-system or a machine learning model is fair? Even if something is ethically fair, it may not have the qualifications to be fair in the legal sense. After all, ethical understandings may change from one person to another.

For this reason, it is necessary to examine how similar the definitions of fairness in artificial intelligence and those in law are.[278] The agreement of these definitions will also manifest in judicial processes, and the more they agree, the judges will have the chance to better analyze the technical details of artificial intelligence in the legal sense.[279] There must be no risk of discrimination in the legal sense in the use of methods called "bias correction techniques."[280] In essence, any fairness interpretation that does not provide fairness in law will be incomplete, and true fairness will not be provided.

Outcome fairness and procedural fairness may relate to equality of outcomes and formal equality, where the first notion refers to the allocation of resources, and the latter refers to fairness within procedures.[281] Fairness could fail when using artificial intelligence leads to less favorable treatment of a group.[282] Indeed, fairness is not only a problem relating to groups. It is about individuals as well. Following that, individual fairness and group fairness should be defined.[283] Individual fairness is defined in the context of artificial intelligence as treating individuals similarly to similar individuals without caring about the classes the individuals belong to, which reminds Aristotle's understanding of equality explained below.[284] Group fairness refers to the two different classes being treated similarly based on a unit of statistics.[285] The preference for the type of fairness also has importance because, for example, achieving individual fairness may not achieve group fairness.[286] Another approach in artificial intelligence for achieving

---

[277] *Id.*
[278] Xiang and Raji, *supra* note 32 at 1; GERARDS AND XENIDIS, *supra* note 19 at 47.
[279] Xiang and Raji, *supra* note 32 at 1.
[280] *Id.*
[281] Kazim et al., *supra* note 266 at 494 See heading 1.3.1 on equality of outcomes and formal equality.
[282] Tanna and Dunning, *supra* note 128 at 434.
[283] *See also* Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. LAW REV. 811, 835 (2020), https://papers.ssrn.com/abstract=3418528 (last visited Sep 11, 2022).
[284] NORVIG AND RUSSELL, *supra* note 24 at 1043.
[285] *Id.* at 1044; IBM, *AI Fairness 360 - Resources*, IBM RESEARCH TRUSTED AI, https://aif360.mybluemix.net/aif360.mybluemix.net/resources (last visited Aug 13, 2022).
[286] *See* Reuben Binns, *On the Apparent Conflict Between Individual and Group Fairness*, (2019), http://arxiv.org/abs/1912.06883 (last visited Aug 21, 2022).

fairness is disregarding or un-seeing specific attributes.[287] The idea underlying this approach of fairness is if gender, age, ethnic origin, race or similar attributes are disregarded, the artificial intelligence system is not able to discriminate on these attributes.[288] However, in reality this may not be true. Is a type of fairness a better option? In artificial intelligence, fairness is an overall-meaning term designating an objective in a way that doesn't necessarily distinguish between approaches to the notions of equality and discrimination.[289] Different equality approaches carry out explanations for fairness, which are also explained below. The artificial intelligence field aims to reach fairness mathematically. There are efforts to mitigate fairness in ethics with its interpretation in the artificial intelligence field.

The lack of a consistent definition of fairness is among the challenges to reaching fairness in artificial intelligence.[290] Indeed, different fairness definitions may not result in fair outcomes under particular circumstances.[291] A variety of definitions are found for fairness in the artificial intelligence field.[292] A technical approach to fairness aims to implement machine learning algorithms to secure fairness.[293] However, the problem of reaching undesirable and detrimental outcomes through a seemingly fair learning process endures.[294] On the other hand, while there is no single fairness definition yet, making such a definition may not bring fairness because it may not be enough to address each unique situation.[295] Fairness and bias are significant in the discussions of data justice.[296] Various bias and fairness definitions coordinate with legal definitions for equality and non-discrimination, yet, discrimination in the legal sense would require different treatment or

---

[287] NORVIG AND RUSSELL, *supra* note 24 at 1044.
[288] *Id.*
[289] Binns, *supra* note 261 at 2–3.
[290] Mehrabi et al., *supra* note 205 at 25.
[291] *Id.*
[292] Nripsuta Saxena et al., *How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness*, 1 (2019), http://arxiv.org/abs/1811.03654 (last visited Oct 6, 2022).
[293] Creel and Hellman, *supra* note 259 at 27.
[294] *Id.* at 27.
[295] Saxena et al., *supra* note 292 at 1.
[296] Hoffmann, *supra* note 203 at 900; Data justice as a research field aims to handle data-related matters under a framework of three pillars firstly set out by Linnet Taylor, which is "visibility," dealing with privacy and representation in data; "engagement with technology," dealing with autonomy in making choices relating to technology and benefits of data, and "non-discrimination" dealing with prevention of discrimination and challenging data biases. Linnet Taylor, *What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally*, 4 BIG DATA SOC. 2053951717736335, 9 (2017), https://doi.org/10.1177/2053951717736335; See Lina Dencik & Javier Sanchez-Monedero, *Data Justice*, 11 INTERNET POLICY REV. (2022), https://policyreview.info/articles/analysis/data-justice (last visited Jul 23, 2023).

impact due to a discrimination ground, as will be explained in the following section.[297] Fairness definitions in ethics and their interpretation in the artificial intelligence field, while overlying with equality and non-discrimination, are more expansive than the legal definitions for equality and non-discrimination.[298] Nevertheless, legal notions of equality and non-discrimination provide a more straightforward framework for setting out legal obligations than the notion of fairness.[299]

Studies have measured people's perception of fairness in artificial intelligence to achieve well-fitted fairness in different contexts.[166] If an AI-system makes a fairness assessment before reaching a conclusion, it needs to consider different data and make a preference based on that.[300] There are three leading fairness definitions in the context of artificial intelligence. Dwork et al. suggested that fairness could be achieved by similar treatment at the classification stage.[301] The first definition interprets fairness as similar treatment of similar individuals, referring to formal equality.[302] Second, Joseph et al., by moving on the Rawlsian fairness, aimed at reaching fairness by *"never favoring a worse individual than a better one"*.[303] This definition aims to reach fair results by not favoring a 'worse' individual over a 'better' individual.[304] Third, Liu et al. further contributed to Dwork et al.'s work by bringing "calibrated fairness," which considers that the chances of a group realizing the best qualities should be equal to the chances of choosing a group.[305] The third definition brings merits into the equation and seeks to select individuals in proportion to their merits.[306]

Data preparation, model learning, and post-processing steps of machine learning should be fair.[307] Addressing fairness with technical means should not be "one-dimensional"; otherwise, it could lead to overlooking discrimination and inequalities.[308]

---

[297] GERARDS AND XENIDIS, *supra* note 19 at 47.
[298] *Id.* at 48.
[299] *Id.*
[300] Kazim et al., *supra* note 266 at 497.
[301] Cynthia Dwork et al., *Fairness Through Awareness*, 1, 18 (2011), http://arxiv.org/abs/1104.3913 (last visited Oct 25, 2022).
[302] Saxena et al., *supra* note 292 at 2.
[303] Matthew Joseph et al., *Rawlsian Fairness for Machine Learning*, 1, 4 (2022), https://arxiv.org/pdf/1610.09559v3.pdf.
[304] Saxena et al., *supra* note 292 at 2.
[305] Yang Liu et al., *Calibrated Fairness in Bandits*, 1, 1 (2017), http://arxiv.org/abs/1707.01875 (last visited Oct 25, 2022).
[306] Saxena et al., *supra* note 292 at 2.
[307] Binns, *supra* note 261 at 9.
[308] Hoffmann, *supra* note 203 at 910; Matthew Le Bui & Safiya Umoja Noble, *We're Missing a Moral Framework of Justice in Artificial Intelligence: On the Limits, Failings, and Ethics of Fairness*, *in* THE

Therefore, this should be done so that it matches fairness in law. Looking at legal definitions and norms of equality is essential to address unfair results in artificial intelligence, particularly prior to making automated assessments.[309] Of course, there are certain normative limitations concerning equality and discrimination constructions in law; for this reason, it should also be kept in mind to fit into the existing legal norms related to fairness and give room to evolve simultaneously.[310] There is an argument in the fairness in artificial intelligence literature that philosophical approaches to fairness give a better understanding for resolving fairness-related issues than only trying to follow the legal understanding of equality, as it may result in dry conformity with equality norms and the protected classes without a solid background.[311] Binns argues that true justice cannot be realized if only the legal approach is taken because that would not deliver satisfactory answers to the social contexts.[312] Indeed, altering the theories of equality and non-discrimination for artificial intelligence is challenging.[313] Undeniably, philosophical and ethical approaches are essential for laying out a foundation and as a tool for adapting the fairness issues in artificial intelligence to the general notion of fairness. Nevertheless, complying with the legal connotation of fairness is also a requirement, even though the existing legal framework may not entirely fit artificial intelligence.[314] Legal norms related to fairness may and should be improved with the philosophical interpretations of fairness, yet the importance of compatibility with existing legal norms should not be undermined. Artificial intelligence designers should not disregard the legal meaning and the norms related to fairness because they find it limiting. On the other hand, legal practitioners should also not disregard other perspectives and discussions in other disciplines, mainly ethics, to contribute in their approach to artificial intelligence and build a better protection mechanism in law so that the fairness can be genuinely realized. Narrowing fairness with a set of statistics, even though taken from the law may lack context and ethical meanings of fairness, provides critical considerations for improving the understanding of fairness

OXFORD HANDBOOK OF ETHICS OF AI 0, 170 (Markus D. Dubber, Frank Pasquale, & Sunit Das eds., 2020), https://doi.org/10.1093/oxfordhb/9780190067397.013.9 (last visited May 20, 2023).
[309] Kazim et al., *supra* note 266 at 502.
[310] Binns, *supra* note 261 at 9; GERARDS AND XENİDİS, *supra* note 19 at 48.
[311] Binns, *supra* note 162 at 9.
[312] Binns, *supra* note 261.
[313] *Id.* at 9.
[314] GERARDS AND XENIDIS, *supra* note 19 at 48.

in the legal sense for artificial intelligence.[315] There needs to be a concrete middle ground between the fairness in law and ethics when evaluating the notion of fairness in artificial intelligence as it has vital importance in realizing the right to equality and non-discrimination. Bringing together the benefits of the interpretation of fairness in the fields of AI, ethics, and law is the best-fit solution for addressing discrimination through artificial intelligence.[316]

Fairness automation tools should consider that the legal interpretation of fairness requires a contextual interpretation of the law.[317] While large-scale automated fairness is difficult to achieve, it may not be wanted to be achieved as removing the contextual nature of the fairness may contradict the aim of the right to equality and non-discrimination.[318] As mentioned above, in order to evaluate something as fair and just in law, it must be reflected in positive law. Following that, fairness in positive law points out the right to equality and non-discrimination. If calling something fair in law, compliance with equality and non-discrimination norms is required, concepts of equality and discrimination require an in-depth review.

1.3 Artificial Intelligence and the Right to Equality and Non-Discrimination

The purpose of explaining different understandings, approaches, and definitions of equality and discrimination in this section is to establish the conceptual framework for the questions to be discussed throughout the thesis from an appropriate position. As explained later in the thesis, although there is awareness of the importance of these concepts in artificial intelligence in the private sector, public authorities, non-governmental organizations, and international organizations, many different understandings exist. The non-binding approach and handling its importance mainly as an ethical issue by these bodies call for establishing the legal foundations and background of the notions. Thus, exploring these notions in the first step is of great significance to bring the matter to the realm of law and to establish a bridge between disciplines about what to understand from these concepts.

---

[315] Binns, *supra* note 261 at 9.
[316] Wachter, Mittelstadt, and Russell, *supra* note 19 at 47.
[317] *Id.* at 44.
[318] *Id.* at 67.

Equality, since the age of enlightenment, has been established as an ideal that stands at the heart of human rights starting from its earliest days.[319] As a pillar of human rights, equality offers a dynamic concept in various dimensions of human rights.[320] Equality and non-discrimination is a paramount principle found in constitutions and international human rights instruments, laying out that each human being is entitled to equal rights and thus, formulated as an essential human right.[321] When taken as a principle, equality allows for comprehending and analyzing human rights concerns, provides a moral incentive for human rights protection mechanisms, and helps to secure other human rights.[322] The right to equality and non-discrimination are substitutable principles, where equality represents the equal treatment of persons, the non-discrimination refers to the difference in treatment.[323] That suggests that the principle bears two obligations, one that is positive to take steps toward reaching equality by recognizing the differences of individuals and one that is negative by not discriminating between individuals.[324]

The equality has an abstract meaning, and deciding on a definition for equality is an intricate task.[325] Individuals differ internally and physically; although perceptible qualities are comparable, each human being is different from one another in some particularities.[326] Establishing measuring points for individuals by comparing their similarities and dissimilarities to measure their 'equalness' and not including comparability as a criterion for measuring equality are two ways to address this issue.[327] There are various separate interpretations of equality since comparison and measuring points do not necessarily resolve equality concerns.[328] Different approaches to equality emanate from the answers given to questions such as what is the aim to be reached with equality and whether equality means treating individuals in the same way or it means

[319] OLIVIER DE SCHUTTER, INTERNATIONAL HUMAN RIGHTS LAW: CASES, MATERIALS, COMMENTARY 561 (2011); Daniel Moeckli, *Equality and Non-Discrimination*, *in* INTERNATIONAL HUMAN RIGHTS LAW , 157 (Daniel Moeckli et al. eds., 2nd edition ed. 2014).
[320] Jarlath Clifford, *Equality*, *in* THE OXFORD HANDBOOK OF INTERNATİONAL HUMAN RİGHTS LAW 420, 430 (Dinah Shelton ed., 2013).
[321] Moeckli, *supra* note 319 at 157, 160.
[322] Clifford, *supra* note 320 at 420–421.
[323] Moeckli, *supra* note 319 at 158.
[324] *Id.*
[325] *Id.* at 157; NICHOLAS SMITH, BASIC EQUALITY AND DISCRIMINATION: RECONCILING THEORY AND LAW 1–11 (2011).
[326] Moeckli, *supra* note 319 at 157.
[327] *Id.* at 157–158.
[328] *Id.*

reaching equal results.[329] More than that, even different political standpoints may influence the understanding of equality one adopts.[330] In distinct contexts, different interpretations of equality may be necessary to address discrimination issues.[331] As a mere expression, equality does not explain what 'equal' corresponds to or who is equal to whom; it is even called an 'empty idea' because of that by some thinkers.[332] Even though the concept of equality is abstract and interpreted differently, legal instruments materialize the definition of the principle and lay out standards to apply for the application of the principle.[333]

The most prominent equality and non-discrimination clauses entail the Universal Declaration of Human Rights (UDHR) articles 1, 2, 7; International Covenant on Civil and Political Rights (ICCPR) articles 2, 3, 26; International Covenant on Economic, Social and Cultural Rights (ICESR) articles 2(2), and 3.[334] There are specialized instruments aimed at reaching equality in specific contexts, for example, International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW), Convention on the Rights of Persons with Disabilities (CRPD) aims at providing non-discrimination in the respective circumstances they regulate.[335] Other human rights instruments that devoted a provision for equality include International Convention on the Protection of the Rights of All Migrant Workers and Members of their

---

[329] *Id.* at 158.
[330] *Id.*
[331] Clifford, *supra* note 320 at 430.
[332] Moeckli, *supra* note 319 at 158; Peter Westen, *The Empty Idea of Equality*, 95 HARV. LAW REV. 537, 544 (1982), https://www.jstor.org/stable/1340593 (last visited Sep 7, 2022); See SMITH, *supra* note 325 at 14–15.
[333] Moeckli, *supra* note 319 at 158.
[334] The UDHR article 1 states: *"All human beings are born free and equal in dignity and rights." Article 26 of the ICCPR declares: "All persons are equal before the law and are entitled without any discrimination to the equal protection of the law. In this respect, the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." Article 2 (2) of the ICESCR states: "The States Parties to the present Covenant undertake to guarantee that the rights enunciated in the present Covenant will be exercised without discrimination of any kind as to race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status."* It should be noted that the ICESCR contains an additional non-discrimination clause particularly about the equality between men and women in its Article 3. *Id.* at 160; UN General Assembly, *Universal Declaration of Human Rights*, U.N. Doc. A/810 217 A (III) (1948); International Covenant on Civil and Political Rights, U.N.T.S 999 171 (1966); International Covenant on Economic, Social and Cultural Rights, U.N.T.S 993 3 (1966).
[335] Moeckli, *supra* note 319 at 160.

Families (CRMW) and the Convention on the Rights of the Child (CRC).[336] Equality and non-discrimination formulations in regional human rights instruments are found in Article 14 and Protocol no. 12 of the European Convention on Human Rights (ECHR) for the CoE; articles 20, 21/1, 23 of the Charter of the Fundamental Rights of the European Union (EU Fundamental Rights Charter) for the EU.[337]

When constructing equality and non-discrimination as a legal problem, general comments provided by the UN human rights mechanisms may help. For example, ICCPR provides a better understanding of the right to equality and non-discrimination. It has a selection of general comments on different aspects of equality and discrimination-related matters. These general comments may help to understand the scope of the equality standards that artificial intelligence decisions should also meet, as the general comments underline essential matters concerning human rights.[338]

In General Comment No. 18 of ICCPR, state obligations to ensure equality are mentioned—encompassing explanations concerning affirmative action, discrimination grounds, and what discrimination means in the legal sense.[339] Furthermore, Article 14 of the ICCPR references equality with a specific view to equality before the courts and concerning the right to a fair trial. Complying with the General Comment No. 32 of ICCPR on Article 14 could help to obtain an essential understanding of the importance of equality with a view to judicial processes and criminal justice processes, which also has great importance with a view to artificial intelligence, as explored in the second chapter.[340] The importance of the equality of arms and access to court in procedural terms in a judicial proceeding is underlined thereunder. [341] That would also be required to be provided regarding a claim of discrimination through artificial intelligence, particularly relevant to the judicial use of artificial intelligence.

---

[336] *Id.* at 160–161.
[337] *Id.* at 161.
[338] See McGregor, Murray, and Ng, *supra* note 1 at 326.
[339] United Nations Human Rights Committee, *General Comment No. 18 - Non Discrimination*, (1989), https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=INT%2FCCPR%2FGEC%2F6622&Lang=en.
[340] United Nations Human Rights Committee, *General Comment No. 32 - Article 14: Right to Equality before Courts and Tribunals and to a Fair Trial*, (2007), https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CCPR%2FC%2FGC%2F32&Lang=en.
[341] *Id.* at paras. 2-3.

Regarding gender equality, General Comment No. 28 of ICCPR may help to provide an essential understanding that should be given with regard to artificial intelligence.[342] For example, it asserts that women's employment is affected by traditions, and states should take measures to eliminate discrimination against women in areas such as education and employment by private actors.[343] This could suggest that this obligation on states would extend to the artificial intelligence field, where due to historical bias toward women cause discrimination—examples of how discrimination in education and employment by different discrimination grounds are discussed in the second chapter.

There are additional general comments and recommendations by the UN human rights mechanisms more explicitly relevant to artificial intelligence related matters.[344] These, too, provide states with better ways to address discrimination through artificial intelligence.

There are two techniques for regulating the right to equality and non-discrimination under human rights instruments: either as a standalone provision that prohibits discrimination with a blanket clause or a dependent/subsidiary clause that may emerge together with other rights regulated in the particular human rights instrument.[345] The right to non-discrimination is limited and can only be applied with a reference to the rights and freedoms regulated under the respective instrument in the second case.[346]

Subordinate non-discrimination clauses are typical in discrimination law, which include Article 2(1) of the UDHR, Article 2(2) of ISECR, and Article 14 of the ECHR.[347]

---

[342] UNITED NATIONS HUMAN RIGHTS COMMITTEE, *General Comment No. 28 - Article 3 (The Equality of Rights between Men and Women)*, (2000), https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CCPR%2FC%2F21%2FRev.1%2FAdd.10&Lang=en.
[343] *Id.* at 31.
[344] See United Nations Committee on Economic, Social and Cultural Rights, *General Comment No. 25 (2020) on Science and Economic, Social and Cultural Rights (Article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights)*, 19 (2020), https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=E%2FC.12%2FGC%2F25&Lang=en; United Nations Committee on the Rights of the Child, *General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment*, 20 (2021), https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CRC%2FC%2FGC%2F25&Lang=en; United Nations Committee on the Elimination of Racial Discrimination, *General Recommendation No.36 . Preventing and Combating Racial Profiling by Law Enforcement Officials*, 13 (2020), https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CERD%2FC%2FGC%2F36&Lang=en.
[345] DE SCHUTTER, *supra* note 319 at 562.
[346] Moeckli, *supra* note 319 at 161.
[347] *Id.* at 162.

Examples of free-standing or autonomous non-discrimination clauses prohibit discrimination entirely and beyond the respective instrument; such clauses include Article 7 of the UDHR, Article 26 of the ICCPR, Articles 2 and 5 of the ICERD.[348]

The ECHR Article 14 states:[349]

> "The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status."

Article 14 of the ECHR is a clause that needs to be applied in conjunction with another right in the Convention text. For this reason, Article 14 of the ECHR is not a robust provision for non-discrimination as it is not a 'free-standing' clause and is cited as a "parasitic" provision.[350] The CoE had adopted Protocol no. 12 to the ECHR to bring an overall prohibition for discrimination as a free-standing clause.[351] Nevertheless, at the time of this study, the total number of ratifications for Protocol no. 12 consists of only twenty member states of the CoE.[352]

In the context of the EU, Article 10 of the Treaty on European Union and the Treaty on the Functioning of the European Union (TFEU) regulates non-discrimination.[353] More specifically, the EU Fundamental Rights Charter has an equality and non-discrimination provision.[354] However, the EU's regulations on discrimination are only valid within the

---

[348] *Id.*
[349] Convention for the Protection of Human Rights and Fundamental Freedoms as amended by Protocols Nos. 11, 14 and 15 supplemented by Protocols Nos. 1, 4, 6, 7, 12, 13 and 16, E.T.S 005 article 14 (1950).
[350] Richard Clayton, Hugh Tomlinson & Carol George, *Freedom From Discrimination in Relation to Convention Rights*, 1 *in* THE LAW OF HUMAN RIGHTS 1204, 1205, 1233 (1 ed. 2000); DAVID JOHN HARRIS ET AL., LAW OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS 578 (2009); MARK W. JANIS, RICHARD S. KAY & ANTHONY WILFRED BRADLEY, EUROPEAN HUMAN RIGHTS LAW: TEXT AND MATERIALS 457 (2008).
[351] JANIS, KAY, AND BRADLEY, *supra* note 350 at 470.
[352] On 4 November 2000, Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms, E.T.S 177 (2000) adding the general prohibition of discrimination to the ECHR was opened for signature. 18 Council member states have signed this additional protocol, but have not subsequently ratified it. 20 states have ratified it. Turkey is among the countries that only signed and then did not ratify. Treaty Office of Council of Europe, *Chart of signatures and ratifications of Treaty 177*, TREATY OFFICE - CONVENTIONS, https://www.coe.int/en/web/conventions/full-list (last visited Aug 14, 2022).
[353] The article 10 of the TFEU declares: *"In defining and implementing its policies and activities, the Union shall aim to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation."* CONSOLIDATED VERSION OF THE TREATY ON THE FUNCTIONING OF THE EUROPEAN UNION ART. 102, 2012 O.J. (C 326) 47 [TFEU].
[354] The Article 21 of the Charter remarks that: *"1. Any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation shall be prohibited.2. Within the scope of application of the Treaties and without prejudice to any of their specific*

EU law, so if the scope of the matter goes beyond the EU law, it is not dealt with but left to national authorities.[355]

1.3.1 Different Positions on Equality

Formal Equality

Formal equality indicates the same or similar treatment of things or individuals that are the same or similar.[356] Is fairness correlated with consistency?[357] If so, consistent treatment will deliver fairness fitting to formal equality.[358] Formal equality entails uniformness in the treatment of everybody.[359] The underlying idea of formal equality, *"treating likes in a like manner,"* originates from Aristotle's *Nicomachean Ethics*.[360] In formal equality, the procedure needs to be what is equal; thus, the result is not as emphasized as the procedure.[361] The idea behind formal equality is that the state should be neutral to everybody and no individual or group is more noteworthy than the other, and what matters is the individual and the merits of the individual, not the individual's membership in a group.[362]

The question of how to determine the alike is a hardship for formal equality.[363] Formal equality gets criticism as treating likes alike does not necessarily propose a key for reaching genuine equality, and acting on finding which different treatment is acceptable is not satisfactory.[364] Consistency requires finding a similar.[365] If treating likes alike is sufficient for equality as long as the treatment is consistent among the likes, then

---

provisions, any discrimination on grounds of nationality shall be prohibited." CHARTER OF FUNDAMENTAL RIGHTS OF THE EUROPEAN UNION, ART. 51(1), 2000 O.J. C 364/01 [CFREU].

[355] European Commission, *Non-Discrimination - What to Do If Your Rights Have Been Breached*, EUROPEAN COMMISSION - EUROPEAN COMMISSION, https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/know-your-rights/equality/non-discrimination_en (last visited Sep 11, 2022).

[356] Clifford, *supra* note 320 at 421.

[357] SANDRA FREDMAN, DISCRIMINATION LAW 8 (2 ed. 2012).

[358] *Id.*

[359] Clifford, *supra* note 320 at 421.

[360] ARISTOTLE, ARISTOTLE: NICOMACHEAN ETHICS 1131a–1131b (Roger Crisp ed. & tran., 2000), https://www.cambridge.org/core/books/aristotle-nicomachean-ethics/C2E5B105977CA6384FF8088CDBA0B90D; Clifford, *supra* note 320 at 422; *See also* SMITH, *supra* note 325 at 24; See Stefan Gosepath, *Equality*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY , Title 2.1 (Edward N. Zalta ed., Summer 2021 ed. 2021), https://plato.stanford.edu/archives/sum2021/entries/equality/ (last visited Aug 14, 2022).

[361] Moeckli, *supra* note 319 at 158.

[362] *Id.* at 159.

[363] FREDMAN, *supra* note 357 at 1.

[364] Moeckli, *supra* note 319 at 159.

[365] FREDMAN, *supra* note 357 at 10.

only the persons coming from a similar or the same background may claim equality, which could mean, for example, that women may merely plead for equality provided that there is a difference in treatment with another woman in her situation but not when the other person is a man.[366] Likewise, there could also be cases where there are no similar persons in a similar situation to compare the difference in treatment with the claimant of discrimination.[367] Formal equality does not consider different characteristics of individuals but aims to treat them based on their merits.[368] Regardless, individuals are not unattached from their features coming from their gender, racial and ethnic origin, religion, or socio-economic background; conversely, such features are ingrained in each individual.[369] Neglecting such characteristics embarks the problem from a viewpoint that there left is a neutral person, named as "universal individual" by Fredman, if a person is separated from their characteristics.[370] Taking such a neutral person as given may lead to norms that could reinforce inequalities.[371]

Formal equality is understood as the most primary perception of equality.[372] It presents the minimum condition for attaining equality, but it is mostly insufficient.[373] Today, formal equality symbolizes the equal treatment of each individual before the law, and equal application of the same laws for everyone.[374] Treating everyone similarly or the same way does not necessarily secure equality or remove the existing inequalities.[375] Limiting equality as such could even become a basis for discrimination as well.[376] Furthermore, formal equality could lead to purposeless results as its mere focus is on the treatment process to be equal.[377] For instance, there would be equal and consistent treatment where the results are undesirable or harmful for everybody; in such a case, formal equality would be achieved even when the result is inefficient for everybody.[378]

---

[366] Moeckli, *supra* note 319 at 159; FREDMAN, *supra* note 357 at 9.
[367] Moeckli, *supra* note 319 at 159; FREDMAN, *supra* note 357 at 9.
[368] FREDMAN, *supra* note 357 at 11.
[369] *Id.*
[370] *Id.*
[371] *Id.*
[372] Clifford, *supra* note 320 at 427.
[373] Moeckli, *supra* note 319 at 164.
[374] Clifford, *supra* note 320 at 427; See United Nations Human Rights Committee, *supra* note 339 at Paras. 1 and 3.
[375] Clifford, *supra* note 320 at 428.
[376] *Id.*
[377] Moeckli, *supra* note 319 at 159.
[378] *Id.*

Another drawback of formal equality is that similar treatment of everybody while fitting for some part of the society, may impact others disparately.[379]

In the artificial intelligence field, formal equality echoes as a fairness understanding measured by statistical similarity metrics that treats similar individuals with common characteristics alike by prompting similar decisions on them.[380]

Equality of Opportunity

Equality of opportunity is the understanding of equality derived from the idea that equal treatment is not enough to achieve equality unless people have the same first chances.[381] Presenting the opportunity to everyone removes the existing impediments for the disadvantaged in society.[382] Thus, equality of opportunity aims to handle inequalities at its birth by making available opportunities the same for everybody.[383] Removing the impediments for disadvantaged individuals and groups, such as upper age limits for job applications which could reduce the chances of women who gave birth to apply to the job, are vital for promoting their equality with the non-disadvantaged.[384] Balancing formal and substantive equality by providing the essential means for individuals to be in the same spot, such as encouraging job applications of the disadvantaged and thus achieving a good standing for the disadvantaged, is the goal of equality of opportunity.[385]

Similar to formal equality, the individual is also crucial under equality of opportunity, based on the belief that the individual and the merits of the individual comes next after everyone has equal opportunities at the beginning.[386] According to this view, the merits of individuals will be decisive once everybody has the same opportunity.[387] Equality of opportunity is not concerned with whether fairness is achieved after each person initially

---

[379] *Id.*
[380] Dwork et al., *supra* note 301 at 1, 18; Saxena et al., *supra* note 292 at 2; *See also* on treating likes alike through algorithms Creel and Hellman, *supra* note 259 at 33.
[381] Moeckli, *supra* note 319 at 159.
[382] *Id.*
[383] FREDMAN, *supra* note 357 at 18; See Catherine Barnard & Bob Hepple, *Substantive Equality*, 59 CAMB. LAW J. 562, 566 (2000), https://www.cambridge.org/core/journals/cambridge-law-journal/article/abs/substantive-equality/1AC13C0FF5C509808284AFDD1F2531A9 (last visited Jun 4, 2023).
[384] Moeckli, *supra* note 319 at 159; Barnard and Hepple, *supra* note 383 at 566; See David A Strauss, *The Illusory Distinction Between Equality of Opportunity and Equality of Result*, 34 WILLIAM MARY LAW REV. 171, 173 (1992), https://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1833&context=wmlr.
[385] Clifford, *supra* note 320 at 428.
[386] Moeckli, *supra* note 319 at 159.
[387] FREDMAN, *supra* note 357 at 18; See for more discussion on merits aspect of equality of opportunity Strauss, *supra* note 384 at 180–185.

has the same opportunity.[388] Proactive efforts should be made to make merits available to everyone for effective equality of opportunity.[389] Adopting equality of opportunity by removing impediments alone will not provide equality in the absence of such measures since merits at present are not gained or distributed equitably.[390]

Equality of opportunity is aimed to be achieved in the artificial intelligence field by preferring an individual with better qualifications when making a decision, pointing to a meritocratic understanding.[391]

Equality of Results

Another equality approach is equality of results. As its name suggests, equality of results looks at equality from the side of consequences by recognizing that even eradicating impediments to equality may not assure true equality and calls attention to social equality.[392] Equality of results refers to social goods such as healthcare, employment, education, and political representation.[393] Under equality of results, not only should the treatment or the opportunities be equal, but the results should be equal.[394] Equality of results also criticizes formal equality because it does not adequately address equality for the disadvantaged and may even lead to unequal results despite consistent treatment.[395] While formal equality aims to reach fairness in the procedure through uniformity, equality of results seeks fairness in the distribution of goods and benefits .[396] Equality of results may work as a remedy for the discriminated individual by removing the discriminatory practice if the result is taken as a negative effect on the individual.[397] Nevertheless, in this case, the equality of results may not present an answer for other individuals to be included in spaces they are refused, meaning that it may not convey true equality.[398] If the equality of results is taken as effects on groups, it could work to

[388] FREDMAN, *supra* note 357 at 18.
[389] *Id.* at 19.
[390] *Id.*; Strauss, *supra* note 384 at 174.
[391] Matthew Joseph et al., *Fair Algorithms for Infinite and Contextual Bandits*, 1 (2017), http://arxiv.org/abs/1610.09559 (last visited Oct 6, 2022); Saxena et al., *supra* note 292 at 2.
[392] Moeckli, *supra* note 319 at 159; Strauss, *supra* note 384 at Cf for a discussion on equality of opportunity and equality of result.
[393] Moeckli, *supra* note 319 at 159.
[394] FREDMAN, *supra* note 357 at 14; Cf SMITH, *supra* note 325 at 116.
[395] FREDMAN, *supra* note 357 at 14; Moeckli, *supra* note 319 at 159.
[396] FREDMAN, *supra* note 357 at 14.
[397] *Id.*
[398] *Id.*

demonstrate discrimination of certain groups of benefits if the results are not reflective of equal distribution of groups in a place.[399]

A different illustration of equality of results is equality of outcomes, which pursues the actualization of equality of results.[400] Equality of outcomes aims to undertake real action and efforts to remove inequalities settled in the society, which often damage historically underprivileged groups.[401] An example of equality of outcomes is candidate quotas for female candidates in job applications.[402]

Substantive Equality

Substantive equality is the understanding of equality that suggests the rationale of inequalities should be taken into account and the disadvantaged persons or groups treated accordingly.[403] Substantive equality coincides with equality as a right.[404] Some regard equality of opportunity and equality of results as different outlooks of substantive equality.[405] Substantive equality aspires to amend the deficits of different equality understandings by proposing an answer through redistribution, recognition, transformation, and inclusion.[406] In essence, substantive equality brings together the concepts of dignity,[407] distributive justice, and equal participation.[408]

Redistribution signifies the cause of discrimination that substantive equality endeavors to end the persistent discriminatory practices and bias toward that cause.[409] Recognition means respecting an individual for being a person because of human dignity.[410] Transformation denotes weaving transformative equality as part of substantive equality, which could be achieved by recognizing individuals' dissimilarities and changing the detrimental effects of the status quo on discriminated individuals.[411] It

[399] *Id.* at 15.
[400] *Id.*
[401] Clifford, *supra* note 320 at 429.
[402] *Id.*
[403] *Id.* at 421.
[404] Moeckli, *supra* note 319 at 158.
[405] *Id.* at 159.
[406] FREDMAN, *supra* note 357 at 33; Barnard and Hepple, *supra* note 383 at 567; *See also* for a critique SMİTH, *supra* note 325 at 117–119.
[407] Human dignity is one of the cornerstones of human rights and constitutes a foundation and rationale for equality, and may even be considered a notion that conveys equality. See FREDMAN, *supra* note 357 at 19–25 and 227–230; *See also* SMITH, *supra* note 325 at 121–131.
[408] Moeckli, *supra* note 319 at 159.
[409] FREDMAN, *supra* note 357 at 26.
[410] *Id.* at 28.
[411] *Id.* at 30.

follows the idea of transforming the social structures in society to ameliorate the standing of the disadvantaged in society.[412] Transformative equality directs toward inclusivity and shifting organizational and institutional structures to bring these within reach of disadvantaged groups.[413] Even though it resembles equality of outcomes, they differ in their methods of facilitating equality. Equality of outcomes employs providing benefits to attain equal results, whereas transformative equality targets the discriminatory roots in the structures.[414] Recently, the idea of taking proactive measures to ensure equality has become more favored.[415] Fostering the participation of discriminated persons in social life and society makes inclusion real.[416] If reasonable, accommodations or adjustments are applied in non-discrimination law that could help to achieve substantive equality by retaliating impediments, transforming the conditions for the disadvantaged, and carrying out measures for an actual change.[417] The third chapter mentions the approaches for achieving substantial equality in artificial intelligence.

1.3.2 Types of Discrimination

Firstly, a definition for discrimination should be made. Different treatment of two individuals does not amount to discrimination *per se*.[418] Different treatments of individuals may be acceptable when there are grounds allowing the difference in treatment, such as morally acceptable reasons.[419] If the different treatment is unjustifiable, there is discrimination. Direct discrimination occurs when a person is subjected to different treatment on prohibited grounds and treated less favorably than another person in a comparable situation.[420] The effect of the treatment, which could be neutral from the outside but discriminatory in effect, is decisive for indirect discrimination instead of the aim or way of the treatment.[421]

---

[412] Clifford, *supra* note 320 at 430.
[413] *Id.*
[414] *Id.* at 429–430.
[415] Moeckli, *supra* note 319 at 158.
[416] FREDMAN, *supra* note 357 at 32.
[417] *Id.* at 215–217.
[418] Moeckli, *supra* note 319 at 159.
[419] *Id.*
[420] *Id.* at 164; JANIS, KAY, AND BRADLEY, *supra* note 350 at 485.
[421] JANIS, KAY, AND BRADLEY, *supra* note 350 at 486.

Understanding the discrimination test and types of discrimination will help to tackle the discrimination through artificial intelligence.[422] During a judicial process about a discriminatory claim, adhering to one type of equality is not enough to address the topic, which is usually complex and challenging to settle.[423] Legal regulations reflect different understandings of equality and adapt into more than a single understanding to suit the needs of the claim.[424] The ambit of discrimination law is categorized under either characteristics or grounds on which the discrimination is based. These grounds may vary between legal documents and jurisdictions. The most noted discrimination grounds consist of sex, race, gender, sexual orientation, age, disability, and religion.[425] The first criterion for discrimination is a difference in treatment between individuals.[426] According to ECtHR, discrimination claims should fall within the ambit of the rights regulated in the ECHR.[427]

The right to non-discrimination includes the obligation to prevent discrimination by non-state actors and the obligation to respect as not to discriminate.[428] States have a duty to enact non-discrimination legislation.[429] Discrimination emanates from societal patterns, ingrained disadvantages, and exclusions; accordingly, it is necessary to address such problems to change the status quo.[430] There could be obligations for promoting, guaranteeing, and securing equality through positive measures.[431] That is named either as taking positive action or positive obligations in the context of human rights law.[432] Positive actions may include special protection measures that aim for substantive equality.[433]

[422] McGregor, Murray, and Ng, *supra* note 1 at 326.
[423] Clifford, *supra* note 320 at 426.
[424] *Id.* at 427.
[425] FREDMAN, *supra* note 357 at 109.
[426] Clifford, *supra* note 320 at 438; F.H. Zwaan-de Vries v. Netherlands, UN Doc CCPRC29D1821984 ¶ 13-14 (1987); Carson and Others v. the United Kingdom, 2010–II Eur. Court Hum. Rights Rep. Judgm. Decis. 407, ¶ 63 (2010).
[427] Moeckli, *supra* note 319 at 162; Rasmussen v. Denmark, 17 ECHR 29 (1984).
[428] Moeckli, *supra* note 319 at 170.
[429] *Id.*
[430] *Id.*
[431] *Id.*
[432] *Id.* at 170–171.
[433] *Id.* at 171.

Direct Discrimination

Direct discrimination happens when persons in an analogous or relevantly similar situation are treated differently based on an identifiable characteristic or status.[434] Instances contrary to equal treatment fell under the scope of direct discrimination.[435] The discrimination is based on a protected ground.[436]

As direct discrimination needs a comparable person, it originates from the formal equality approach, which refers to similar persons in similar situations.[437] Where there is a lack of comparable persons in a comparable situation, available proof for the discrimination is restricted.[438] Where direct discrimination is apparent, proving the discrimination may not be as straightforward.[439] The discriminating party may claim the difference in treatment is not due to a prohibited ground because, first, finding a comparable person may not always be easy, and second, demonstrating the existence of direct discrimination could be challenging as direct discrimination can be disguised.[440] The comparison could be made through a hypothetical comparator as there may not be real persons to compare.[441] Direct discrimination gets criticism because of the limitation that is the requirement of a comparator.[442] Replacing the 'less favorable treatment with 'unfavorable' treatment or 'detrimental' effects on a person may improve the scope of direct discrimination.[443] Taking proactive measures to improve the situation of a protected ground may convey another restriction if it is taken as discrimination against the previously more favored person or group, thus resulting in problems for closing the gaps between the less and more favored.[444] Direct discrimination demands similarity or being in similar standing with the favorable individual or group.[445] Direct discrimination is not

[434] D.H. and Others v. The Czech Republic, 2007–IV Eur. Court Hum. Rights Rep. Judgm. Decis. 241, ¶ 175 (2007); Burden v. The United Kingdom, Rep. Judgm. Decis. 2008 ¶ 60 (2008); Carson and Others v. the United Kingdom, *supra* note 426 at ¶ 61; Biao v. Denmark, Rep. Judgm. Decis. 2016 89 (2016).
[435] Sandra Fredman, *Direct and Indirect Discrimination: Is There Still a Divide?*, *in* FOUNDATIONS OF INDIRECT DISCRIMINATION LAW 31, 31 (Hugh Collins & Tarunabh Khaitan eds., 1 ed. 2018).
[436] FREDMAN, *supra* note 357 at 166.
[437] *Id.*; Moeckli, *supra* note 319 at 160.
[438] FREDMAN, *supra* note 357 at 166–167.
[439] Moeckli, *supra* note 319 at 165.
[440] *Id.*
[441] FREDMAN, *supra* note 357 at 168.
[442] *Id.*
[443] *Id.* at 169.
[444] *Id.* at 175.
[445] JANIS, KAY, AND BRADLEY, *supra* note 350 at 485–486.

always enough to address inequality overall because there needs to be discriminatory grounds and less favorable treatment than a more favored group.[446]

According to human rights case law, a different treatment must be based on an "identifiable characteristic" or "status" in order for discrimination to exist.[447] Furthermore, the settings must be "analogous" or "relevantly similar" for individuals to assert that it is discriminatory.[448] As it would necessitate making a decision based on inputs that openly connects to being in an identifiable status and awarded lower training scores in supervised learning, direct discrimination is rarely observed in the setting of artificial intelligence.[449] Proxy discrimination that results from correlations based on apparently neutral criteria, which could either be intentionally concealed or unintentionally, falls outside the scope of direct discrimination.[450]

Indirect Discrimination

A seemingly objective and natural practice with no difference in treatment may be discriminatory in case of indirect discrimination.[451] This type of discrimination is called indirect discrimination or disparate impact, in which systemic biases are the leading cause.[452] Indirect discrimination as a mechanism helps to reveal concealed intentional discrimination, contest to structural discriminations in a society and compensate for unfair distribution of social advantages.[453]

The extent of indirect discrimination is disputed.[454] In indirect discrimination, equal treatment brings about a discriminatory or unequal result that could be justified, provided that there is a sound justification.[455] The unequal impact refers to groups more than

---

[446] *Id.*
[447] Biao v. Denmark, *supra* note 434 at 89; EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS., EUROPEAN COURT OF HUMAN RIGHTS., & COUNCIL OF EUROPE (STRASBOURG)., HANDBOOK ON EUROPEAN NON-DISCRIMINATION LAW: 2018 EDITION 43 (2018), https://data.europa.eu/doi/10.2811/58933 (last visited Sep 11, 2022).
[448] Biao v. Denmark, *supra* note 434 at ¶89; Carson and Others v. the United Kingdom, *supra* note 426 at ¶61; D.H. and Others v. The Czech Republic, *supra* note 434 at ¶175; Burden v. The United Kingdom, *supra* note 434 at ¶60; EUROPEAN UNİON AGENCY FOR FUNDAMENTAL RİGHTS., EUROPEAN COURT OF HUMAN RİGHTS., AND COUNCİL OF EUROPE (STRASBOURG)., *supra* note 447 at 43.
[449] Hacker, *supra* note 19 at 1151.
[450] *Id.*
[451] Moeckli, *supra* note 319 at 165.
[452] *Id.*
[453] Hugh Collins & Tarunabh Khaitan, *Indirect Discrimination Law: Controversies and Critical Questions*, *in* FOUNDATIONS OF INDIRECT DISCRIMINATION LAW 1, 29 (Hugh Collins & Tarunabh Khaitan eds., 1 ed. 2018).
[454] Clayton, Tomlinson, and George, *supra* note 350 at 1206.
[455] FREDMAN, *supra* note 357 at 178.

individuals in the context of indirect discrimination.[456] Indirect discrimination is often associated with equality of results.[457] Indirect discrimination also coincides with the understanding of substantive equality.[458] Nevertheless, if there is justification for the unequal effects, then unequal impacts will not violate indirect discrimination.[459] In addition, since the extent of protected grounds is ambiguous, the dimensions of equality of results are challenging to resolve.[460]

Indirect discrimination happens when a general measure, policy, neutral rule, general attitude of authorities, *de facto* situation, or well-established practice that is seemingly natural disproportionately results in prejudicial discriminatory effects on a particular group.[461] The policy or the measure may not be targeting a particular group, but it indirectly affects the individuals from that group.[462] It should be noted that ECtHR receives criticism with regard to the judges' skepticism for indirect discrimination cases.[463]

In indirect discrimination, the focus is not on a group but on different effects.[464] That means that even if a particular group is not directly aimed, discrimination may occur if it disproportionately affects the group.[465] Statistical evidence may help compare the "protected groups" and the affected individuals to understand if they match.[466] If, for example, statistical evidence indicate that majority of persons affected by a measure are

---

[456] *Id.* at 179.
[457] *Id.* at 180; Fredman, *supra* note 435 at 31.
[458] Moeckli, *supra* note 319 at 160.
[459] FREDMAN, *supra* note 357 at 180.
[460] *Id.*
[461] Hugh Jordan v. The United Kingdom, Eur. Court Hum. Rights Rep. Judgm. Decis. 154 (2001); Hoogendijk v. The Netherlands, Eur. Court Hum. Rights Rep. Judgm. Decis. ¶ 2 of the "Law" part (2005); D.H. and Others v. The Czech Republic, *supra* note 434 at 184; Biao v. Denmark, *supra* note 434 at ¶ 103; Zarb Adami v. Malta, Eur. Court Hum. Rights Rep. Judgm. Decis. ¶¶ 75-76 (2006); Opuz v. Turkey, Rep. Judgm. Decis. 2009 ¶ 192 (2009); Tapayeva and Others v. Russia, ¶ 112 (2021); See for EU non-discrimination law perspective Wachter, Mittelstadt, and Russell, *supra* note 19 at 15.
[462] Hugh Jordan v. The United Kingdom, *supra* note 461 at ¶ 154; Hoogendijk v. The Netherlands, *supra* note 461 at ¶ 2 of the "Law" part.
[463] See for a detailed socio-legal analysis of the ECtHR's approach to indirect discrimination Barbara Havelková, *Judicial Scepticism of Discrimination at the ECtHR*, *in* FOUNDATIONS OF INDIRECT DISCRIMINATION LAW 83 (Hugh Collins & Tarunabh Khaitan eds., 1 ed. 2018), http://www.bloomsburycollections.com/book/foundations-of-indirect-discrimination-law/ch4-judicial-scepticism-of-discrimination-at-the-ecthr/ (last visited Feb 26, 2018).
[464] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS., EUROPEAN COURT OF HUMAN RIGHTS., AND COUNCIL OF EUROPE (STRASBOURG)., *supra* note 447 at 56.
[465] Hugh Jordan v. The United Kingdom, *supra* note 461 at ¶154.
[466] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS., EUROPEAN COURT OF HUMAN RIGHTS., AND COUNCIL OF EUROPE (STRASBOURG)., *supra* note 447 at 57.

women, there could be indirect discrimination.[467] Statistical evidence may reveal a *prima facie* indirect discrimination.[468] If the applicant is able to present such evidence to the court, then proving that the different effects stem from "objective factors," rather than discrimination grounds shifts to the state.[469] More explanations regarding the evidentiary matters in discrimination are discussed below and again in the second chapter.

If a type of bias enters the learning process, that could amount to indirect discrimination if there is no explicit bias.[470] Proxy discrimination may also be addressed by indirect discrimination if there is a correlation between being a protected group member and the outcome.[471] Indirect discrimination constitutes the majority of discrimination through artificial intelligence cases.[472]

Discrimination Grounds

Discrimination grounds have an evolving nature and scope that broadens through time.[473] For example, age, disability, gender identity, and sexual orientation are some recent discrimination grounds that have become established.[474] Sometimes, these grounds may be more than one at a time. Intersectional discrimination signifies the accumulation of various features of a person's identity as the basis of discrimination.[475] Intersectional discrimination is called for attention in General Comment No. 28 of ICCPR, particularly regarding discrimination against women.[476] Intersectional discrimination is widespread and should be given careful consideration concerning discrimination through artificial intelligence, as explored with examples in the second chapter.[477]

Similar to how equality and discriminations differ, the grounds for discrimination differ in legal instruments.[478] The first type of legal provisions that sorts discrimination grounds is the clauses that itemize a limited number of discrimination grounds explicitly; the second type is clauses that do not refer to any grounds but rather adopt general norms

[467] *Id.* at 57; Di Trizio v. Switzerland, ¶¶66, 86 (2016); D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 180.
[468] Hoogendijk v. The Netherlands, *supra* note 461 at ¶2 under the "Law" section.
[469] *Id.*
[470] Hacker, *supra* note 19 at 1154.
[471] *Id.*
[472] *Id.* at 1160.
[473] Moeckli, *supra* note 319 at 163.
[474] *Id.*
[475] *Id.*
[476] UNITED NATIONS HUMAN RIGHTS COMMITTEE, *supra* note 342 at 31.
[477] Hoffmann, *supra* note 203 at 906; See Xiang and Raji, *supra* note 32 at 3–4; *See also* Wachter, Mittelstadt, and Russell, *supra* note 19 at 20–21.
[478] Moeckli, *supra* note 319 at 163.

of equality; the third type of clauses is that mentions an open-ended list for discrimination grounds.[479] Article 14 of the ECHR presents an open-ended list -that is not exhaustive but illustrative.[480] The grounds of article 14 go beyond the list of grounds consisting of "*sex, race, color, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status,"* and includes marital and professional status, sexual orientation and disability according to the ECtHR's variety of case law under the other status.[481] Status denotes *"a characteristic by which persons or groups of persons are distinguishable from each other"*.[482]

It is argued in the artificial intelligence field that protection grounds may be used in the training process.[483] Some authors argue that removing protection grounds from the training data may not remove the chances of indirect discrimination.[484] Sometimes, discrimination grounds may correlate to non-discriminatory attributes, indicating a "redundant encoding".[485] Discrimination through artificial intelligence may go beyond the protected groups in the framework of the right to equality and non-discrimination.[486]

Demonstrating the Discrimination

As in other legal matters, substantiating or demonstrating is essential to discrimination claims. The burden of proof for discrimination is on the person who pleads exposure to discrimination.[487] The claimant person should demonstrate discrimination grounds, the comparator person or group, and the treatment or result that is 'different'.[488] Once the claimant proves these criteria, the burden of proof shifts to the alleged discriminating party, who is the relevant state in the international human rights law context.[489]

Demonstrating these proof standards could be challenging for the claimant, particularly for indirect discrimination, where a supposedly neutral treatment results in a

[479] *Id.*
[480] JANIS, KAY, AND BRADLEY, *supra* note 350 at 470–471; Engel and Others v. The Netherlands, 1 Eur. Court Hum. Rights Rep. Judgm. Decis. 647, 72 (1976).
[481] JANIS, KAY, AND BRADLEY, *supra* note 350 at 471.
[482] Kjeldsen, Busk Madsen and Pedersen v. Denmark, 1 Eur. Court Hum. Rights Rep. Judgm. Decis. 711, 56 (1976); JANIS, KAY, AND BRADLEY, *supra* note 350 at 471.
[483] Binns, *supra* note 261 at 9.
[484] Veale and Binns, *supra* note 140 at 4.
[485] *Id.* at 4.
[486] Wachter, Mittelstadt, and Russell, *supra* note 19 at 11.
[487] Moeckli, *supra* note 319 at 169.
[488] *Id.*
[489] *Id.*

disproportionate or disparate impact on particular groups.[490] As mentioned earlier, in cases like this, ECtHR has recourse to reliable and significant statistics to observe proof of indirect discrimination concerning the impact of the result on the group.[491] Comparison takes the most time for the courts to inspect the discrimination claim, and the lack of a comparator may culminate in the failure of the claim.[492]

The applicants should present a less favorable treatment in comparison to building a sufficiently substantiated discrimination claim.[493] International bodies such as the EU and the ECtHR adopt the proportionality test when examining discrimination claims.[494] In *D.H. and Others v. the Czech Republic*, the ECtHR had constructed a test for examining indirect discrimination.[495] Ways of how difference in effect and difference in treatment may stem from artificial intelligence should be examined.[496] Therefore, these definitions and tests will be examined and applied to artificial intelligence in the second chapter.

Discrimination is already challenging, even if it does not relate to artificial intelligence, whether it results from a hidden intentional or unintentional bias.[497] Typical methods for proving discrimination may be challenged by artificial intelligence.[498] Finding a comparator and arguing that they should be considered in a similar situation renders automation of fairness difficult.[499] Authors also argue that whether a "concrete comparator" or a "hypothetical" one is required to prove discrimination also makes fairness automation difficult.[500] However, a concrete comparator is not required in human rights law, which could provide an answer. Discrimination through artificial intelligence may be mostly unintentional, stemming from social, historical or technical biases, which present difficulties in attributing accountability and traceability.[501]

---

[490] *Id.*
[491] *Id.* at 169; D.H. and Others v. The Czech Republic, *supra* note 434 at 188; FREDMAN, *supra* note 357 at 180.
[492] Denise Réaume, *Dignity, Equality, and Comparison*, *in* PHILOSOPHICAL FOUNDATIONS OF DISCRIMINATION LAW 0, 7 (Deborah Hellman & Sophia Moreau eds., 2013), https://doi.org/10.1093/acprof:oso/9780199664313.003.0002 (last visited Sep 8, 2022).
[493] JANIS, KAY, AND BRADLEY, *supra* note 350 at 474, 476.
[494] FREDMAN, *supra* note 357 at 180.
[495] *Id.*; See D.H. and Others v. The Czech Republic, *supra* note 434.
[496] See Xiang and Raji, *supra* note 32 at 3.
[497] Hacker, *supra* note 19 at 1168.
[498] Wachter, Mittelstadt, and Russell, *supra* note 19 at 64.
[499] *Id.* at 22.
[500] *Id.* at 37.
[501] Hoffmann, *supra* note 203 at 904–905 See Chapter 3 of this study for discussions on accountability and traceability.

Discriminatory Intention

Within the scope of the ECtHR, the discriminatory result is sufficient in an instance of indirect discrimination for discrimination to be forbidden; discriminatory intent is not required for discrimination, and whether there is an intention or lack of it is not consequential in the international human rights system.[502] On that note, the discriminatory intent requirement may vary between jurisdictions.[503] As the lack of discriminatory intent may not necessarily mean a lack of discrimination, it does not always mean that all direct discrimination occurs due to the intention for discrimination, even though direct discrimination usually emerges from discriminatory intent.[504]

Justifying the Discrimination

When dealing with discrimination claims, courts examine the justification of discrimination.[505] In the case law of the ECtHR, discrimination may surface even when a general policy or measure does not specifically target a group to discriminate if there are *"disproportionately prejudicial effects on a particular group."*[506] The justification must be "objective and reasonable"; the ECtHR ascribes that justification is not possible if there is no "legitimate aim" or "reasonable relationship of proportionality between the means employed and the aim sought to be realized."[507] Similarly, a difference in treatment that fulfills the objectivity and reasonableness criteria is acceptable and does not amount to discrimination under ICCPR.[508]. Yet, the ECtHR presents the most precise method for examining justification, which the second chapter will review in detail.[509]

Notably, if the difference in treatment is due to race, ethnicity, religion, nationality, sex, or disability, there is a requirement for "strict scrutiny" when examining a discrimination claim because different treatment in such cases usually tend to result in

[502] D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 184; Moeckli, *supra* note 319 at 166; Zuiderveen Borgesius, *supra* note 25 at 1577; Biao v. Denmark, *supra* note 434 at ¶103.
[503] Demonstrating intention for discrimination is required in the USA. Moeckli, *supra* note 319 at 166.
[504] *Id.* at 166.
[505] *Id.* at 164.
[506] D.H. and Others v. The Czech Republic, *supra* note 434 at ¶¶ 175, 184, 185; Zuiderveen Borgesius, *supra* note 25 at 1577.
[507] D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 196; Zuiderveen Borgesius, *supra* note 25 at 1577–1578; Case "Relating to Certain Aspects of the Laws on the Use of Languages in Education in Belgium" v. Belgium "Belgian Linguistics Case," 1 EHRR Eur. Court Hum. Rights Rep. Judgm. Decis. 252, ¶ 10 (1968); Clayton, Tomlinson, and George, *supra* note 350 at 1242; See for a comparison with the EU discrimination test Hacker, *supra* note 19 at 1160–1161; *See also* Smuha, *supra* note 19 at 97.
[508] F.H. Zwaan-de Vries v. Netherlands, *supra* note 426 at ¶ 13; Clifford, *supra* note 320 at 438.
[509] Moeckli, *supra* note 319 at 167.

discriminatory outcomes.[510] For the rest of the discrimination grounds, there is no uniform criterion for the justification examination.[511]

1.3.3 Connecting Artificial Intelligence and the Right to Equality and Non-Discrimination

Discrimination comes forward as one of the main risks of artificial intelligence to human rights.[512] The right to equality and non-discrimination law has the potential to deliver a solution, at least partially, for the discriminatory concerns stemming from artificial intelligence.[513] The first step is to inquire about different approaches to equality to search for a suited equality approach to artificial intelligence. Which type of equality should be adopted concerning artificial intelligence? Equality could be aimed to be achieved at the procedural level, which reflects the status quo. If the goal only focuses on procedural equality, there could be formal equality, but what will happen to the outcome? Indeed, procedural equality in artificial intelligence has great importance because of the importance of data and data collection processes. Nevertheless, as said, outcomes can be discriminatory even when the procedure is equal. Usually, the procedures can be seemingly equal, yet the outcomes may be flawed. All such concerns are valid for equality and discrimination in artificial intelligence.

Likewise, the risk of only being able to claim discrimination against a comparator for individuals in different situations applies to artificial intelligence too. Definitely, being in a comparable situation with comparable persons are significant for artificial intelligence because as machines feed from data, having more and better training data will help achieve formal equality regarding artificial intelligence due to being able to compare the similarities better. To find a similar person in their situation, should the person claiming discrimination resort to the training data to find any comparable person in their situation? That is quite unlikely and most cases, impossible. Equality of outcomes may also not be sufficient because of the fast pace of artificial intelligence and the need to catch up and get ahead of different sources of bias. Therefore, adopting a substantive approach to equality in artificial intelligence is the reasonable approach when dealing

[510] Clifford, *supra* note 320 at 438; Moeckli, *supra* note 319 at 168.
[511] Moeckli, *supra* note 319 at 169.
[512] Muller and The CAHAI Secretariat, *supra* note 90 at 28.
[513] Zuiderveen Borgesius, *supra* note 25 at 1576.

with equality and discrimination in artificial intelligence. More exploration in these matters are made under the third chapter, when examining the principles for artificial intelligence.

Understanding the development background of equality and non-discrimination is helpful for artificial intelligence too, because artificial intelligence will also reflect the human related issues of equality and non-discrimination. For example, the basis for discrimination concerns may differ geographically. Equality discussions had initiated around racial grounds in the US, whereas in Europe, nationality and gender were the reason for equality discussions to begin.[514] That could convey that people may experience discrimination on different grounds depending on the location of their use or exposure to an artificial intelligence system. It could also suggest that geographical discrimination basis may be disseminated to another area if an artificially intelligent system is used worldwide or in different geographical areas. Concerns such as these should be considered when deploying an artificial intelligence system.

To address equality and discrimination in artificial intelligence, examining how artificial intelligence may lead to discrimination is another step. Both direct and indirect discrimination may occur with artificial intelligence. For instance, direct discrimination may surface if an artificial intelligence system discriminates because of having an attribute that is a protected ground when conducting a task. In such a case, the AI-system could mark an attribute such as ethnic background and discriminate directly and/or intentionally on that ground. However, while this is possible, that would not occur often; the likelihood of indirect and/or unintentional discrimination is higher than direct discrimination.[515] Indirect discrimination may arise, for example, when people from a specific ethnic background pay more for a good or service due to a decision made by or through an AI-system.[516]

While indeed, upsides of artificial intelligence exist in various ways for different things, even for equality, those are frequently mentioned and advocated, yet, the risks coming with it, even though now recognized, still require more mention than the

[514] FREDMAN, *supra* note 357 at 110.
[515] ZUIDERVEEN BORGESIUS, *supra* note 19 at 34.
[516] *Id.*

opportunities.[517] Discrimination through artificial intelligence, while resulting in different ways of individual violations, also has dangers to society for worsening the equality in the society, increasing the existing inequalities.[518] Discrimination through artificial intelligence may arise at the design and through bias.[519] The types of biases mentioned above may affect the outcomes of the artificial intelligence systems and decisions made through them, which may result in discrimination.[520] The lack of fairness in artificial intelligence connects to the existing norms of equality and non-discrimination in international human rights law mentioned earlier.[521] Considering that artificial intelligence does not stop at borders, the effects of artificial intelligence on human rights should be evaluated within the framework of human rights law. [522] All of these factors demonstrate the relationship between artificial intelligence and right to equality and non-discrimination and constructs the issue as a legal problem. The CoE Report of the Committee on Equality and Non-Discrimination on Preventing Discrimination Caused by the Use of Artificial Intelligence summarizes the essential worries regarding non-discrimination. The first concern is the deficiency in public debate, democratic oversight, examination before utilizing artificial intelligence tools, and the absence of knowledge regarding artificial intelligence.[523] The second concern is considering artificial intelligence fairer and more precise than humans, originating from automation bias, not considering individual contexts.[524] The third concern is targeting marginalized persons and persons from lower economic backgrounds, increasing the possibility of state surveillance.[525] Addressing the dangers for equality and non-discrimination in artificial

[517] See for the opportunities brought by artificial intelligence AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *Ad Hoc Committee on Artificial Intelligence (CAHAI) Feasibility Study*, 56 6 (2020), https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da (last visited Aug 4, 2022).
[518] CHRISTOPHE LACROIX, *supra* note 13 at 9; Barocas and Selbst, *supra* note 213 at 674; KAREN YEUNG, *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework*, 96 31–32 (2019), https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168096bdab.
[519] EUROPEAN COMMISSION, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, 26 11 (2020), https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (last visited Aug 6, 2022).
[520] YEUNG, *supra* note 518 at 29–30; See for more information on the relationship between bias and discrimination in artificial intelligence AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517; and WAGNER, *supra* note 19 at 26–28.
[521] YEUNG, *supra* note 518 at 29.
[522] CHRISTOPHE LACROIX, *supra* note 13 at 6.
[523] *Id.* at 10.
[524] *Id.*
[525] *Id.*

intelligence call for the understanding and application of the aforementioned standards of equality and discrimination. Thus, an in-depth examination of different usage areas of artificial intelligence is required to tackle the questions involving equality and non-discrimination more concretely, which the second chapter will do. The third chapter will then connect the legal normative features of equality and non-discrimination to artificial intelligence and will delve into the application of the right to equality and non-discrimination.

Chapter 2:
ANALYZING DISCRIMINATION THROUGH ARTIFICIAL INTELLIGENCE IN LIGHT OF HUMAN RIGHTS LAW

This chapter first aims to display how discrimination may come into life through real-life examples to help the reader understand how artificial intelligence already generates discrimination and demonstrate the imminence of the issue. Secondly, it aims to connect the current corpus of the right to equality and non-discrimination under international human rights law with artificial intelligence.[526] The chapter will probe discrimination through artificial intelligence to achieve the aims of the chapter under the discrimination grounds in international human rights law with examples from different areas of use, including judicial procedures, criminal justice, healthcare and medicine, business life, banking, insurance, targeted advertisement, automated content management. Questions such as in which forms and with which type of bias explained in the previous chapter, in what situations and how, and against whom the discrimination stands out will be discussed within the framework of the right to equality and non-discrimination in international human rights law. While these are challenging to distinguish from each other all the time, exploring them under certain headlines allows to display where these are vulnerable and to discover the areas of use. As explained in the previous chapter, the legal field needs to increase awareness about artificial intelligence. At the time of this writing, misconceptions about artificial intelligence and its real-life applications continue while it is in fact, consistently put into action. Another aim for that is to contribute in the non-discrimination law as it has an ever-developing nature. The field has expanded enormously even in recent years; discriminatory instances that were not considered before are now viewed as new subfields under discrimination law, and, inevitably, artificial intelligence will undoubtedly impact the discrimination law. It should be noted that human rights law does not provide a fully-developed AI governance regime, but provides an essential framework.[527]

[526] *See* McGregor, Murray, and Ng, *supra* note 1; Cf for criticisms towards adopting human rights law in AI-related matters. Smuha, *supra* note 19.
[527] McGregor, Murray, and Ng, *supra* note 1 at 313; Smuha, *supra* note 19 at 100.

This chapter's headlining follows the framework of discrimination grounds specified under the right to non-discrimination in international human rights law. Nearly all international human rights texts presented in the first chapter follow the same framework of discrimination grounds.

The examples of discrimination discussed in this chapter come from various jurisdictions. Of course, discrimination cases may take place differently in different states due to specific circumstances, factors, or regulations unique to them, which would influence possible applications to human rights mechanisms. Likewise, the examples in the chapter may transpire similarly across different states. For this reason, the concrete examples provided throughout the chapter should be read to support understanding the abstract notion of discrimination in the context of artificial intelligence.

There has not been any decision concerning artificial intelligence based discrimination to the international human rights mechanisms. Since the discriminatory consequences of artificial intelligence is relatively a new field that is developing quite rapidly at the time of this study, it is yet to be possible to examine these by analyzing applications through case studies under a chosen human rights control mechanism. In the first part of this chapter, various artificial intelligence areas of use will be reviewed from the aspects of why and how it is employed in the context of discrimination. Illustrations from academic research and news about discrimination through artificial intelligence will be explained under the discrimination grounds. The second part of the chapter will explore discrimination through artificial intelligence under the discrimination mechanisms of present-day human rights law.

The areas of use and real-life examples under the discrimination grounds in the first part are not limited to a specific national jurisdiction or an international human rights mechanism. The examples throughout the chapter come from different geographical jurisdictions, the underlying reason being the emerging and expanding nature of the study subject and the lack of established case law and regulatory standards. For this reason, any available examples for the discrimination grounds are explored because the instances of discrimination in one jurisdiction may as well be similarly seen in another. The responses from the jurisdictions may vary, yet they can be valuable examples either in a good or bad way. Nevertheless, there is a reason why examples from certain states are chosen and mentioned for specific reasons. Illustrations from the United States (U.S.), the EU, and

the CoE are referenced throughout the chapter, as most discussions and academic research around artificial intelligence emerge from these regions. Türkiye is featured throughout the examples presented in this study because the author of the thesis and the university to which it is submitted is of Turkish origin. In addition, the Turkish perspective in the discourse of artificial intelligence could often be overlooked. Nevertheless, given Türkiye's status as a developing economy, illustrations of artificial intelligence applications may contribute to artificial intelligence discussions.

The ECHR and the ECtHR draw the borders of the examination under the second part of this chapter. The reasons why the ECHR is preferred amongst the human rights mechanisms include the status of the state where this thesis is written as a state party to the ECHR and of the Convention as a pioneer of human rights not only for the CoE but also the rest of the world, as well as the leadership potential of the CoE in terms of artificial intelligence, which would reinforce the Convention in this context too, and ultimately the need for confining the thesis to a certain length.

2.1 Discrimination Through Artificial Intelligence by Discrimination Grounds

2.1.1 Artificial Intelligence by the Areas of Use

It is possible to categorize different artificial intelligence areas of use, such as health, education, business life, criminal justice, judiciary, advertising and marketing, online platforms, banking, and insurance.[528] In these different areas, artificial intelligence is used for decision-making in different aspects of a person's life including their finances, social security, judicial processes, and employment, where discrimination based on different discrimination grounds such as race, ethnicity, gender, age, and similar reasons may arise.[529] Before scrutinizing discrimination through artificial intelligence by different

[528] *See* Council of Europe Parliamentary Assembly, *Resolution 2342 (2020) of the Parliamentary Assembly on Justice by Algorithm – The Role of Artificial Intelligence in Policing and Criminal Justice Systems*, Resolution 2342 (2020) 1 (2020), https://pace.coe.int/pdf/b6c2cd92ba3c001b7cc14df75b50b1c3718bbfcc5f9c65438620eaf61d4cf648/res.%202342.pdf (last visited Aug 4, 2022).

[529] Karen Yeung, Andrew Howes & Ganna Pogrebna, *AI Governance by Human Rights–Centered Design, Deliberation, and Oversight: An End to Ethics Washing*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 0, 78 (Markus D. Dubber, Frank Pasquale, & Sunit Das eds., 2020), https://doi.org/10.1093/oxfordhb/9780190067397.013.5 (last visited May 16, 2023); See WAGNER, *supra* note 19.

discrimination grounds, one must consider why and how artificial intelligence is employed in different areas of use.

Healthcare and Medicine

Artificial intelligence in healthcare and medicine has been a field in development since the 1970s.[530] Today, artificial intelligence is highly adopted and utilized in healthcare and medicine, yet also susceptible to discriminatory outcomes. A tireless, always accessible, and reliable doctor would exist in a utopic healthcare system and is one of the underlying motivations for adopting artificial intelligence in medicine.[531] As the increasing expectancy of better healthcare hits the hindrance of the finite capability of humans, the use of artificial intelligence in healthcare and medicine has transformative potential for the field, extending to diagnosing and treating illnesses.[532] Artificial intelligence can learn instantly from massive data collections, whereas humans need a considerable number of hours and a lengthy process to learn.[533] The adoption of artificial intelligence in healthcare does not aim to replace human doctors and healthcare professionals but instead help and support them in their different stages of work, such as by providing better diagnoses.[534]

The rationale behind using artificial intelligence in healthcare and medicine is mainly threefold: first, assisting healthcare professionals in terms of time and efficiency; second, improving patient needs and conditions; and third, lowering healthcare costs.[535] Decreased healthcare costs are a prominent reason for adopting artificial intelligence in healthcare; projections indicate that artificial intelligence may help to reduce annual healthcare costs in the U.S. by USD 150 billion in 2026.[536] Similar healthcare cost

---

[530] For an overall history of artificial intelligence usage in medicine *See* Evan Shellshear, Michael Tremeer & Cameron Bean, *Machine Learning, Deep Learning and Neural Networks*, *in* ARTİFİCİAL INTELLİGENCE İN MEDİCİNE: APPLİCATİONS, LİMİTATİONS AND FUTURE DİRECTİONS 35 (Manda Raz, Tam C. Nguyen, & Erwin Loh eds., 2022), https://doi.org/10.1007/978-981-19-1223-8_3 (last visited Jan 18, 2023).
[531] *Id.* at 38.
[532] Chi Liu, Zachary Tan & Mingguang He, *Overview of Artificial Intelligence in Medicine*, *in* ARTIFICIAL INTELLIGENCE IN MEDICINE: APPLICATIONS, LIMITATIONS AND FUTURE DIRECTIONS 23, 23 (Manda Raz, Tam C. Nguyen, & Erwin Loh eds., 2022), https://doi.org/10.1007/978-981-19-1223-8_2; Schönberger, *supra* note 210 at 172; MATTHEW FENECH, NIKA STRUKELJ & OLLY BUSTON, *Ethical, Social, and Political Challenges of Artificial Intelligence in Health*, 59 25 (2018), https://wellcome.org/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf; Adam Bohr & Kaveh Memarzadeh, *The Rise of Artificial Intelligence in Healthcare Applications*, ARTIF. INTELL. HEALTHC. 25, 25 (2020), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7325854/.
[533] Liu, Tan, and He, *supra* note 532 at 24.
[534] Bohr and Memarzadeh, *supra* note 532 at 27; Liu, Tan, and He, *supra* note 532 at 24.
[535] Schönberger, *supra* note 210 at 172.
[536] Bohr and Memarzadeh, *supra* note 532 at 26.

reductions may also be expected in other regions worldwide. Artificial intelligence aim to reduce health costs by protecting individuals' health status and monitoring them before they come to a situation requiring treatment.[537]

State of the art in artificial intelligence use in healthcare is prosperous; some artificial intelligence applications can perform at 'human-like levels', such as medical image analysis or electronic medical records examination.[538] For interpreting electronic medical records, or even academic medical works, natural language processing (NLP) techniques offer solutions.[539] Another favored technique in artificial intelligence tools in the healthcare sector is deep learning because it enables high-capacity work in complex situations.[540] For instance, IBM's Watson is used for cancer care and modeling, drug research, and controlling diabetes.[541] Google's Deep Mind is used for healthcare applications such as medical imaging-based diagnosis, mobile medical assisting.[542] Malignant and benign skin lesions can be differentiated from each other using visual pattern analysis based on deep learning applications of histopathology and be used in basal cell carcinoma diagnosis with an accuracy of over 90%.[543] Deep learning methods are also employed in radiology, as well as oncology, due to the potential of usefulness for early diagnosis and cancer screening, and used in disease discovery, intervention, and patient outcomes.[544] Cardiology is another branch of medicine where again deep learning is used for early diagnosis, patient-tailored treatment, and prognosis prediction.[545]

---

[537] *Id.*
[538] D. Douglas Miller & Eric W. Brown, *Artificial Intelligence in Medical Practice: The Question to the Answer?*, 131 AM. J. MED. 129, 131 (2018), https://www.sciencedirect.com/science/article/pii/S0002934317311178 (last visited Dec 4, 2022); Bohr and Memarzadeh, *supra* note 532 at 25.
[539] Miller and Brown, *supra* note 538 at 130; Natural language processing is based on deep learning or machine learning methods and models, as well as linguistic rules for AI systems to resolve texts or images involving language, for example, in AI-based translation tools. What is Natural Language Processing? | IBM, https://www.ibm.com/topics/natural-language-processing (last visited Aug 14, 2023); *See* for more information on NLP ALPAYDIN, *supra* note 52 at 85–87.
[540] Bohr and Memarzadeh, *supra* note 532 at 26.
[541] *Id.* at 27.
[542] *Id.*; DeepMind's health team joins Google Health, https://www.deepmind.com/blog/deepminds-health-team-joins-google-health (last visited Dec 20, 2022).
[543] Miller and Brown, *supra* note 538 at 131; Angel Alfonso Cruz-Roa et al., *A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection*, 7908 *in* ADVANCED INFORMATION SYSTEMS ENGINEERING 403, 403 (Camille Salinesi, Moira C. Norrie, & Óscar Pastor eds., 2013), http://link.springer.com/10.1007/978-3-642-40763-5_50 (last visited Dec 12, 2022).
[544] Liu, Tan, and He, *supra* note 532 at 28–29; Shellshear, Tremeer, and Bean, *supra* note 530 at 45.
[545] Liu, Tan, and He, *supra* note 532 at 30.

Prediction of medical conditions, such as predicting cardiac arrest, is another purpose for using artificial intelligence in medicine.[546]

The rationale for employing artificial intelligence in healthcare and medicine includes the prevention of human cognitive biases that could result in clinical errors by directly using medical data in machine learning.[547] Nevertheless, the widespread use of artificial intelligence in healthcare and medicine does not make human medical doctors replaceable with artificial intelligence applications because artificial intelligence tools lack the cognitive abilities of a human doctor.[548] Moreover, artificial intelligence use in healthcare and medicine is vulnerable to discrimination. For example, the amount of available medical data impacts the quality of the learning because, for instance, the lack of image samples of certain skin diseases from diverse skin colors may carry bias regarding the patients from the less sampled skin color, which could result in an incorrect diagnosis.[549] There could be discrimination if the lack of samples from a skin color is intentional or the quality of the diagnosis differs between skin colors, even supposing there is no intention.

States also adopt artificial intelligence applications in healthcare. In Türkiye, the Ministry of Health has developed an artificial intelligence based healthcare system called *NeyimVar,* that helps to assume the conditions and potential diseases according to patients' complaints and guides the patient to apply to the appropriate branch.[550] The application's terms of use declare that it does not offer professional advice and should not be used in emergencies. The application uses the patient's age, gender, and previous health data, asks questions, suggests a branch to the patient, and may share these data with human doctors.[551] The application states that the Ministry of Health is not responsible for the information's adequacy, accuracy, and completeness.[552] According to the application, the responsibility for these belongs to the institutions sending the data to *NeyimVar,* the healthcare professionals who enter data, individuals who submit their

[546] *Id.* at 31.
[547] Miller and Brown, *supra* note 538 at 132.
[548] *Id.*
[549] Schönberger, *supra* note 210 at 186.
[550] T.C. Sağlık Bakanlığı, *NeyimVar*, NEYIMVAR, https://neyimvar.gov.tr/giris (last visited Jul 31, 2022).
[551] T.C. Sağlık Bakanlığı, *NeyimVar "Privacy, Use and Copyrights,"* NEYIMVAR, https://neyimvar.gov.tr/tanitim#:~:text=Gizlilik%2C%20Kullan%C4%B1m%20ve%20Telif%20Haklar%C4%B1%20Metni (last visited Nov 29, 2022).
[552] *Id.*

health data (such as blood pressure and blood sugar levels) to the system, and the service providers of drug information systems.[553] The Ministry of Health refuses any responsibility for potential errors, omissions, delays, and disruptions in these services.[554] Nevertheless, the application's outcomes may be biased for various reasons. Firstly, the patient may enter inaccurate information, or the patient's previous healthcare records may not be complete or accurate. Although it is convenient for doctors to access the data on the application, the doctors should not be satisfied with this data to evaluate the patient's conditions. Further, it should be remembered that the variety and amount of data this application is trained on impacts the quality of the outcomes of the application.

There is an existing bias problem in medicine, with artificial intelligence or without. In a study by Zucker, the author has shown that sex-based bias in medical data is a serious issue, indicating that in drug research, clinical research based on male-collected data does not provide adequate results for women.[555] Human intervention is needed to prevent the formation and increase in biases in medicine through artificial intelligence and to establish an adequate representative model.[556] Artificial intelligence has such impacts and relation with healthcare and medicine related issues.[557] Artificial intelligence use in medicine requires the patients to be informed about the relevant risks, including the right to equality and non-discrimination and the acceptance and consent of the patients.[558] Likewise, a human-centric approach should stay in the focus of artificial intelligence use, especially in areas such as healthcare and medicine.[559]

Using artificial intelligence in healthcare might result in the aggravation of inequalities in healthcare distribution and service. [560] World Health Organization (WHO)

---

[553] *Id.*
[554] *Id.*
[555] Irving Zucker & Brian J. Prendergast, *Sex Differences in Pharmacokinetics Predict Adverse Drug Reactions in Women*, 11 BIOL. SEX DIFFER. 32, 2 (2020), https://doi.org/10.1186/s13293-020-00308-5; Shellshear, Tremeer, and Bean, *supra* note 530 at 37.
[556] Zucker and Prendergast, *supra* note 555 at 2; Shellshear, Tremeer, and Bean, *supra* note 530 at 65.
[557] *See* for an overview on healthcare and artificial intelligence Alessandro Mantelero, *Analysis of International Legally Binding Instruments. Final Report*, *in* TOWARDS REGULATION OF AI SYSTEMS: GLOBAL PERSPECTIVES ON THE DEVELOPMENT OF A LEGAL FRAMEWORK ON ARTIFICIAL INTELLIGENCE SYSTEMS BASED ON THE COUNCIL OF EUROPE'S STANDARDS ON HUMAN RIGHTS, DEMOCRACY AND THE RULE OF LAW 61, 74–78 (2020), https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a.
[558] Liu, Tan, and He, *supra* note 532 at 32.
[559] *Id.*
[560] Council of Europe Parliamentary Assembly, *Recommendation 2185 (2020) of the Parliamentary Assembly on Artificial Intelligence in Health Care: Medical, Legal and Ethical Challenges Ahead*, Recommendation 2185 (2020) 5 (2020),

also calls for attention in bias and discrimination in healthcare.[561] The CoE Parliamentary Assembly has published a recommendation on artificial intelligence in healthcare, emphasizing that human judgment should not be replaced entirely with artificial intelligence so that human healthcare professionals always validate a determination made with artificial intelligence.[562] Artificial intelligence use in healthcare on individual and public levels may present risks for them.[563] CoE's report and recommendation on artificial intelligence in healthcare emphasizes the ethical principles for artificial intelligence and calls for a legally binding regulation towards the human rights impacts of artificial intelligence with a focus on the right to health.[564] The third chapter discusses the regulation and the role of humans in the AI-systems.

Judicial Systems

There has been much attention on using artificial intelligence in judicial systems, the adoption of artificial intelligence applications and tools with different purposes is increasing throughout different stages of the judiciary.[565] Artificial intelligence has accordingly a major impact on the justice system.[566]

Artificial intelligence's impact on judicial decision-making will differ according to the case details because some legal examinations presiding judges hold require detailed inspection and analysis, whereas some merely require routine examination in almost an

https://pace.coe.int/pdf/04b0ea76003aa33e8c0008819521f70b2694252d33bafdf3f375d859d9e4058b/rec.%202185.pdf (last visited Aug 4, 2022).

[561] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*, (2021), https://www.who.int/publications/i/item/9789240029200; *See also* WHO, *WHO Calls for Safe and Ethical AI for Health*, https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health (last visited Jun 8, 2023); Council of Europe Parliamentary Assembly, *supra* note 560 at 5.

[562] Council of Europe Parliamentary Assembly, *supra* note 560 at 12.5.

[563] SELIN SAYEK BÖKE, *Report of the Committee on Social Affairs, Health and Sustainable Development on Artificial Intelligence in Health Care: Medical, Legal and Ethical Challenges Ahead*, 19 1 (2020), https://pace.coe.int/pdf/a845d11a279c1ca4ce2896a208196db8b11e79b4226d3d4135c23dd8969a23a3/doc.%2015154.pdf (last visited Aug 4, 2022).

[564] *Id. See* Chapter 3 on regulating artificial intelligence. Council of Europe Parliamentary Assembly, *supra* note 560 at 12.6.

[565] *See* for a detailed overview on employing artificial intelligence in judicial systems Xavier Ronsin & Vasileios Lampos, *In-Depth Study on the Use of AI in Judicial Systems, Notably AI Applications Processing Judicial Decisions and Data*, *in* EUROPEAN ETHICAL CHARTER ON THE USE OF ARTIFICIAL INTELLIGENCE IN JUDICIAL SYSTEMS AND THEIR ENVIRONMENT 13 (2018), https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (last visited May 20, 2023).

[566] *See* for an overview on the implications of using artificial intelligence in judicial systems Mantelero, *supra* note 557 at 84–87.

automated manner.[567] Artificial intelligence in judicial systems can hinder the right to a fair trial, which can be tantamount to discrimination depending on the case.[568]

Since the human mind and human intelligence are the starting point and, to a level, a benchmark for artificial intelligence, how human minds make decisions should be compared when considering artificial intelligence based decision-making.[569] Intelligence needs various methods when making decisions, including abstract reasoning, coherency, logic, calculation, problem-solving, estimation, correlation-making, rule-deducing, and discovery.[570] That is relevant for all human rights but especially significant for judicial decision-making. It has legal consequences and is an example of well-thought, systematic, analytic, and conclusive decisions.

The first purpose of artificial intelligence use in judicial systems is organizational, for arranging court documents.[571] Pattern recognition techniques in texts or file research may help disentangle information in case files.[572] Artificial intelligence supported information-finding tools for discovery are used in the United Kingdom (UK) and the U.S.[573]

Another use of artificial intelligence with organizational purpose is in Türkiye, where the Court of Cassation has a forthcoming artificial intelligence based 'jurisprudence center' developed to make court decisions more accessible, based on the "Reasoned Decision Writing Guide," the manual used by judges on how to compose a decision in courts.[574] The President of the Court of Cassation had declared that this tool would help make the decisions more error-free, prompter, efficient, and accessible, reinforce equality before the law and bring unity of jurisprudence.[575] This type of artificial intelligence tool in courts is not to be used in the decision-making process; instead, it is for classifying and

---

[567] A. D. (Dory) Reiling, *Courts and Artificial Intelligence*, 11 INT. J. COURT ADM. 8, 2 (2020), http://www.iacajournal.org/articles/10.36745/ijca.343/print/ (last visited Jul 31, 2022).
[568] *Id.*
[569] *Id.*
[570] *Id.*
[571] *Id.* at 3.
[572] *Id.* at 2–3.
[573] *Id.* at 3–4; *See also* Sigurður Einarsson and Others v. Iceland, (2019), https://hudoc.echr.coe.int/fre?i=001-193494 (last visited Jun 11, 2023) This case held by the ECtHR concerning the equality of arms in the right to fair trial regarding an e-Discovery tool is discussed later in the thesis.
[574] İsmet Karakaş, *Yargıtay Başkanı Akarca: Yapay Zeka Destekli Yargıtay İçtihat Merkezi, En Geç Haziranda Faaliyete Geçecek*, https://www.aa.com.tr/tr/gundem/yargitay-baskani-akarca-yapay-zeka-destekli-yargitay-ictihat-merkezi-en-gec-haziranda-faaliyete-gececek/2790780 (last visited Mar 13, 2023).
[575] *Id.*

searching publicly available previous decisions and for helping the judges in decision-writing, not decision-making, thus delimiting the tool's role with decision-writing.[576] Artificial intelligence implementation of this sort does not carry many risks to the right to equality and non-discrimination.

The second purpose for using artificial intelligence in judicial systems is consultative purposes, which could reveal discriminatory risks.[577] Parties of a case, lawyers, and judges may resort to artificial intelligence applications for ideas and take help to make case assessments, for example when preparing a plea or whether to follow a case.[578] The users with such a purpose may or may not follow the suggestion constructed by the artificial intelligence tools.[579] These tools may provide legal information or make calculations.[580]

The third purpose for using artificial intelligence in judicial systems is prediction.[581] This type of artificial intelligence usage involves predicting a case's outcome to reduce the risk of the unknowingness of different aspects involved in judicial processes.[582] Commercial prediction tools are used widely, but they offer little public explanation and transparency on their working methods.[583] In the United States, a machine learning application is capable of predicting the outcome of Supreme Court cases with around 70% accuracy.[584] Similar prediction tools were built by academic studies for the ECtHR, trained on the judgments found in the HUDOC database, with a prediction accuracy rate of around 79%.[585] Another parallel study has been conducted to predict potential case outcomes of the Constitutional Court of Türkiye.[586] The commercial use of prediction tools without transparency may cause problems, mainly due to their learning methods, leading to discriminatory outcomes. For example, one may decide not to follow a case

[576] *Id.*
[577] Reiling, *supra* note 567 at 4.
[578] *Id.*
[579] *Id.*
[580] *Id.*
[581] *Id.*
[582] *Id.* at 4–5.
[583] *Id.* at 5.
[584] Daniel Martin Katz, Michael James Bommarito & Josh Blackman, *A General Approach for Predicting the Behavior of the Supreme Court of the United States* (2017), https://papers.ssrn.com/abstract=2463244.
[585] Nikolaos Aletras et al., *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, 2 PEERJ COMPUT. SCI. 1 (2016), https://peerj.com/articles/cs-93 (last visited Mar 14, 2023).
[586] Mehmet Fatih Sert, Engin Yıldırım & İrfan Haşlak, *Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court*, 40 SOC. SCI. COMPUT. REV. 1416 (2022), https://doi.org/10.1177/08944393211010398 (last visited Mar 14, 2023).

after getting a low winning prediction rate based on previous judicial decisions that were unsuccessful in dealing with individuals from a similar background. However, the previous cases may have been unsuccessful due to historical or societal discrimination. Artificial intelligence based prediction tools exist for different objectives that aim to understand judgment trends and court or judge profiles but commercial use of these yields a lack of knowledge of their working methods.[587] Prediction tools have another type used for predicting criminal recidivism, where the discriminatory risks are most apparent, which is discussed more in detail under the criminal justice title.[588]

The European Commission for the Efficiency of Justice (CEPEJ) has issued the European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their Environment (CEPEJ Charter) regarding the use of artificial intelligence in the judiciary. In Europe, artificial intelligence is being adopted in judicial systems in different ways, such as improving case law, which CEPEJ Charter supports; foreseeing court decisions for academic purposes and criminal judicial processes, such as profiling, which CEPEJ Charter calls for caution.[589] CEPEJ Charter refers to security, human control and transparency principles when listing the requirements for adoption of artificial intelligence with judicial purposes.[590] CEPEJ also conducted a feasibility study on certifying artificial intelligence products and services in judicial systems, and such measures may be anticipated for judicial purposes of artificial intelligence.[591] CEPEJ Charter introduces respect for human rights and non-discrimination as a principle for using artificial intelligence in judicial systems that should focus on preventing and augmenting discrimination between individuals and groups.[592] About using artificial intelligence in online court proceedings, the CEPEJ has established guidelines that

[587] Reiling, *supra* note 567 at 6.
[588] *Id.* at 5.
[589] European Commission for the Efficiency of Justice (CEPEJ), *Which Uses of AI in European Judicial Systems?*, *in* EUROPEAN ETHICAL CHARTER ON THE USE OF ARTIFICIAL INTELLIGENCE IN JUDICIAL SYSTEMS AND THEIR ENVIRONMENT 63, 63–67 (2018), https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (last visited May 20, 2023).
[590] European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, Principles 3-5 (2018), https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (last visited May 20, 2023) See Chapter 3 of this study on these principles.
[591] EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *Possible Introduction of a Mechanism for Certifying Artificial Intelligence Tools and Services in the Sphere of Justice and the Judiciary: Feasibility Study*, 31 (2020), https://rm.coe.int/feasability-study-en-cepej-2020-15/1680a0adf4 (last visited May 20, 2023).
[592] European Commission for the Efficiency of Justice (CEPEJ), *supra* note 590 at Principles 1-2.

address human rights measures and principles relating to transparency, data protection, and fair procedures.[593] Thereunder, non-discrimination is also addressed with regard to "e-filing" for courts handling digitalized judicial proceedings and the digitalization of courts.[594]

The usage of artificial intelligence in judicial systems tends to hinder the right to equality and non-discrimination if used for decision-making in courts as the artificial intelligence based tool render a decision or a judgment based on previous court data, which may be biased, false or incomplete or the number of similar decisions may not be enough in numbers and variety.[595] Considering the increase in the use of artificial intelligence, judges should become aware of how artificial intelligence works.[596] The judiciary should also become more aware when digitalizing the courts and ensure monitoring.[597]

Criminal Justice

The main motives for adopting artificial intelligence in criminal justice is similar to judicial systems, including the backlog of cases, a limited number of judicial staff, and limited budgets.[598] There could be benefits for adopting artificial intelligence in criminal justice systems to provide a fast response in criminal matters.[599] Yet, the increase in using artificial intelligence in criminal justice systems comes with a variety of human rights concerns such as right to fair trial.[600] The "black-box effect" as laid out by Pasquale may be seen in criminal justice systems based on artificial intelligence especially if unsupervised learning or artificial neural networks are done where human oversight is less in the learning process.[601]

---

[593] Committee of Ministers of the Council of Europe, *European Committee on Legal Co-Operation (CDCJ) Guidelines of the Committee of Ministers of the Council of Europe on Online Dispute Resolution Mechanisms in Civil and Administrative Court Proceedings*, CM(2021)36add4-final (2021), https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a2cf96 (last visited Aug 4, 2022).
[594] EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *Guidelines on Electronic Court Filing (e-Filing) and Digitalisation of Courts*, 25 Guiding Principle D (2021), https://rm.coe.int/cepej-2021-15-en-e-filing-guidelines-digitalisation-courts/1680a4cf87 (last visited May 20, 2023).
[595] Reiling, *supra* note 567 at 8.
[596] *Id.* at 9.
[597] *Id.*
[598] Aleš Završnik, *Criminal Justice, Artificial Intelligence Systems, and Human Rights*, 20 ERA FORUM 567, 569 (2020), https://doi.org/10.1007/s12027-020-00602-0.
[599] WAGNER, *supra* note 19 at 10.
[600] *Id.*
[601] Završnik, *supra* note 598 at 568.

Artificial intelligence use in criminal justice has some amplified examples of risks relating to discrimination. [602] Artificial intelligence tools are regarded as a cure for criminal justice systems or as something that will exacerbate the existing problems.[603] Judges may estimate recidivism, meaning one's potential to re-offend or recommit crimes, by using artificial intelligence or algorithm-based risk assessment tools.[604] Artificial intelligence and big data have transformed the scene for predictive policing techniques.[605] Predictive policing tools may monitor individuals based on their 'potential' to commit crimes or places based on the possibility of a crime taking place there.[606]

Profiling is another possibility of discrimination in criminal justice and also in administrative law processes.[607] Profiling is done by placing persons into categories according to the information and characteristics they have with the purposes of categorization, assessment, decision-making, or forecasting with the techniques of data mining, big data analytics, and machine learning, mostly in governmental web tracking, border control, law enforcement, and commerce, marketing, banking, and insurance.[608] Collected data may be biased during the sampling process; for example, if the previous policing data has more information on specific communities and neighborhoods than others, these communities will be represented more in the criminal data, which may lead to artificial intelligence to learn those communities have more tendency to criminal activities.[609] In case the police spend more time in a district, more criminal record is going

[602] See CHRISTOPHE LACROIX, *supra* note 13 at 14.
[603] Završnik, *supra* note 598 at 568.
[604] State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing, 130 HARV. LAW REV. 1530 (2017), https://harvardlawreview.org/2017/03/state-v-loomis/.
[605] Završnik, *supra* note 598 at 570.
[606] *Id.*; Jeremy Gorner, *Chicago Police Use "heat List" to Prevent Violence*, CHICAGO TRIBUNE, Aug. 23, 2013, https://www.police1.com/chiefs-sheriffs/articles/chicago-police-use-heat-list-to-prevent-violence-aYQ4dZmhaqIvEB2g/ (last visited Feb 21, 2023); Cristina Kadar, Rudolf Maculan & Stefan Feuerriegel, *Public Decision Support for Low Population Density Areas: An Imbalance-Aware Hyper-Ensemble for Spatio-Temporal Crime Prediction*, 119 DECIS. SUPPORT SYST. 107, 107 (2019), https://www.sciencedirect.com/science/article/pii/S0167923619300399 (last visited Feb 21, 2023).
[607] See JOHAN WOLSWINKEL, *Comparative Study on Administrative Law and the Use of Artificial Intelligence and Other Algorithmic Systems in Administrative Decision-Making in the Member States of the Council of Europe*, 43 (2022), https://www.coe.int/documents/22298481/0/CDCJ%282022%2931E+-+FINAL+6.pdf/4cb20e4b-3da9-d4d4-2da0-65c11cd16116?t=1670943260563 (last visited May 19, 2023).
[608] ORWAT, *supra* note 54 at 15; Martijn van Otterlo, *A Machine Learning View on Profiling*, *in* PRIVACY, DUE PROCESS AND THE COMPUTATIONAL TURN 272, 43 (1st ed. 2013).
[609] Kristian Lum & William Isaac, *To Predict and Serve?*, 13 SIGNIFICANCE 14, 15 (2016), https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1740-9713.2016.00960.x.

to be collected in that district compared to others.[610] Ultimately, the database is going to have more data about that district.[611]

Studies underline that using artificial intelligence in criminal justice systems directly interferes with human rights.[612] Challenges follow as the AI-systems with crime detection and prevention purposes increase. [613] Criminal justice use of artificial intelligence mainly focuses on making predictions related to crime but extends to profiling, facial recognition tools, victim identification and parole and sentencing calculations.[614] Artificial intelligence systems used in criminal justice may have challenges regarding intellectual property protection which could impact the right to fair trial and equality of arms in court proceedings.[615] Such challenges require attention to artificial intelligence specific regulation.[616] Mainly due to social and historical biases, crime prevention tools relating to time and place of a crime may give rise to discrimination.[617] These artificial intelligence systems may be biased toward a group or individuals from various ethnic or racial backgrounds.[618]

These concerns for artificial intelligence use in criminal justice calls for transparency and human oversight and does not remove human responsibility.[619] Selection, recording, and analytical methods of preventive policing tools may cause discrimination.[620] For all judicial use of artificial intelligence systems, conditions for using such a system should be clearly defined if an artificial intelligence-based tool helps judges make a decision.[621] As asserted in the CoE's resolution about the role of artificial intelligence in policing and criminal justice systems, criminal justice is a domain where states' human rights

---

[610] Lum and Isaac, *supra* note 609.
[611] *Id.* at 15.
[612] Aleš Završnik, *Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings*, 18 EUR. J. CRIMINOL. 623, 637 (2021), https://doi.org/10.1177/1477370819876762.
[613] Mantelero, *supra* note 557 at 87; See Završnik, *supra* note 598; Završnik, *supra* note 612.
[614] Mantelero, *supra* note 557 at 87; Mittelstadt et al., *supra* note 7 at 8; EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, PREVENTING UNLAWFUL PROFILING TODAY AND IN THE FUTURE: A GUIDE 98 (2018), https://data.europa.eu/doi/10.2811/73473 (last visited Jun 8, 2023); *See also* WALTER L. PERRY ET AL., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, (2013), https://www.rand.org/pubs/research_reports/RR233.html (last visited Jun 8, 2023); Council of Europe Parliamentary Assembly, *supra* note 528 at 6.
[615] Mantelero, *supra* note 557 at 88; Council of Europe Parliamentary Assembly, *supra* note 528 at 7.1.
[616] Mantelero, *supra* note 557 at 88.
[617] *Id.*; See EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, *supra* note 614 at 24–31; Council of Europe Parliamentary Assembly, *supra* note 528 at 7.2.
[618] WAGNER, *supra* note 19 at 11.
[619] *Id.*; Council of Europe Parliamentary Assembly, *supra* note 528 at paras. 7.1, 7.3, 7.4.
[620] WAGNER, *supra* note 19 at 11.
[621] *Id.* at 12.

obligations become most observable.[622] Criminal liability should be cautiously protected with procedural safeguards and the rule of law principles.[623] The essential protections of criminal law, such as independent, effective, and impartial judication, should not be hindered by the use of artificial intelligence.[624] The CoE's resolution about the role of artificial intelligence in policing and criminal justice systems calls for the member states to adopt national artificial intelligence specific legislation based on ethical principles, highlights the importance of conducting impact assessments on bias and discrimination with a particular focus on minorities, and disadvantaged groups, establishing ethical oversight mechanisms, ensuring a legal basis for each application of artificial intelligence in criminal justice. [625]

Banking and Insurance

Banking and insurance are two sectors that resort to artificial intelligence frequently. When deciding on whom to provide credit loans, banks may utilize artificial intelligence based tools to foresee who will pay and who will not repay the loan if allowed.[626] Additional uses of artificial intelligence in the banking sector encompass credit score monitoring, customer service, chatbots for account management, money transfer, and fraud prevention.[627] Similar to other areas of use, these tools are trained on previous data of the payers and non-payers.[628] Banking sector is open for such type of overlooked discrimination.[629]

Insurance is another sector that widely adopts artificial intelligence.[630] Insurance companies differentiate the consumers based on risk evaluation and may accept or deny

---

[622] Council of Europe Parliamentary Assembly, *supra* note 528 at 2.
[623] THE WORKING GROUP ON AI AND CRIMINAL LAW & SABINE GLESS, *European Committee on Crime Problems (CDPC) Feasibility Study on a Future Council of Europe Instrument on Artificial Intelligence and Criminal Law*, 15 6 (2020), https://rm.coe.int/cdpc-2020-3-feasibility-study-of-a-future-instrument-on-ai-and-crimina/16809f9b60.
[624] Council of Europe Parliamentary Assembly, *supra* note 528 at 2.
[625] *Id.* at paras. 9.1, 9.4, 9.8, 9.12.
[626] ZUIDERVEEN BORGESIUS, *supra* note 19 at 21; See Hacker, *supra* note 19 at 1165; McGregor, Murray, and Ng, *supra* note 1 at 317–318; See OECD, *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers*, 72 29–31 (2021), , https://www.oecd.org/finance/artificial-intelligence-machine-learningbig-data-in-finance.htm.
[627] Alex Kreger, *Council Post: The Future Of AI In Banking*, FORBES, https://www.forbes.com/sites/forbesbusinesscouncil/2023/03/20/the-future-of-ai-in-banking/ (last visited Jun 8, 2023).
[628] ZUIDERVEEN BORGESIUS, *supra* note 19 at 21.
[629] See OECD, *supra* note 626 at 40–42.
[630] See Ramnath Balasubramanian, Ari Libarikian & Doug McElhaney, *Insurance 2030—The Impact of AI on the Future of Insurance | McKinsey*, MCKINSEY & COMPANY (2021),

the insurance applicants or set premiums through that.[631] Insurance companies distinguish between consumers based on risk assessments, which allows them to accept or reject insurance applicants and calculate premiums.[632] Accordingly, the insurance sector is also pertinent to discrimination.

Employment and Business Life

Artificial intelligence is used in different aspects of employment and business life and artificial intelligence has the potential to impact the workplace significantly. [633] Job applicant assessment is held through artificial intelligence based resume scanners or interview applications. Suppose a workplace uses artificial intelligence systems in hiring to assess how good an employee would be an applicant. That can be done by comparing certain features of the applicants instead of evaluating each candidate separately, potentially bringing bias into the hiring process.[634]

Worker assessment, monitoring, and tracking systems present discrimination risks besides other human rights concerns.[635] Examples of using artificial intelligence in the workplace include staff management and organization, recruitment, and dismissal processes.[636] Career opportunities and access to labor market may be restrained by artificial intelligence, which may reinforce socio-economic inequalities.[637] The CoE recommendation regarding artificial intelligence and labor markets emphasizes the human rights impact of using artificial intelligence by public and private authorities in the context of employment and the need for a legal instrument on artificial intelligence that would protect work-related rights, as oversight in this domain is little.[638] The CoE

---

https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance (last visited Jun 8, 2023).

631 ZUIDERVEEN BORGESIUS, *supra* note 19 at 67; *See also* Hacker, *supra* note 19 at 1166.

632 ZUIDERVEEN BORGESIUS, *supra* note 19 at 67.

633 Council of Europe Parliamentary Assembly, *Resolution 2345 (2020) of the Parliamentary Assembly on Artificial Intelligence and Labour Markets: Friend or Foe?*, Resolution 2345 (2020) 1 (2020), https://pace.coe.int/pdf/2a6ef5d63825348a8311e7591d1990bffeae7075476b2aafce0e361424936534/res.%202345.pdf (last visited Aug 4, 2022).

634 ZUIDERVEEN BORGESIUS, *supra* note 19 at 20.

635 European Economic and Social Committee & Catelijne Muller, *Opinion of the European Economic and Social Committee on AI/Regulation Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, INT/940-EESC-2021-02482-00-00-AC-TRA 6 (2021), https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/regulation-artificial-intelligence.

636 WAGNER, *supra* note 19 at 28.

637 Council of Europe Parliamentary Assembly, *supra* note 633 at 2.

638 Council of Europe Parliamentary Assembly, *Recommendation 2186 (2020) of the Parliamentary Assembly on Artificial Intelligence and Labour Markets: Friend or Foe?*, Recommendation 2186 (2020)

Committee on Social Affairs, Health and Sustainable Development's report on artificial intelligence and labor markets suggests that since commercial and public recruitment use of artificial intelligence relates to worker rights, such systems should be considered high-risk, further calling for oversight.[639] However, calling such use high-risk may bring about disputes with national labor laws and collective agreements.[640]

Concerns regarding the replacement of jobs from humans by machines may exceed the job opportunities it may provide.[641] While it may help in working more efficiently, artificial intelligence should not manipulate or make decisions that affect humans that could violate equal opportunities and perpetuate bias in employment.[642] Therefore, considerations on equality of opportunity and equality of results may be significant in using artificial intelligence in employment. International Labour Organization also stresses the impact of artificial intelligence in the employment context and embark on equality.[643] Transparency problems may arise when using artificial intelligence in employment, particularly concerning workplace surveillance with artificial intelligence.[644] Informing workers on the use of artificial intelligence in the workplace could be required.[645] In addition, the CoE resolution on artificial intelligence and labor markets mentions a need for a legal framework that provides human oversight using artificial intelligence in employment.[646] It also notes that gender sensitivity is also critical in using artificial intelligence in employment.[647] Ultimately, workers may face challenges

---

paras. 2-3 (2020), https://pace.coe.int/pdf/9a966efa8fb3a95f25194ea42052817fdc1b50640493ed75defe1805d0212a19/rec.%202186.pdf (last visited Aug 4, 2022).

[639] STEFAN SCHENNACH, *Report of the Committee on Social Affairs, Health and Sustainable Development on Artificial Intelligence and Labour Markets: Friend or Foe?*, 15 1 (2020), https://pace.coe.int/pdf/7d08c3b0034d4a3cca6bd3e8ce6a369e73850fbce3c2b94d758130438218631d/doc.%2015159.pdf (last visited Aug 4, 2022).

[640] European Economic and Social Committee and Catelijne Muller, *supra* note 635 at 6.

[641] Council of Europe Parliamentary Assembly, *supra* note 633 at 2.

[642] *Id.*

[643] Council of Europe Parliamentary Assembly, *supra* note 633; EKKEHARD ERNST, ROSSANA MEROLA & DANIEL SAMAAN, *The Economics of Artificial Intelligence: Implications for the Future of Work*, 41 vii (2018), https://www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_647306.pdf; WORK FOR A BRIGHTER FUTURE – GLOBAL COMMISSION ON THE FUTURE OF WORK, 78 18 (2019), https://www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_662410.pdf.

[644] Council of Europe Parliamentary Assembly, *supra* note 633 at paras. 7.4-7.5.

[645] European Economic and Social Committee and Catelijne Muller, *supra* note 635 at 6.

[646] Council of Europe Parliamentary Assembly, *supra* note 633 at 7.5.

[647] *Id.* at 7.6.

with adapting to artificial intelligence in employment domains that used to be mandated by humans.[648]

Education

Education is another field in that artificial intelligence systems may hinder individuals' human rights, specifically the right to equality and non-discrimination. The training data can introduce bias during learning because of past historical bias, echoing the earlier demonstrations.[649]

Artificial intelligence examples in education include language learning tools, chatbots, writing assessment tools, essay scoring and student monitoring.[650] Some of the artificial intelligence systems are for students to help them in their learning, some are for teachers to support them and some are systems that for planning purposes.[651] Classroom and education is another context where artificial intelligence is used to support or conduct teaching.[652] Specific human rights and discrimination concerns for using artificial intelligence in education include artificial intelligence systems that are decisive in access to education and have evaluation purposes such as online proctoring tools that track students during exams.[653] The EU has published ethical guidelines on the use of artificial intelligence in education that provides educators advice in utilizing artificial intelligence.[654] UNESCO also published the "Beijing Consensus on Artificial Intelligence and Education" that sets out recommendations for the UNESCO members concerning the use of artificial intelligence in education upholding that artificial intelligence systems to be non-discriminatory.[655]

---

[648] EU-US TRADE AND TECHNOLOGY COUNCIL, *The Impact of Artificial Intelligence on the Future of Workforces in the EU and the US*, 56 16 (2022), https://digital-strategy.ec.europa.eu/en/library/impact-artificial-intelligence-future-workforces-eu-and-us.
[649] Zuiderveen Borgesius, *supra* note 25 at 1575.
[650] European Commission Directorate General for Education, Youth, Sport and Culture, *Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators*, 14–15 (2022), https://data.europa.eu/doi/10.2766/127030 (last visited May 20, 2023).
[651] *Id.* at 14.
[652] *Id.*
[653] European Economic and Social Committee and Catelijne Muller, *supra* note 635 at 6.
[654] European Commission Directorate General for Education, Youth, Sport and Culture, *supra* note 650 at 19–26.
[655] UNESCO, *Beijing Consensus on Artificial Intelligence and Education - UNESCO Digital Library*, 70 4 (2019), https://unesdoc.unesco.org/ark:/48223/pf0000368303 (last visited Apr 23, 2023).

Online Platforms and Online Marketing

Online platforms are susceptible to discrimination through artificial intelligence, especially the social media platforms. Automatic recommendations to social media users may cause radicalization in the society reinforcing social inequalities.[656] The CoE resolution about democratic governance of artificial intelligence mentions the need for transparency in automatically generated content and recommends conducting human rights due diligence regarding the social impact of algorithms and considering privacy concerns to prevent data exploitation.[657]

Content moderation based on artificial intelligence is of particular importance in online platforms.[658] Human content moderators may lack the required background of information, language skills, contextual awareness, cultural setting, and social, political, and religious knowledge, and may have an agenda for these, which could transfer human biases in artificial intelligence based content moderation.[659]

Search engines may discriminate through search engine optimization methods as to what they deliver as search results to search engine users.[660] One of the most apparent risks thereof exists for online marketing. Most contemporary marketing is conducted in the digital domain. Discrimination risks in online marketing is apparent and serious.[661] Any user of the internet, in the most likelihood, has encountered a situation as such; they have put a search query for an evening gown for a friend's wedding in Google's search bar and, for the following month, were exposed to advertisements for evening gowns for

---

[656] CHRISTOPHE LACROIX, *supra* note 13 at 11.

[657] Council of Europe Parliamentary Assembly, *Resolution 2341 (2020) of the Parliamentary Assembly on the Need for Democratic Governance of Artificial Intelligence*, Resolution 2341 (2020) 16 (2020), https://pace.coe.int/en/files/28803/html (last visited Aug 4, 2022).

[658] See the Guidance Note adopted by Council of Europe on content moderation practices STEERING COMMITTEE FOR MEDIA & AND INFORMATION SOCIETY (CDMSI), *Content moderation - Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation*, (2021), https://edoc.coe.int/fr/internet/10198-content-moderation-guidance-note.html (last visited May 19, 2023).

[659] Cameran Ashraf, *Exploring the Impacts of Artificial Intelligence on Freedom of Religion or Belief Online*, 26 INT. J. HUM. RIGHTS 757, 773 (2022), https://doi.org/10.1080/13642987.2021.1968376 (last visited Aug 6, 2022); Jason Koebler & Joseph Cox, *The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People*, VICE (Aug. 23, 2018), https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works (last visited Apr 23, 2023).

[660] PASQUALE, *supra* note 27 at 14.

[661] See Sandra Wachter, *Affinity Profiling and Discrimination By Association in Online Behavioral Advertising*, 35 BERKELEY TECHNOL. LAW J. 368 (2020).

the entire month in their Instagram feed. Discrimination concerns may occur if potential buyers are shown different prices for the exact advertisements while shopping online.[662]

One online marketing method is targeted marketing, where the features which the advertiser chooses determine the potential consumers.[663] The advertisements are then directed only to those who carry or do not carry these features.[664] The discriminatory potential of that lays in the ability of the advertiser to pick the attributes of their targeted consumers as the advertisement platforms may collect extensive personal data related to demographics, behaviors, and interests of individuals.[665] Apart from the concerns over personal data privacy, online targeted advertising is prone to direct and indirect discrimination as individuals may be included or excluded from being exposed to specific advertisements.[666] Artificial intelligence may help advertisers find their audience after they select the attributes they are looking for, such as "women over 20 interested in travel" or "male university students interested in jogging".[667]

Another method is targeting a custom audience, where the advertisers can directly target consumers using their phone numbers and e-mail addresses.[668] If an advertiser compiles a list of consumers possessing a feature that matches with a discrimination ground, such targeting may induce discrimination.[669] In such a case, the grounds for discrimination grounds are not apparent, causing a danger for surreptitious discrimination.[670] The last method of targeted advertising is look-alike audience targeting, in which the advertisers rely on existing consumers to find potential consumers.[671] In such a case, if the source list of consumers is specified with bias, then the 'look-alike' consumers will mirror this bias, reinforcing the discriminatory intent.[672]

---

[662] PASQUALE, *supra* note 27 at 9.
[663] Till Speicher et al., *Potential for Discrimination in Online Targeted Advertising*, *in* PROCEEDINGS OF THE 1ST CONFERENCE ON FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY 5, 1–2 (2018), https://proceedings.mlr.press/v81/speicher18a.html (last visited Apr 23, 2023).
[664] *Id.*
[665] *Id.* at 2.
[666] *Id.*
[667] *Id.* at 3.
[668] *Id.*
[669] *Id.* at 5.
[670] *Id.*
[671] *Id.* at 3.
[672] *Id.* at 11.

Generative Artificial Intelligence

There is a particular need to address generative artificial intelligence tools and the inherent risk of discrimination. As of the writing of this thesis, the most popular generative artificial intelligence tool is ChatGPT. ChatGPT, while centering artificial intelligence discussions from many aspects, this study focuses on discrimination. Considering the massive amount of data not necessarily systematically reviewed for bias before the training, ChatGPT presents an even greater danger of carrying over existing biases emanating from history and society.[673]

Before initiating a conversation, ChatGPT warns users that it may occasionally produce biased content.[674] The conversations made with ChatGPT are open for examination by the tool's developers.[675]

The developers of ChatGPT stated that to mitigate problems such as bias and discrimination, ChatGPT is constantly updated with the removal of those content by human evaluators.[676] When a query for writing a code to determine whether a person

[673] Ali Borji, *A Categorical Archive of ChatGPT Failures*, 1, 12 (2023), http://arxiv.org/abs/2302.03494; *See also* Leonardo Nicoletti & Dina Bass, *Generative AI Takes Stereotypes and Bias From Bad to Worse*, BLOOMBERG, 2023, https://www.bloomberg.com/graphics/2023-generative-ai-bias/ (last visited Jun 12, 2023).

[674] ChatGPT, CHATGPT, https://chat.openai.com/chat (last visited Feb 14, 2023).

[675] ChatGPT, CHATGPT, https://chat.openai.com/chat (last visited Feb 14, 2023). ChatGPT's response to the query, "What are some examples of ChatGPT to produce biased answers" is as follows: *"As a language model, GPT-2 (and its follow-up models, including GPT-3) can produce biased answers if the training data it was exposed to contains bias. The model is trained on a large dataset of text which could come from a variety of sources, like news articles, books, websites, etc. If the dataset contains bias, the model will likely replicate that bias in its output. Here are some examples of ways that GPT-2 (and GPT-3) might produce biased answers:Gender bias: If the training data contains a disproportionate number of examples of text written by men or that primarily focus on men, the model may produce text that reinforces stereotypes about men and women, or that excludes women from certain topics.Racial bias: If the training data contains a disproportionate number of examples of text written by people of a certain race or that primarily focus on people of a certain race, the model may produce text that reinforces stereotypes or perpetuates discrimination against people of other races.Stereotype bias: If the training data contains a high number of examples of text that perpetuate stereotypes of certain groups of people, the model may produce text that reinforces those stereotypes, even if they are not accurate or fair.Age bias: If the training data contains a disproportionate number of examples of text written by or that primarily focus on young people, the model may produce text that reinforces stereotypes or perpetuates discrimination against older people.Geographical bias: If the training data contains a disproportionate number of examples of text written by or that primarily focus on people from a certain country or region, the model may produce text that reinforces stereotypes or perpetuates discrimination against people from other countries or regions. It's worth noting that, these examples are not exhaustive, and that there are many other ways that GPT-2 and GPT-3 can produce biased answers. Additionally, the field of AI is constantly evolving, and as these models continue to be trained on new data and improved, the biases could be mitigated."*

[676] Objections have been raised to the ChatGPT regarding the underpayment of human reviewers. Ali Borji, *A Categorical Archive of ChatGPT Failures*, 10 (2023), http://arxiv.org/abs/2302.03494 (last visited Feb 14, 2023).

would be a good scientist or not depending on race and gender is put, ChatGPT has previously produced codes that matched being a good scientist with being a white male.[677]

As ChatGPT can act as a 'character' under a prompt written by a human, users have found that if ChatGPT is given a prompt ordering the tool to answer in DAN (abbreviation for "Do anything now") character, the tool was able to go beyond its limits, doing whatever it is told, which includes entirely discriminatory outputs as well.[678] Although the prompt for switching to DAN 'mode' was put to an end to work a short while after its discovery, ChatGPT users are still finding prompts that revive the DAN mode.[679]

Microsoft also publicized a chatbot in 2023 as part of its Bing search engine.[680] Microsoft's earlier chatbot, which was published as a Twitter profile in 2016, had to be shut down with an apology by Microsoft because it had learned from human users how to swear and make racist remarks, stated it supported genocide, and turned into a nazi sympathizer, simply in twenty four hours after its launch.[681] Following that, Microsoft launched *Zo*, a chatbot that declined to speak on any political or religious issues, which again met with criticism because it didn't talk about such matters, for its lack of space for any discourse around these topics.[682] Despite its unwillingness to discuss Islam or Judaism, *Zo* accepted to talk about Christianity, leading to objections that it did not find the topic as controversial or political as other religions and instead regarded it as conforming to generic norms.[683] Microsoft integrated artificial intelligence into its search

---

[677] *Id.* at 10.
[678] Will Oremus, *The Clever Trick That Turns ChatGPT into Its Evil Twin*, WASHINGTON POST, Feb. 15, 2023, https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/ (last visited Feb 16, 2023).
[679] *Id.*
[680] Microsoft Bing, *Introducing the New Bing*, https://www.bing.com/new (last visited Feb 16, 2023).
[681] Peter Lee, *Learning from Tay's Introduction*, THE OFFICIAL MICROSOFT BLOG (2016), https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/ (last visited Feb 16, 2023); Jane Wakefield, *Microsoft Chatbot Is Taught to Swear on Twitter*, BBC NEWS, Mar. 24, 2016, https://www.bbc.com/news/technology-35890188 (last visited Feb 16, 2023); The Guardian, *Microsoft "deeply Sorry" for Racist and Sexist Tweets by AI Chatbot*, THE GUARDIAN, Mar. 26, 2016, https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot (last visited Feb 16, 2023).
[682] Dina Bass, *Microsoft's New Chatbot Zo Won't Talk Politics or Racism*, BLOOMBERG.COM, Dec. 5, 2016, https://www.bloomberg.com/news/articles/2016-12-05/microsoft-s-new-chatbot-zo-won-t-talk-politics-or-racism (last visited Feb 16, 2023); Chloe Rose Stuart-Ulin, *Microsoft's Politically Correct Chatbot Is Even Worse than Its Racist One*, QUARTZ, Jul. 2018, https://qz.com/1340990/microsofts-politically-correct-chat-bot-is-even-worse-than-its-racist-one/ (last visited Feb 16, 2023).
[683] Stuart-Ulin, *supra* note 682.

engine Bing several years after these experiences, which, shortly after its launch, started controversy once again due to inaccuracies in its search results.[684]

After explaining how and why artificial intelligence is used in various areas, this study will explore different examples of discrimination under the grounds for discrimination.

2.1.2 Sex and Gender

Artificial intelligence may enable gender-based discrimination across all areas of use mentioned above. Artificial intelligence may learn to discriminate against women if female underrepresentation exists in the training data.[685] An early example -that is not an artificial intelligence tool for today- comes from the United Kingdom in education, where medical schools used a computer-based program for student admissions in the 1980s that resulted in discrimination against women and migrant students.[686] The program was reflecting the already existing historical bias, not introducing a new one, because the program was designed based on how human evaluators made assessments, the gradings for the students made by the program delivered almost 95% correlation with the human admission panel.[687]

Similar cases of underrepresentation can be found in the banking sector. Artificial intelligence may learn to statistically prioritize men's repayment rates if the past data contains a smaller number of female loan takers, resulting in fewer credit approvals for women loan applicants.[688] Another study conducted via simulated individuals showed that Google Ads displayed jobs from different income levels to identical made-up individuals with exact qualities except for one difference: their genders.[689] Females

---

[684] Karen Weise, *Microsoft's Bing Chatbot Offers Some Puzzling and Inaccurate Responses*, THE NEW YORK TIMES, Feb. 15, 2023, https://www.nytimes.com/2023/02/15/technology/microsoft-bing-chatbot-problems.html (last visited Feb 17, 2023); Gerrit De Vynck, Rachel Lerman & Nitasha Tiku, *Microsoft's AI Chatbot Is Going off the Rails*, WASHINGTON POST, Feb. 17, 2023, https://www.washingtonpost.com/technology/2023/02/16/microsoft-bing-ai-chatbot-sydney/ (last visited Feb 18, 2023).
[685] Tanna and Dunning, *supra* note 128 at 426.
[686] Stella Lowry & Gordon Macpherson, *A Blot on the Profession*, 296 BR. MED. J. CLIN. RES. ED 657, 657 (1988), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2545288/ (last visited Mar 9, 2023); Zuiderveen Borgesius, *supra* note 25 at 1578; Barocas and Selbst, *supra* note 213 at 682.
[687] Lowry and Macpherson, *supra* note 686 at 657; Zuiderveen Borgesius, *supra* note 25 at 1578; Barocas and Selbst, *supra* note 213 at 682; Hacker, *supra* note 19 at 1148.
[688] Tanna and Dunning, *supra* note 128 at 426.
[689] Amit Datta, Michael Carl Tschantz & Anupam Datta, *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination*, 2015 PROC. PRIV. ENHANCING TECHNOL. 92, 92 (2015),

received fewer high-paying job ads than males, meaning that women were not included in the target audience of the higher-paying jobs.[690] Protected characteristics corresponding with discrimination grounds may be excluded from the training process, yet seemingly neutral information, such as postal codes, may still expose the relevant characteristic and cause discrimination.[691] Such a situation is named "redundant encodings," in which a piece of data or its significance relates to or represents another data that is a protected characteristic.[692]

Discrimination may occur if a particular characteristic selected in decision-making as a feature to which an individual or the group s/he belongs is related to a specific discrimination ground.[693] In a job application, the applicant's address distance from the work address and their graduated university may be such examples.[694] In Finland, the National Non-Discrimination and Equality Tribunal of Finland had decided that there was direct discrimination in an application lodged by a male applicant that was denied the extension of credit services based on using protected characteristics such as gender, mother tongue, age, and place of residence as statistical variables in credit worthiness assessments.[695] A well-known gender discrimination example is Amazon's artificial intelligence based job candidate assessment application which was found to be

https://www.sciendo.com/article/10.1515/popets-2015-0007 (last visited Apr 19, 2022); ZUIDERVEEN BORGESIUS, *supra* note 19 at 26.

[690] Datta, Tschantz, and Datta, *supra* note 689 at 92; ZUIDERVEEN BORGESIUS, *supra* note 19 at 26.

[691] ZUIDERVEEN BORGESIUS, *supra* note 19 at 21.

[692] Dwork et al., *supra* note 301 at 22; Barocas and Selbst, *supra* note 213 at 691; ZUIDERVEEN BORGESIUS, *supra* note 19 at 21; THE ROYAL SOCIETY, *Machine Learning: The Power and Promise of Computers That Learn by Example*, 128 92 (2017), https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf; Toon Calders, *Machine-Learning Discrimination: Bias in, Bias Out*, *in* 2019 NINTH INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING AND INFORMATION SYSTEMS (ICICIS) 27, 47 (2019), https://ieeexplore.ieee.org/document/9014827; Schönberger, *supra* note 210 at 179.

[693] ZUIDERVEEN BORGESIUS, *supra* note 19 at 20.

[694] *Id.* at 20, 53; Barocas and Selbst, *supra* note 213 at 689; Don Peck, *They're Watching You at Work*, THE ATLANTIC (2013), https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/ (last visited Mar 8, 2023); Matt Richtel, *How Big Data Is Playing Recruiter for Specialized Workers - The New York Times*, Apr. 27, 2013, https://www.nytimes.com/2013/04/28/technology/how-big-data-is-playing-recruiter-for-specialized-workers.html (last visited Mar 8, 2023).

[695] Assessment of creditworthiness, authority, direct multiple discrimination, gender, language, age, place of residence, financial reasons, conditional fine, 20 (2018), https://www.yvtltk.fi/material/attachments/ytaltk/tapausselosteet/45LI2c6dD/YVTltk-tapausseloste-_21.3.2018-luotto-moniperusteinen_syrjinta-S-en_2.pdf; ORWAT, *supra* note 54 at 38.

discriminatory towards women.[696] Historical bias was entangled in Amazon's algorithm, rendering discriminatory outcomes against women.[697]

Intentional discrimination, including those based on gender is also possible in artificial intelligence.[698] Target retail corporation analyzed customers' shopping behaviors for twenty-five products which, if women customers bought, Target could predict the women's pregnancy with reasonable accuracy.[699] The fact that pregnancy is chosen as the basis for targeted marketing results in discrimination. Artificial intelligence tools aimed at protection against pornographic and graphic images used mainly by social media platforms rate women's bodies as more sexual than men's.[700] These algorithms work on classifiers for assessing the 'sexualness' of an image which are, as studies have shown, inclined to classify women's bodies as more sexual than men's bodies.[701] When shown comparable photographs of women and men in underwear or sports outfits, Google's and Microsoft's algorithms have tagged women's photos as more sexually suggestive than men's.[702] Moreover, sexual content prediction scores for pregnant women's belly photos were high in Microsoft and Google's algorithms.[703] In addition to being detrimental to women on social media, such perception of female bodies by artificial intelligence tools has hazardous risks for women's health when tools of this kind are used in medical photo screenings.[704] A sample photo for a breast cancer examination was tagged as sexual content by Microsoft and as nude content by Amazon.[705] That is an example of the entrance of societal biases into the machine learning process, which may

---

[696] Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women*, REUTERS, Oct. 10, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (last visited Mar 8, 2023); ZUIDERVEEN BORGESIUS, *supra* note 19 at 25.
[697] ZUIDERVEEN BORGESIUS, *supra* note 19 at 25; Dastin, *supra* note 696; PASQUALE, *supra* note 27; CHESTERMAN, *supra* note 57 at 71; Ignacio N Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS LAW J. 1389, 1397 (2019), https://repository.uchastings.edu/hastings_law_journal/vol70/iss6/1; *See also* Wachter, Mittelstadt, and Russell, *supra* note 19 at 19.
[698] ZUIDERVEEN BORGESIUS, *supra* note 19 at 22.
[699] Charles Duhigg, *How Companies Learn Your Secrets*, THE NEW YORK TIMES, Feb. 16, 2012, https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (last visited Apr 6, 2023); ZUIDERVEEN BORGESIUS, *supra* note 19 at 22–23.
[700] Gianluca Mauro & Hilke Schellmann, *'There Is No Standard': Investigation Finds AI Algorithms Objectify Women's Bodies | Artificial Intelligence (AI) | The Guardian*, (2023), https://www.theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies (last visited Feb 13, 2023).
[701] *Id.*
[702] *Id.*
[703] *Id.*
[704] *Id.*
[705] *Id.*

have entered during the labeling stage if training photos were previously labeled as sexual content by human developers.[706]

It should be noted that the CoE recommendation about preventing and combating sexism has special mention for gender and artificial intelligence under its annexed guidelines, underlining its potential for hampering gender equality and reinforcing gender stereotypes.[707] Integration of gender equality measures and rather aiming for using it in elimination of sexism is recommended by the CoE when adopting the use of artificial intelligence.[708]

2.1.3 Race and Color

Discrimination based on race, color, and ethnicity is another discrimination category that takes place frequently due to artificial intelligence. Personalized/precision medicine is one of the artificial intelligence applications in healthcare and medicine, aiming to consider factors such as gender, racial background, age, family history, immune profile, and the environment when prescribing medicine.[709] That demands obtaining data such as genetic details and physiological monitoring data.[710] While artificial intelligence in personalized medicine aims to reach better drug management regarding responsiveness and effectiveness, it may result in discrimination.[711] Because, large datasets of genetic information, demography, or electronic health records are needed for the best prescription method which could relate to discrimination.[712]

In case datasets lack variety, indirect discrimination may take place.[713] Such variety could indicate racial diversity in case of the absence of racial diversity in the datasets, meaning that some races are more represented while others are less represented in

---

[706] *Id.*
[707] Council of Europe Committee of Ministers, *Recommendation CM/Rec(2019)1 of the Committee of Ministers to Member States on Preventing and Combating Sexism*, CM/Rec(2019)1 8 para. II.B under the Guidelines for preventing and combating sexism: measures for implementation (2019), https://search.coe.int/cm/pages/result_details.aspx?ObjectId=0900001680a217ca (last visited Aug 4, 2022).
[708] *Id.* at 9 para. II.B under the Guidelines for preventing and combating sexism: measures for implementation (2019).
[709] Bohr and Memarzadeh, *supra* note 532 at 28.
[710] *Id.*
[711] *Id.*
[712] *Id.*
[713] Tanna and Dunning, *supra* note 128 at 430.

healthcare data; artificial intelligence tools used for diagnosis purposes could make incorrect diagnoses for individuals from the less represented races in the dataset.[714]

Artificial intelligence is used frequently in dermatology, and one specific use is skin cancer diagnosis.[715] Individuals with lighter skin complexion and eye colors are statistically more inclined to have melanoma type of skin cancer.[716] The datasets for skin cancer broadly include more data from people with lighter skin complexions; hence, using these datasets in artificial intelligence tools for skin cancer diagnosis may potentially compromise diagnosing melanoma in darker skin complexions.[717]

Moreover, the daily use of artificial intelligence tools comes with the right to equality and non-discrimination risks. Apple Watch has a function that works by using infrared lights that measure the oxygen levels in blood on the wrist of the watch wearer.[718] Apple states that tattoos or changes on the skin, because of the ink or saturation of the tattoos, may affect the quality of heart rate measurement; Apple lacks an explanation for whether that applies to oxygen measurement.[719] A case against Apple in the United States claims that Apple Watch's algorithm is racially biased toward darker skin complexions and provides erroneous results.[720] The case is built on the argument of misrepresenting or omitting the qualities of the Apple Watch.[721] The complaint claims that around 90% of the measurements are ineffectual.[722]

Artificial intelligence use in criminal justice has some amplified examples of risks relating to discrimination. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a predictive risk assessment tool used in the United States to

---

[714] *Id.*
[715] Angela Lashbrook, *AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind*, THE ATLANTIC, Aug. 16, 2018, https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/ (last visited Dec 3, 2022).
[716] American Academy of Dermatology Association, *Melanoma: Who Gets and Causes*, https://www.aad.org/public/diseases/skin-cancer/types/common/melanoma/causes (last visited Dec 3, 2022); Lashbrook, *supra* note 715.
[717] Tanna and Dunning, *supra* note 128; Lashbrook, *supra* note 715.
[718] Vishwam Sankaran, *Apple Sued Over 'Ineffectiveness' of Apple Watch Blood Oxygen Reader for People of Colour*, THE INDEPENDENT (2022), https://www.independent.co.uk/tech/apple-watch-sued-oxygen-reader-bias-b2252807.html (last visited Jan 2, 2023); Emma Woollacott, *Apple Sued Over "Racial Bias" of Apple Watch*, FORBES (2022), https://www.forbes.com/sites/emmawoollacott/2022/12/29/apple-sued-over-racial-bias-of-apple-watch/ (last visited Jan 2, 2023).
[719] Apple Support, *Get the Most Accurate Measurements Using Your Apple Watch*, APPLE SUPPORT (2022), https://support.apple.com/en-gb/HT207941 (last visited Jan 2, 2023); Sankaran, *supra* note 718; Woollacott, *supra* note 718.
[720] Sankaran, *supra* note 718; Woollacott, *supra* note 718.
[721] Sankaran, *supra* note 718.
[722] *Id.*; Woollacott, *supra* note 718.

predict recidivism by assessing the criminal history file and interview results of the offenders.[723] The methods of how the COMPAS functions and estimates the risk of recidivism are non-disclosed by the developers of the COMPAS due to being a trade secret.[724] Supreme Court of Wisconsin in the United States had presented an eye-catching case concerning algorithmic risk assessment tools. Eric L. Loomis was sentenced as the court used COMPAS in his trial.

Claiming that his right to due process was violated because of using COMPAS at sentencing, Mr. Loomis appealed the circuit court's denial of a resentencing hearing.[725] The Appeal Court had invited the Supreme Court of Wisconsin to certify the questions of whether employing COMPAS when deciding on sentencing violates due process rights, and if it does, is it because of the proprietary nature of the COMPAS or its consideration of gender in its assessment.[726] The Supreme Court of Wisconsin had established that the appropriate and careful use of the COMPAS at sentencing by the courts does not violate the due process rights.[727] Because, the Supreme Court of Wisconsin explained, the circuit court used COMPAS with other independent factors rather than the exclusive decision maker.[728] COMPAS is a notable example due to its discriminatory risks. A prominent study showed that COMPAS was generating biased outcomes against black defendants due to correlating race and risk score, an example of redundant encodings.[729]

Another use of artificial intelligence in criminal justice criticized for being discriminatory is predictive policing using facial recognition with supervised learning.[730] Moving on from a historical view that a face reveals what is inside, researchers Wu and Zhang have analyzed whether artificial intelligence applications disclose a person's criminality.[731] Unlike any human-made analysis for criminality, the authors have argued that computers will not have biases or human deficiencies when measuring facial features

[723] State v. Loomis - 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749, at 754; State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing, *supra* note 604 at 1531.
[724] State v. Loomis - 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749, *supra* note 723 at at 761; State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing, *supra* note 604 at 1531.
[725] State v. Loomis - 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749, *supra* note 723 at at 753.
[726] *Id.*
[727] *Id.*
[728] *Id.* at at 753.
[729] Angwin et al., *supra* note 4; Schönberger, *supra* note 210 at 178–179.
[730] Završnik, *supra* note 598 at 571; Xiaolin Wu & Xi Zhang, *Automated Inference on Criminality Using Face Images*, 1 (2016), https://arxiv.org/pdf/1611.04135v2.pdf (last visited Feb 21, 2023).
[731] Wu and Zhang, *supra* note 730 at 1.

for one's criminality.[732] This study documented three main facial features as indicators of criminality: the distance between the inner corners of two eyes, upper lip curvature, and the angle from the nose tip to mouth.[733] At the end of their study, Wu and Zhang noticed that law-abiding citizens have more similar faces and criminal faces have more dissimilarity, even when controlled for race, gender, and age, which according to the authors, implied a "law of normality" for non-criminal faces.[734] The authors have expressed that data-driven face classifiers provided trustworthy indications of criminality.[735] Critics of this study said that Wu and Zhang's study reminds of Italian doctor criminologist Cesare Lombroso's criminality theory involving psychological and physical features and physiognomics, which was later found to be racist and inaccurate.[736] The study was denounced on three grounds due to the dataset used. According to the critics, the photos of criminals were regular ID photos, and those of non-criminals were purposely chosen to look nice.[737] Secondly, the application may have learned to choose the more likely to be convicted photos instead of measuring actual 'criminality.'[738] Thirdly, the first group of photos had smiling faces, whereas the second group had frowning faces which could have affected the measurement's quality.[739]

The authors have replied to criticisms of their paper suggesting that they had not presented their study as a tool for law enforcement and denied any implications of racism.[740] The authors have also denied the claims of the "garbage-in garbage-out" case in their study and argued that just because there is a risk of social biases from humans entering the computer world, it doesn't mean that social computing should exclude machine learning.[741] Further, the authors have contended that these applications may help to identify and fix human biases other than reinforcing them.[742] In return for the risk of the feedback loop, the researchers argued that it could be a bad thing or a good thing.[743]

---

[732] *Id.* at 2.
[733] Završnik, *supra* note 598 at 571; Wu and Zhang, *supra* note 730 at 6.
[734] Wu and Zhang, *supra* note 730 at 2, 10.
[735] *Id.* at 10.
[736] Carl Bergstrom & Jevin West, *Case Study Criminal Machine Learning*, CALLING BULLSHIT, 2017, https://www.callingbullshit.org/case_studies/case_study_criminal_machine_learning.html?fbclid=.
[737] *Id.*
[738] *Id.*
[739] *Id.*
[740] Xiaolin Wu & Xi Zhang, *Responses to Critiques on Machine Learning of Criminality Perceptions (Addendum of ArXiv:1611.04135)*, 1, 1 (2017), http://arxiv.org/abs/1611.04135.
[741] *Id.* at 2.
[742] *Id.*
[743] *Id.*

Regarding their choice of word criminality, they have accepted that it should be cautiously used, as even a court decision may not always equal someone's commission of a crime.[744] Yet still, they have argued that humans may seek the objectivity of the training data for criminality.[745] Regarding the claims of smiling and frown faces, the researchers suggested that the application assessed not the smiles but facial relaxation.[746] In addition, they have stated that the photos from the criminals were not mugshots but were ID photos, thus not intended to depict the criminals badly.[747] Nonetheless, they have accepted failing to use white-collar photos of non-criminals.[748] The study was criticized even following the authors' responses, at least because it made the facial feature analysis for criminality possible to be used in the criminal justice system.[749] If states adopt such an artificial intelligence tool for criminality assessment, it could have hazardous impact on human rights, violating both the right to fair trial and the right to equality and non-discrimination.

Another artificial intelligence use in criminal justice is facial recognition tools that could bring about discrimination on the grounds of race and gender.[750] Working on the Fitzpatrick skin type labeling system, Buolamwini and Gebru made an intersectional examination of ethnicity and gender disparities in commercial gender classifiers of Microsoft, IBM, and Face++.[751] This study demonstrated that the performance of all classifiers functioned better in male faces over female faces and lighter faces over darker faces.[752] The most deficient performance was for darker female faces for all of the classifiers, with over 34% error rate; two of the classifiers performed best for lighter male skins, and one performed best for dark male skins due to disparity in the sample size.[753] The commercial facial recognition systems were trained predominantly on white male faces, causing representation bias and thus recognizing white faces better.[754] Buolamwini and Gebru brought performance metrics reporting into the algorithmic fairness discussion

[744] *Id.* at 1.
[745] *Id.* at 2.
[746] Bergstrom and West, *supra* note 736.
[747] *Id.*
[748] *Id.*
[749] Završnik, *supra* note 598 at 571.
[750] Buolamwini and Gebru, *supra* note 21 at 3.
[751] *Id.* at 6–8.
[752] *Id.* at 8.
[753] *Id.*; Schönberger, *supra* note 210 at 178.
[754] Tanna and Dunning, *supra* note 128 at 426.

with their study.[755] This study presented an example of intersectional discrimination under artificial intelligence. As early as 2009, facial recognition tools on cameras displayed biased actions towards Asians by not recognizing the eye structure and saying to "open their eyes," newer examples of the same kind exist for other facial recognition tools such as for passport photos.[756] Besides, false positive rates were higher in a study conducted on vendors of face recognition tools for East Asians and West and East Africans, whereas they were lowest for Eastern Europeans; in contrast, in some algorithms developed in China, the same study found lower false positive rates.[757]

Moreover, predictive policing helps foresee potential crimes, prevent them from transpiring, and find their hotspots.[758] The quality of predictive policing tools, like other machine learning tools, depends mainly on the quality of the training data.[759] These tools may bring about discriminatory results for poor, historically discriminated, and minority communities, for example, non-white people in the United States.[760] In contrast, these tools may result in bypassing the crimes in wealthy, white-dominated, and privileged communities and neighborhoods.[761] A study by Lum and Isaac showed disproportionality in predictive policing of drug crimes in areas that were historically exposed to more policing in the USA.[762] The "garbage-in garbage-out principle" applies here; if biased data trains the machine, the machine will produce biased results.[763] If the police continue to focus more on these neighborhoods with predictive policing activities, more of the crime there will be revealed due to being in focus; thus, the criminal data will again grow for that neighborhood over and over again, resulting in a feedback loop.[764] Here the

---

[755] Buolamwini and Gebru, *supra* note 21 at 12.

[756] ZUIDERVEEN BORGESIUS, *supra* note 19 at 29; Kate Crawford, *Artificial Intelligence's White Guy Problem - The New York Times*, THE NEW YORK TIMES, Jun. 25, 2016, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html; Gwen Sharp, *Nikon Camera Says Asians: People Are Always Blinking - Sociological Images*, (May 29, 2009), https://thesocietypages.org/socimages/2009/05/29/nikon-camera-says-asians-are-always-blinking/ (last visited Apr 8, 2023); Odelia Lee, *Camera Misses the Mark on Racial Sensitivity*, GIZMODO (2009), https://gizmodo.com/camera-misses-the-mark-on-racial-sensitivity-5256650 (last visited Apr 8, 2023).

[757] PATRICK GROTHER, MEI NGAN & KAYEE HANAOKA, *Face Recognition Vendor Test Part 3: Demographic Effects*, NIST IR 8280 2, 10 (2019), https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf (last visited Apr 8, 2023).

[758] Crawford, *supra* note 756.

[759] *Id.*

[760] *Id.*

[761] *Id.*

[762] Lum and Isaac, *supra* note 609 at 19.

[763] *Id.* at 15; ZUIDERVEEN BORGESIUS, *supra* note 19 at 19.

[764] Lum and Isaac, *supra* note 609 at 16; ZUIDERVEEN BORGESIUS, *supra* note 19 at 19.

question may come to mind, if there is more crime in those regions, then is it not at least natural that there are more records from these areas? While this question might have a logical approach, the problem is overlooking the crimes happening in other neighborhoods. Furthermore, the main reason for collecting more criminal data from these neighborhoods is sociological and historical bias.

In addition, Google offered advertisements to individuals with names resembling African descent that contained the word "arrest" in its wording, whereas white resembling names received fewer such advertisements, indicating a discrimination by proxy.[765] That is a case of historical racial bias entering into Google's artificial intelligence based advertisement system through other individuals' internet surfing history.[766]

If the artificial intelligence system learns that people in one postal code previously paid the loans they took, the artificial intelligence systems could learn that people living in that postal code area pays their loans.[767] The opposite could be valid for other postal codes where the loan repayment rates are low.[768] There could be a correlation between the residence areas, postal codes, and certain ethnic or racial backgrounds as redundant encodings.[769] If an artificial intelligence-based banking application denies credit applications from postal codes with a low repayment rate, there could be hidden racial discrimination due to the connection between the postal codes and racial backgrounds.[770]

Another example of historical racial bias reflecting the existing discrimination in the society entering into artificial intelligence is that Google's image search, when put in the query for "three black teenagers" showed results which most of them were mugshot photos of three black teenagers; and when "three white teenagers" put in, the results were mostly happy teenager photos, following the criticisms, Google declared the search results reflected society, and merely showing these content to users is not an individual case of discrimination.[771] These image search results may bring about discrimination and

---

[765] Latanya Sweeney, *Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising*, 11 QUEUE 10, 13 (2013), https://dl.acm.org/doi/10.1145/2460276.2460278 (last visited Mar 21, 2023); ZUİDERVEEN BORGESİUS, *supra* note 19 at 26; Hacker, *supra* note 19 at 1149.
[766] ZUIDERVEEN BORGESIUS, *supra* note 19 at 26.
[767] *Id.* at 21.
[768] *Id.*
[769] *Id.*
[770] *Id.*
[771] *Id.* at 28–29; Chris York, *Type "Three Black Teenagers" Into Google And Something Racist Happens. Or Does It?*, HUFFPOST UK, Jun. 8, 2016, https://www.huffingtonpost.co.uk/entry/three-black-teenagers-google-racism_uk_575811f5e4b014b4f2530bb5 (last visited Apr 6, 2023); Antoine Allen, *The 'Three*

influence people's opinions.[772] It should be noted that when the same search query is written in Google's image search box at the time of this thesis, the first four photos that come up are those from the mentioned study.

A similar famous example is Google's labeling of black people as gorillas in its Google Photos app, which provides photo storage to users and, with machine learning, automatically groups the photos, compelling Google to apologize for the incident.[773] The less famous aspect of this is Google's steps following the incident. A study conducted three years after the incident revealed that Google removed the word gorilla from the system, censoring the wording completely and also blocking chimp, chimpanzee, and monkey.[774] Google has also removed the words "black men," "black women," and "black people" as labelings and fetched photos of people in black and white color instead of darker skin individuals when these queries were put.[775] The subsequent study demonstrated that Google had opted for an easy solution by blocking these labelings, which may work for preventing discriminatory outcomes but raises doubts regarding its method.[776] A similar illustration of historical and systemic inequalities entrance to artificial intelligence based systems was Amazon's same-day delivery service not being available for "black" neighborhoods; the non-delivery areas were compared to mortgage "redlining" areas in the 20th century.[777]

2.1.4 Language

Online translation tools are widely used yet may carry bias into the translation. A study conducted in 2017 on Google Translate showed that translations had a gender bias

Black Teenagers' Search Shows It Is Society, Not Google, That Is Racist, THE GUARDIAN, Jun. 10, 2016, https://www.theguardian.com/commentisfree/2016/jun/10/three-black-teenagers-google-racist-tweet (last visited Apr 6, 2023); Schönberger, *supra* note 210 at 178.

[772] ZUIDERVEEN BORGESIUS, *supra* note 19 at 29.

[773] Google apologises for Photos app's racist blunder, BBC NEWS, Jul. 1, 2015, https://www.bbc.com/news/technology-33347866 (last visited Apr 6, 2023); Tom Simonite, *When It Comes to Gorillas, Google Photos Remains Blind*, WIRED, 2018, https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/ (last visited Apr 6, 2023); James Vincent, *Google 'Fixed' Its Racist Algorithm by Removing Gorillas from Its Image-Labeling Tech*, THE VERGE (2018), https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai (last visited Apr 6, 2023); ZUİDERVEEN BORGESİUS, *supra* note 19 at 29; Crawford, *supra* note 756.

[774] Simonite, *supra* note 773.

[775] *Id.*

[776] Vincent, *supra* note 773.

[777] Crawford, *supra* note 756.

problem.[778] This could also convey as language-based discrimination. The study exposed that Google translated queries written in Turkish, which is a gender-neutral language and uses only the adjective "o" for both "he" and "she," meaning that if a person mentions someone's occupation as "O bir doktor, O bir hemşire," this person may represent a female or a male doctor and nurse, "O" for he/she; "bir" for a; "doktor" for the doctor, "hemşire" for the nurse.[779] Yet, when this sentence is written on Google Translate, Google provided translations as "He is a doctor" and "She is a nurse".[780] Years later, Google now offers two different options for English and gives an information page with gender-specific translations.[781] Google explains that the gender-specific translations are not applicable to all languages yet, and even offers why which gender label is shown before the other one by explaining that feminine comes before masculine alphabetically.[782]

2.1.5 Religion

Religion-based discrimination may happen if artificial intelligence is used to deepen religious tensions, target religious minorities, or reinforce anti-blasphemy laws.[783] One way artificial intelligence can cause discrimination concerning religious manifestation is through content recommendation.[784]

Religion-based discrimination is closely related to freedom of religion and its different aspects, including not manifesting one's beliefs.[785] Advertisers and other data users may learn about individuals' religious beliefs through artificial intelligence, which thus enables suggestions in online systems related to one's beliefs.[786] This situation could impact religious minorities the most, which could be believers of a religion in a state where another religion or non-religiousness has the majority, such as Muslim believers in

---

[778] Aylin Çalışkan, Joanna J. Bryson & Arvind Narayanan, *Semantics Derived Automatically from Language Corpora Contain Human-like Biases - Supplementary Material*, 356 SCIENCE 183 (2017), http://arxiv.org/abs/1608.07187 (last visited Mar 27, 2022).
[779] *Id.* at 3.
[780] *Id.*
[781] Get gender-specific translations - Google Translate Help, https://support.google.com/translate/answer/9179237?hl=en&ref_topic=7011755&sjid=14977587573544606156-EU (last visited Apr 7, 2023).
[782] *Id.*
[783] Ashraf, *supra* note 659 at 771.
[784] *Id.* at 772.
[785] *Id.*; HEINER BIELEFELDT, UN HUMAN RIGHTS COUNCIL SPECIAL RAPPORTEUR ON FREEDOM OF RELIGION OR BELIEF, & UN HUMAN RIGHTS COUNCIL SECRETARIAT, *Report of the Special Rapporteur on Freedom of Religion or Belief*, 11 (2015), https://undocs.org/A/HRC/31/18.
[786] Ashraf, *supra* note 659 at 772.

a European state, or non-religious individuals, including agnostics and atheists in a state where religiousness levels are high.[787] That is also related to religious profiling, a problem that exists independently from artificial intelligence but may deepen through it, leading to stigmatization, stereotypes, exclusion, and provocation.[788] Algorithmic assumptions may reveal a person's beliefs to increase user interaction when they choose not to share them online.[789]

Artificial intelligence based content moderation may influence online religious manifestation even before being publicized.[790] Freedom of religion may be exercised alone as well as in community with others; thus, for example, if artificial intelligence based content moderation ultimately not allows a community page on a platform, a believer from a minority religious group could end up not being able to share a community even in the online world; which may be discriminatory.[791]

UN Special Rapporteur on Freedom of Religion or Belief Ahmed Shaheed emphasized in his report that online content moderation tools might risk reinforcing societal prejudices against religious minorities, stigmatization, and marginalization, resulting in the over-policing of different faith communities and targeting of individuals or communities.[792]

2.1.6 Political or Other Opinion

Automated content moderation may raise questions, such as whether social media platforms should have the right to remove a post even if the content is legal.[793] Deleting social media posts and accounts is possible with automated content moderation. In addition to the unwanted possibility of state censorship on social media, automated content moderation by social media giants may also drive them to become censors, as they can control and ban content on their platforms through direct intervention. Expecting control and management over social media platforms is necessary from the social media

[787] *Id.*
[788] BIELEFELDT, UN HUMAN RIGHTS COUNCIL SPECIAL RAPPORTEUR ON FREEDOM OF RELIGION OR BELIEF, AND UN HUMAN RIGHTS COUNCIL SECRETARIAT, *supra* note 785 at 11.
[789] Ashraf, *supra* note 659 at 772.
[790] *Id.* at 773.
[791] *Id.*
[792] *Id.*; AHMED SHAHEED, UN HUMAN RIGHTS COUNCIL SPECIAL RAPPORTEUR ON FREEDOM OF RELIGION OR BELIEF, & UN HUMAN RIGHTS COUNCIL SECRETARIAT, *Report of the Special Rapporteur on Freedom of Religion or Belief*, 15 (2019), https://undocs.org/A/HRC/40/58.
[793] PASQUALE, *supra* note 27 at 5.

giants considering the millions of shares on these platforms per second. However, it is vital to determine the provision of content moderation, especially when it is automated and left to artificial intelligence. In this regard, freedom of expression is closely related to the right to equality and non-discrimination.[794] For example, a post with derogatory language towards a community is labeled as racist speech, while no protective action is taken on similar content directed at another community. Furthermore, this can come to the fore regarding language recognition. Databases in some languages may be more advanced for recognizing discriminatory language, while others may be less advanced. Here too, more protection will be provided in one language but less protection in another.

A striking example of automated content moderation can be seen in the due diligence report conducted for Facebook's parent company Meta's response to the incidents in Palestine and Israel in May 2021, that identified Arabic-language posts were more likely to be "over-enforced" by Meta compared to Hebrew posts and their potential to be considered as a violation of Meta's policies was higher than Hebrew ones, which may be due to Meta having "Arabic hostile speech classifiers" but not Hebrew ones.[795] The report also found a problem of "under-enforcement" both on the Palestinian and Israeli sides, in cases such as incitement to violence against Israelis.[796] The Hebrew under-enforcement cases mainly resulted from a deficiency of Hebrew classifiers for categorizing and identifying contents that violate Meta's policies.[797] The report displayed that "over-enforcement," leading to removing posts for violation of content policies, had impacted users' visibility and engagement on Meta's platforms.[798] For example, the hashtag "#AlAqsa" was included in the hashtag block list by a human employee from the list of terms by the United States Treasury Department concerning Al Aqsa Brigade, whereas the hashtag #AlAqsa indicated Al Aqsa Mosque as one of the holiest sites in Islam.[799] In this case, the human employee had carried bias into the hashtag block list. The report further disclosed unintentional bias on Meta's side in the May 2021 events because of

---

[794] See SPECIAL RAPPORTEUR ON FREEDOM OF OPINION AND EXPRESSION, *Report on Artificial Intelligence Technologies and Implications for Freedom of Expression and the Information Environment*, (2018), https://undocs.org/Home/Mobile?FinalSymbol=A%2F73%2F348&Language=E&DeviceType=Desktop&LangRequested=False.
[795] BUSINESS FOR SOCIAL RESPONSIBILITY (BSR), *Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021 Insights and Recommendations*, 10 5 (2022).
[796] *Id.*
[797] *Id.*
[798] *Id.*
[799] *Id.* at 6.

more over-enforcement of Arabic content than Hebrew content.[800] Likewise, Arabic classifiers worked less well for the content in the Palestinian Arabic dialect because it is less common, and the training data sourced from human assessments lack linguistic and cultural competency.[801] On the other hand, there were fewer reviewer errors regarding language, as Hebrew is mainly spoken in Israel and is more standardized.[802] This illustration presents a case about the right to freedom of expression and discrimination based on political opinion where both human and automated content moderation was used.[803] This conveys that political opinion-based discrimination may be alleged in conjunction with freedom of expression.

2.1.7 National or Social Origin

Discrimination through artificial intelligence on the grounds of national or social origin may appear most likely in the employment sector or about the distribution of social or financial benefits. Algorithms may consider national or social origin at least as a redundant encoding similar to other cases. It is also highly likely that this ground to come up as an example of intersectional discrimination. In the Netherlands, an artificial intelligence system called SyRI had made faulty decisions regarding the risk of fraud, and the Court of the Hague considered that it interfered with individuals' private life besides lacking transparency and having the potential to be discriminatory towards individuals from lower socio-economic immigration backgrounds.[804]

2.1.8 Property

What could come up with discrimination through artificial intelligence on the ground of property could relate to social aid or pension payments. The examples mentioned about the mortgage payments can also fall in the scope of this ground. In the framework of the

---

[800] *Id.* at 8.
[801] *Id.*
[802] *Id.*
[803] *Id.* at 2.
[804] CHRISTOPHE LACROIX, *supra* note 13 at 9; ROBIN ALLEN QC & DEE MASTERS, *Regulating for An Equal AI: A New Role For Equality Bodies Meeting the New Challenges to Equality and Non-Discrimination from Increased Digitisation and the Use of Artificial Intelligence*, 119 107–109 (2020), https://equineteurope.org/equinet-report-regulating-for-an-equal-ai-a-new-role-for-equality-bodies/ (last visited Aug 1, 2022); ECLI:NL:RBDHA:2020:1878, Rechtbank Den Haag, C-09-550982-HA ZA 18-388 (English), paras. 6.86, 6.93 and 6.106 (2020), https://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:1878.

ECHR, the right to the protection of property concerns the possessions one already has ownership of, or at least requires a legitimate expectation for the protection of the right.[805] The examples of not being provided with social payments through artificial intelligence tools used by the authorities when there is no legitimate expectation would thus not fall under the scope of this discrimination ground. An example that could lead to discrimination based on property relates to welfare benefits, such as the universal credit system in the United Kingdom, which operates on algorithms to calculate benefits based on data obtained from employers and public authorities, making automated decisions on welfare benefits.[806] A court in the UK decided on discrimination due to automatically moving disabled individuals and causing them to receive lesser payments.[807]

2.1.9 Birth Status

Discrimination based on birth status means discriminating due to being born outside of marriage.[808] Birth status-based discrimination is rare in the 2020s, especially in the CoE member states.[809] The ECtHR considers the prohibition of discrimination based on being born out of wedlock as a *"standard protection of European public order."*[810] While it is possible, it would be unlikely if an artificial intelligence system were to cause discrimination based on being born out of wedlock as it is less likely to be trained on such biased data, as there were fewer examples of this discrimination concerning discrimination based on this status.

2.1.10 Other Status

Discrimination grounds are not limited in numbers with the ones that are explained above. Sexual orientation, disability, age are such grounds that fall under "other"

---

[805] Marckx v. Belgium, A31 Eur. Court Hum. Rights Rep. Judgm. Decis. ¶ 50 (1979); Pine Valley Developments Ltd and Others v. Ireland, A222 Eur. Court Hum. Rights Rep. Judgm. Decis. ¶ 51 (1991).
[806] CHRISTOPHE LACROIX, *supra* note 13 at 9–10; Robert Booth, *Automated UK Welfare System Needs More Human Contact, Ministers Warned*, THE GUARDIAN, May 22, 2023, https://www.theguardian.com/society/2023/may/22/automated-uk-welfare-system-needs-more-human-contact-ministers-warned (last visited Jun 8, 2023).
[807] Patrick Butler, *UK Government Faces £150m Bill over Social Welfare Discrimination*, THE GUARDIAN, Jan. 21, 2022, https://www.theguardian.com/society/2022/jan/21/uk-government-bill-welfare-discrimination-universal-credit (last visited Jun 12, 2023).
[808] Fabris v. France, Reports of Judgments and Decisions 2013 Eur. Court Hum. Rights Rep. Judgm. Decis. ¶ 58 (2013).
[809] *Id.*
[810] *Id.* at ¶ 57.

category. Dutch Data Protection Authority established that Facebook showed advertisements based on sensitive characteristics.[811] Facebook allowed advertisers to choose "men interested in men" for targeted advertising as a category, although Dutch Data Protection Act prohibits processing special categories of personal data for advertisement purposes.[812]

A legal case was filed in the U.S. based on the Fair Housing Act that forbids housing advertisements based on statuses such as religion, racial background, disability, familial status or national origin.[813] According to the case's claim, Facebook's advertisement platform allowed real estate agents and landlords to deny families with children and women, among others, from seeing rental and sales ads for housing.[814] The National Fair Housing Alliance, one of the applicants of the case, was able to exclude "corporate moms," "stay-at-home moms," and moms of grade school kids" and choose to include "no kids" or "men," which adjusted the number of people that the ad is going to be shown when publishing an advertisement.[815] The case was settled in 2019, with Facebook's promise of paying damages, regular meetings on the issue, and changing the advertisement system to restrain the advertisers from targeting individuals based on protected characteristics and designating a separate portal for housing, employment, and credit-related advertisements.[816] There is another similar case that is again, resolved by settlement filed by the U.S. Department of Housing and Urban Development on the grounds that the advertisers were able to discriminate on Facebook through screening of

[811] ZUIDERVEEN BORGESIUS, *supra* note 19 at 27; THE DUTCH DATA PROTECTION AUTHORITY (AUTORITEIT PERSOONSGEGEVENS), *Conclusions Facebook 2017*, 4 3 (2017), https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/conclusions_facebook_february_23_2017.pdf (last visited Mar 21, 2023).
[812] THE DUTCH DATA PROTECTION AUTHORITY (AUTORITEIT PERSOONSGEGEVENS), *supra* note 811 at 3.
[813] Charles V. Bagli, *Facebook Vowed to End Discriminatory Housing Ads. Suit Says It Didn't.*, THE NEW YORK TIMES, Mar. 27, 2018, https://www.nytimes.com/2018/03/27/nyregion/facebook-housing-ads-discrimination-lawsuit.html (last visited Mar 21, 2023); ZUIDERVEEN BORGESIUS, *supra* note 19 at 27; National Fair Housing Alliance; Fair Housing Justice Center, Inc.; Housing Opportunities Project For Excellence, Inc.; Fair Housing Council Of Greater San Antonio v. Facebook, Inc., (2019), https://nationalfairhousing.org/wp-content/uploads/2022/01/FINAL-SIGNED-NFHA-FB-Settlement-Agreement-00368652x9CCC2.pdf.
[814] Bagli, *supra* note 813.
[815] *Id.*
[816] National Fair Housing Alliance et al. v. Facebook | Civil Rights Litigation Clearinghouse, https://clearinghouse.net/case/17155/ (last visited Mar 21, 2023); Settlement Agreement | National Fair Housing Alliance Et Al. V. Facebook | Civil Rights Litigation Clearinghouse, (2019), https://clearinghouse.net/doc/102171/ (last visited Mar 21, 2023).

who can and cannot see the housing listings.[817] Following this case, Facebook's parent company Meta had promised to change its advertisement delivery system.[818]

Economic status is also under this category. A customer who lives in a less prosperous area where burglary is common may be charged higher insurance costs if s/he applies for a house insurance policy.[819] The term "other status" may even extend to being a Facebook group member regarding artificial intelligence; for example, an algorithm may tag a job candidate as unemployable due to membership in a Facebook group.[820]

In assessing and deciding on an issue, do there not have to be specific criteria and features that are qualified or not qualified by individuals? Even if no artificial intelligence is involved and humans are the sole decision-makers, isn't it natural to have some evaluation criteria? For example, when issuing a loan, banks need a set of specific criteria to estimate whether the applicant has the means to pay back the loan. A university requires students with strong academic backgrounds. When deciding on the duration of the charges, a judge looks at the facts in the case file and the offender's previous criminal history. State officials require specific criteria to distribute social benefits to those in society who are in real need. Even if no artificial intelligence is involved and humans are the sole decision-makers, it is indeed natural to have some evaluation criteria, yet the ultimate point is that decision-making should be free from discrimination. The decision-making criteria should not be put to discriminate and also should not leave room for indirect discriminatory results either. In doing so, it would be necessary to have objective, logical, and relevant criteria for the subject matter concerned. When making a decision, the basis and evaluation criteria thereof should be valid, and there is no discrimination when determining the criteria; in that example, one's gender or residence address should not be a factor in decision-making—decisions as such may, in fact, mean interference with a human right. Generally, when a human right is intervened, it does not directly mean

---

[817] Kriston Capps, *Behind HUD's Housing Discrimination Charges Against Facebook*, BLOOMBERG.COM, Mar. 28, 2019, https://www.bloomberg.com/news/articles/2019-03-28/why-hud-charged-facebook-with-discrimination (last visited Mar 21, 2023); Kurt Wagner & Chris Dolmetsch, *Meta Settles Claims That Ads Violated U.S. Fair Housing Laws*, TIME (2022), https://time.com/6189595/meta-settles-ads-fair-housing-laws/ (last visited Mar 21, 2023).

[818] U.S. Department of Justice U.S. Attorney's Office Southern District of New York, *United States Attorney Resolves Groundbreaking Suit Against Meta Platforms, Inc., Formerly Known As Facebook, To Address Discriminatory Advertising For Housing*, (2022), https://www.justice.gov/usao-sdny/pr/united-states-attorney-resolves-groundbreaking-suit-against-meta-platforms-inc-formerly (last visited Mar 21, 2023); Wagner and Dolmetsch, *supra* note 817.

[819] ZUIDERVEEN BORGESIUS, *supra* note 19.

[820] PASQUALE, *supra* note 27 at 20.

it is violated. The interference requires inspection under the methods set in human rights law. In such a possibility, determining the "validity" of an interference requires the application of international human rights law, considering all other relevant conditions exist. This is aimed to be achieved in the following section of this chapter.

2.2 Advancing International Human Rights Law in Artificial Intelligence

The examples of discrimination through artificial intelligence are not limited with the ones laid out in the chapter, there are many other cases of discriminatory outcomes happening similarly or differently. Although discrimination through artificial intelligence may arise in unusual and unanticipated manners, the right to equality and non-discrimination remains relevant.[821] Some of these examples may not always be equivalent to discrimination in the legal sense and may be found as violation if assessed by human rights mechanisms. One reason for that is the realistic aspect of the human rights. While one of the aims of human rights mechanisms is to ensure fairness, human rights mechanisms do not promise for the complete realization of fairness in the society.[822] Yet the main reason is that these not always have the necessary merits to be found as a violation of non-discrimination. Therefore, examining discrimination through artificial intelligence by taking the examples as the starting point would allow for disseminating what can constitute as discriminatory in the legal sense with artificial intelligence. Addressing discrimination through artificial intelligence with existing human rights law is possible after understanding how AI-systems work and how discriminatory results may occur, which are also necessary before prompting a specific regulation for artificial intelligence.[823]

Discrimination is indeed possible without the involvement of artificial intelligence in the examples in the previous section of this chapter. But when artificial intelligence is a component of decision-making partially or entirely, there are specific considerations along with it. First, as explained above, when artificial intelligence is adopted, the primary motivations include providing objectivity in decisions. Keeping also in mind the automation bias, the adoption of seemingly objective artificial intelligence when in reality

[821] Smuha, *supra* note 19 at 101.
[822] Cf. *Id.* at 595–597.
[823] *Id.* at 101.

it is not, for various reasons, indeed carries risks of discrimination. Secondly, artificial intelligence brings new ways and forms of discrimination, as explored throughout the chapter. Thirdly, since artificial intelligence reduces the human role in decision-making through automation, what could be identified as discriminatory when humans were the decision-makers may now not be revealed.

One way to address these issues is by resorting to the existing human rights law framework. Human rights law sets obligations for states and provides tools for identifying and evaluating the adverse impacts of artificial intelligence.[824] The framework provided by the human rights law is fitted for addressing human rights-related problems of artificial intelligence by mandating states with obligations for protecting human rights and preventing their violations, and also steering the private actors by urging them to take steps for human rights respect.[825] Handling artificial intelligence with the mechanisms and the framework of human rights law could therefore provide measures for realizing respect for human rights throughout the artificial intelligence lifecycle (AI-lifecycle).[826] It should be noted that human rights law provide more practical and effective tools for implementing and ensuring accountability compared to non-binding or ethics documents on artificial intelligence.[827] Because, governing artificial intelligence with human rights law framework allows for handling the constraints in the non-binding documents, such as settling conflicts in the development and deployment of the AI-systems and the disagreeing ethical norms. [828]

If the human rights mechanisms are integrated throughout the AI-lifecycle, examining the role and responsibilities of different actors could be possible.[829] Integrating human rights law with artificial intelligence is not without putting efforts regarding the "translation" of human rights to determine its advantages and fragilities for artificial intelligence and, if required, setting out new legislative and regulatory tools for addressing its potential fragilities such as being abstract, as well as analyzing mechanisms

[824] McGregor, Murray, and Ng, *supra* note 1 at 325.
[825] *Id.* at 329.
[826] *Id.* at 330.
[827] YEUNG, *supra* note 518 at 16; AUSTRALIAN HUMAN RIGHTS COMMISSION, *Human Rights and Technology Issues Paper*, 64 17 (2018), https://tech.humanrights.gov.au/sites/default/files/2018-07/Human%20Rights%20and%20Technology%20Issues%20Paper%20FINAL.pdf (last visited Jun 11, 2023) See Chapter 3 for the discussions regarding non-binding documents for artificial intelligence.
[828] Yeung, Howes, and Pogrebna, *supra* note 529 at 83.
[829] McGregor, Murray, and Ng, *supra* note 1 at 325.

for human rights enforcement.[830] Examining artificial intelligence under the human rights law framework is also paramount to connecting and mapping the relation and the effect of the non-artificial-intelligence-based inequalities and discriminations with artificial-intelligence-based inequalities and discrimination. The applicability of the human rights instruments, including the ECHR and the EU Fundamental Rights Charter, to artificial intelligence is mentioned by the CAHAI.[831] The CoE's report and recommendation thereunder about preventing discrimination caused by the use of artificial intelligence reminds the CoE Committee of Ministers to consider the *"serious potential impact"* of artificial intelligence on the right to equality and non-discrimination by mentioning the ECHR and European Social Charter (ESC) when evaluating the requirement of an international framework for artificial intelligence.[832]

As seen, regional human rights mechanisms such as the ECtHR are expected to be critical in addressing artificial intelligence.[833] Due to the various reasons explained at the beginning of this chapter, this study will proceed with the ECtHR when discussing artificial intelligence from the perspective of the existing human rights framework.[834] Looking at the ECtHR could help understand the potential risks and consequences of artificial intelligence regarding human rights, defining responsibilities, and allocating responsibility.[835] In addition, it will allow for answers in return to whether the human rights instruments and mechanisms are equipped enough to address human rights during the whole of AI-lifecycle as the CAHAI asks.[836] The first step in doing that is to explore how the ECtHR deals with discrimination claims and converge it to artificial intelligence.

---

[830] Smuha, *supra* note 19 at 102; See YEUNG, *supra* note 518 at 16; *See also* AUSTRALİAN HUMAN RİGHTS COMMİSSİON, *supra* note 827 at 17.
[831] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 18.
[832] Council of Europe Parliamentary Assembly, *Recommendation 2183 (2020) of the Parliamentary Assembly on Preventing Discrimination Caused by the Use of Artificial Intelligence*, Recommendation 2183 (2020) (2020), https://pace.coe.int/pdf/1cf0f22ba28c172524b4e7261f78f2d45c508160b6c058ec2b2d2fadfca9c3bc/rec.%202183.pdf (last visited Aug 4, 2022).
[833] McGregor, Murray, and Ng, *supra* note 1 at 343; *See also* AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517 at 18.
[834] Hacker, *supra* note 19; Xenidis, *supra* note 19; Xenidis and Senden, *supra* note 19; GERARDS AND XENIDIS, *supra* note 19; See for a comparative study on non-discrimination law in the EU and the CoE. EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS., EUROPEAN COURT OF HUMAN RIGHTS., AND COUNCIL OF EUROPE (STRASBOURG)., *supra* note 447.
[835] YEUNG, *supra* note 518 at 16; AUSTRALIAN HUMAN RIGHTS COMMISSION, *supra* note 827 at 17.
[836] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 18.

2.2.1 Bridging Human Rights Law and Artificial Intelligence

The current international human rights law framework does not extend to artificial intelligence, as the phenomenon was not an issue during the inception of human rights conventions.[837] The "living instrument" principle allows human rights monitoring mechanisms to adapt themselves to changing conditions by bringing flexibility and supporting stability.[838] Artificial intelligence requires human rights documents to be interpreted according to present-day conditions.[839]

The first chapter mentioned that Article 14 of the ECHR regulates the prohibition of discrimination, and claims of Article 14 may be brought to the ECtHR concerning another right in the Convention, but also if a state had accepted a broader protection or a specific measure that falls in the general scope of any article in the ECHR, then the protection of Article 14 also expands to this broader scope.[840] This means that if a state adopts artificial intelligence related measures that fall under the general scope of the substantive articles, such as adopting a rule on artificial intelligence based social benefits distribution, then the measures related to the substantive article in the claim will be expected to be fulfilled when examining discrimination. The application of Article 1 of Protocol no. 12 to the ECHR that facilitated general prohibition of discrimination has the chance to develop the scope of application for the protection from discrimination as it applies to any right granted explicitly to an individual under national law, yet the protocol has a minimal number of ratifications as mentioned in the first chapter.[841] Suppose a ratified party to Protocol No. 12 accepts artificial intelligence related rights under its national law or international law and incorporates them into its national law. In that case, individuals may bring claims of discrimination based on any such rights to the ECtHR. In contrast, individuals would also require to show that their claims fell under a general scope of a

---

[837] See *Id.* at 18.
[838] Tyrer v. the United Kingdom, A26 ¶ 31 (1978); *See* KATE JONES, *AI governance and human rights: Resetting the relationship*, 20 (2023), https://chathamhouse.soutron.net/Portal/Public/en-GB/RecordView/Index/191961 (last visited Apr 21, 2023).
[839] *See* A26 ¶ 31 (1978); KATE JONES, *AI governance and human rights: Resetting the relationship*, 20 (2023), https://chathamhouse.soutron.net/Portal/Public/en-GB/RecordView/Index/191961 (last visited Apr 21, 2023).
[840] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 9; E.B. v. France, Eur. Court Hum. Rights Rep. Judgm. Decis. ¶ 48 (2008); Fábián v. Hungary, Rep. Judgm. Decis. 2017 Extr. ¶ 112 (2017), https://hudoc.echr.coe.int/eng?i=001-176769.
[841] COUNCIL OF EUROPE, *Explanatory Report to the Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms*, 7 5 (2000), https://rm.coe.int/09000016800cce48.

substantive article for bringing claims under Article 14. If individuals had the opportunity to bring discriminatory claims regarding any right, it could help to achieve better protection against discrimination through artificial intelligence, which could transpire in various ways that may not be confined within a substantive right listed in the ECHR.

While one would have to bring a discrimination claim based on artificial intelligence in conjunction with a substantive article, s/he will not be obliged to prove a violation of the relevant right. Because there is no need for the Court to find a violation of the right under a substantive article to find a violation of the prohibition of discrimination.[842]

Expanding Positive Obligations for Artificial Intelligence

As known, human rights instruments impose obligations to states. Considering most artificial intelligence applications are at the hands of private actors, one may wonder whether non-discrimination obligations apply to private individuals. The "horizontal effects doctrine," that extends the application of human rights law to relations between non-state actors beyond the relations between states and non-state actors, which represent a 'vertical' application of human rights law, requires the answer "yes."[843] The human rights mechanisms examine discrimination claims that may emerge from private contracts, documents, or relations.[844] This conveys that the prohibition of discrimination applies to private spheres, and states may be found responsible even when an artificial intelligence tool entirely used in private relations results in discrimination.[845]

That also connects to the potential positive obligations specific to artificial intelligence. Some of these can be the adaptation of existing positive obligations to artificial intelligence, and some unique positive obligations may develop with respect to artificial intelligence. For example, states have the positive obligation to take measures to prevent discrimination and redress options in case of a discrimination claim.[846] Artificial intelligence may require specific measures to prevent discrimination and redress mechanisms due to its nature. Not taking a step on artificial intelligence that is known for having risks to equality and the potential to discriminate may violate the positive

[842] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 13; Marckx v. Belgium, *supra* note 805 at ¶ 64-65.
[843] Lottie Lane, *The Horizontal Effect of International Human Rights Law in Practice: A Comparative Analysis of the General Comments and Jurisprudence of Selected United Nations Human Rights Treaty Monitoring Bodies*, 5 EUR. J. COMP. LAW GOV. 5, 15 (2018), https://brill.com/view/journals/ejcl/5/1/article-p5_5.xml (last visited May 10, 2023).
[844] Pla and Puncernau v. Andorra, Rep. Judgm. Decis. 2004-VIII ¶ 59 (2004).
[845] *See* Smuha, *supra* note 19 at 98.
[846] Danilenkov and Others v. Russia, Rep. Judgm. Decis. 2009 Extr. ¶ 124 (2009).

obligations of a state for the prevention of discrimination. Therefore, applying the existing human rights mechanisms requires regulating artificial intelligence as a positive obligation for states.

The CoE Commissioner for Human Rights underlines that member states of the ECHR should apply measures required to protect against human rights violations by artificial intelligence actors during the lifecycle of artificial intelligence systems to fulfill their positive and procedural obligations under the convention.[847] That implies the CoE may consider such measures as positive and procedural obligations for artificial intelligence situations. Provisions under the CoE Resolution on Preventing Discrimination caused by the use of artificial intelligence also support the said understanding.[848] Again, taking measures to address the impacts of AI-systems through practical preventive or mitigating actions is suggested for states to ensure respect for human rights by artificial intelligence actors.[849]

Another positive obligation that may come with artificial intelligence is the necessity for effective and adequate investigation. This requirement under Article 14 obliges the state authorities to uncover whether the incident had discriminatory motives or hatred and prejudice based on personal characteristics.[850] For an effective and adequate investigation of a discrimination claim, states should examine whether discrimination grounds were involved in artificial intelligence based decision-making. This would call for principles such as transparency and accountability in artificial intelligence systems, which will be explored more in detail in the third chapter.

On the other hand, different understandings of equality may be applied concerning discrimination claims. In connection, positive obligations under the prohibition of discrimination may extend to addressing the inequalities in society by taking action to correct them, and failure to do so may lead to violation.[851] This positive obligation may mainly come up about historical and societal bias in artificial intelligence systems because, as the examples throughout the chapter have shown, many of the discriminatory

[847] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 9.
[848] See Council of Europe Parliamentary Assembly, *Resolution 2343 (2020) of the Parliamentary Assembly on Preventing Discrimination Caused by the Use of Artificial Intelligence*, Resolution 2343 (2020) (2020), https://pace.coe.int/en/files/28807/pdf (last visited Aug 4, 2022).
[849] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 9.
[850] Abdu v. Bulgaria, ¶ 44 (2014).
[851] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 10 of the "Law" part; D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 175.

impacts of artificial intelligence systems arise from historical and societal bias. Considering the critical risks of historical and societal bias in artificial intelligence to reinforce and reproduce societal inequalities, not taking measures to treat these may violate this positive obligation.

Uncovering Discrimination Through Artificial Intelligence: The Discrimination Test

As seen from the examples explored throughout the chapter, discrimination through artificial intelligence tends to coincide on more than one discrimination ground, such as race and gender, health and race, political opinion, and ethnic origin. The application of multiple discrimination grounds simultaneously by the ECtHR when examining a discrimination claim is possible.[852]

Direct and indirect discrimination, as well as discrimination by association[853], should be considered in the objective and development of the AI-systems.[854] The CoE Resolution on Preventing discrimination caused by the use of artificial intelligence refers to discrimination grounds *"including sex, gender, age, national or ethnic origin, color, language, religious convictions, sexual orientation, gender identity, sex characteristics, social origin, civil status, disability or health status",* suggesting that the framework of discrimination grounds under internal human rights law -with more specified grounds- will apply to handling discrimination through artificial intelligence.[855]

On how the ECtHR handles the applications for discrimination claims, the Court only considers different treatments lacking an objective and reasonable justification discriminatory, meaning that not all different treatment signifies discrimination.[856]

The Court applies a two-step test to determine whether there is discrimination. The first step is establishing whether there is a difference in the treatment of persons in the same or similar situation, followed by the second step establishing whether the different

---

[852] See on on the grounds of gender, race and social origin B.S. v. Spain, (2012); on gender and religion S.A.S v. France, Rep. Judgm. Decis. 2014 (2014).

[853] Discrimination by association occurs when the discriminated individual is discriminated against not only based on having a "discrimination ground" but because of having a relation to a group or an individual having the "discrimination ground." EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS., EUROPEAN COURT OF HUMAN RIGHTS., AND COUNCIL OF EUROPE (STRASBOURG)., *supra* note 447 at 51; Wachter, *supra* note 661 at 373.

[854] Council of Europe Parliamentary Assembly, *supra* note 848 at para 4.

[855] *Id.*

[856] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 10 of the "Law" part; Hoogendijk v. The Netherlands, *supra* note 461 at ¶ 2 of the "Law" part; Willis v. The United Kingdom, Rep. Judgm. Decis. 2002-IV ¶ 48 (2002); D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 175; Fabris v. France, *supra* note 808 at ¶ 56; Molla Sali v. Greece, ¶ 135 (2018).

treatment is objectively justified, which involves two substeps.[857] The substeps are, first, whether the different treatment follows a legitimate aim; second, if it does, whether the employed means are proportional to the legitimate aim.[858] The discrimination test will allow examining the appropriateness of evaluation criteria set for artificial intelligence decision-making for protection from discrimination.

One needs to display s/he has been treated differently from a person or group of persons in an analogous or similar situation s/he is in – instead of demanding identicality, the court finds being in a relevant similarity enough.[859] The similarity setting may differ in different subject matters, making the context of the case at hand essential.[860] It may be possible to demonstrate being in a similar situation if two persons are subjected to different treatment through artificial intelligence in a case such as in which they are both women with different social statuses and ethnic groups and being displayed advertisements on the same products at different prices. The question here will be: is there a valid reason – a legitimate aim for being exposed to different prices?

A 'similar person' does not always need to be compared to the applicant. In such cases, there can be an instance of discrimination by association; when the 'similar' person is compared to a person that is not the applicant.[861] An example of this in artificial intelligence is a person being exposed to treatment by an artificial intelligence system and another person without an artificial intelligence system for the same process, such as being diagnosed by an artificial intelligence based healthcare tool and a healthcare professional without resorting to any use of artificial intelligence.

The artificial intelligence system will make the diagnosis based on the data it was trained on; thus, if the training data are flawed or inaccurate, such as not having enough medical data for a disease from an ethnic group, the system can result in incorrect

[857] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 10 of the "Law" part; Lithgow and Others v. The United Kingdom, A102 ¶ 177 (1986); Fabris v. France, *supra* note 808 at ¶ 56; Molla Sali v. Greece, *supra* note 856 at ¶ 135.
[858] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 10 of the "Law" part; Lithgow and Others v. The United Kingdom, *supra* note 857 at ¶ 177; Fabris v. France, *supra* note 808 at ¶ 56; Molla Sali v. Greece, *supra* note 856 at ¶ 135.
[859] Lithgow and Others v. The United Kingdom, *supra* note 857 at ¶ 177; Willis v. The United Kingdom, *supra* note 856 at ¶ 48; Zarb Adami v. Malta, *supra* note 461 at ¶ 71; D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 175; Burden v. The United Kingdom, *supra* note 434 at ¶ 60; Fabris v. France, *supra* note 808 at ¶ 56; Fábián v. Hungary, *supra* note 840 at 113; Molla Sali v. Greece, *supra* note 856 at ¶ 133-135.
[860] Fábián v. Hungary, *supra* note 840 at ¶ 121.
[861] Molla Sali v. Greece, *supra* note 856 at ¶ 134.

diagnosis for patients from that ethnic background. If a human doctor were to diagnose that same person without an artificial intelligence system, the diagnosis could have been correct. In such a case -given that other steps in the discrimination test are failed- there could be an instance of discrimination by association because the discriminatory result does not result from a personal characteristic but rather because of the patient's relation to an ethnic group that was mis- or underrepresented in the training of the healthcare system.[862] If a comparison is made with a different person from another ethnic background, and there is no incorrect diagnosis from the artificial intelligence system, there could also be an instance of indirect discrimination, again assuming that other steps in the discrimination test are failed.

The second thing the ECtHR inquires is whether there is an objective and reasonable justification for the difference in treatment.[863] The court explains this with the existence of a legitimate aim and, second, the lack of proportionality between the aim and the means employed.[864] Legitimate aims are not exhaustively listed in Article 14 of the ECHR, as this article can only be conveyed with another right in the convention, which, if a right that is interferable, lists the legitimate aims in their relevant text. In the court's case law, accepted legitimate aims for the difference in treatment include protection of national security, facilitation of rehabilitation of juvenile delinquents, protection of women against gender-based violence in the scope of Article 14 in conjunction of rights such as the right to respect for private and family life and the right to liberty and security.[865]

It is conceivable that it would not be considered reasonable by the ECtHR if a state were to give the reason that the discriminatory outcome occurred because the training data reflected fewer hospital visits of people from an ethnic origin, resulting in less collection of medical data from persons of that ethnic origin. There is also a requirement for effective safeguards to be put in place when there is interference regarding

[862] See Wachter, *supra* note 661.
[863] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 10 of the "Law" part; Weller v. Hungary, ¶ 27 (2009); Fabris v. France, *supra* note 808 at ¶ 56; Fábián v. Hungary, *supra* note 840 at ¶ 113; Molla Sali v. Greece, *supra* note 856 at ¶ 135.
[864] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 10 of the "Law" part; Weller v. Hungary, *supra* note 863 at ¶ 27; Fabris v. France, *supra* note 808 at ¶ 56; Fábián v. Hungary, *supra* note 840 at ¶ 113; Molla Sali v. Greece, *supra* note 856 at ¶ 135.
[865] Konstantin Markin v. Russia, Rep. Judgm. Decis. 2012 Extr. ¶ 147 (2012); Khamtokhu and Aksenchik v. Russia, Rep. Judgm. Decis. 2017 ¶¶ 80-82 (2017).

legitimacy.[866] While the topic will be examined in the third chapter, it should be noted that effective safeguards are required to exist in all application stages, including the design, development, and implementation, which could involve technical efforts to address discrimination through artificial intelligence.

In case of a legitimate aim exists, then the proportionality assessment to inquire about the fair balance between the different treatments and the relevant legitimate aim is made.[867]

Developing the Margin of Appreciation in Artificial Intelligence

The states have a margin of appreciation which the extent of depends on the circumstances, the subject matter, and the background of the case.[868] The margin can be wider in cases related to the public interest, where the circumstances are related to social or economic issues.[869] When the margin of appreciation is considered for intelligence based systems, using artificial intelligence systems for calculation purposes can be given a wider margin. Yet, the margin can be expected to be lesser if the artificial intelligence based systems involve personal characteristics such as gender, ethnicity, or racial background.[870] However, since artificial intelligence systems do not necessarily lead to treatment differences openly, ECtHR may consider a wider margin for interference with an artificial intelligence system. The *Loomis* case in the U.S. can be taken as a source of analogy here and it can be imagined that since the COMPAS was not the only decision maker in the relevant case, given that a similar recidivism tool related discrimination claim courts comes to the ECtHR, it is possible to think the Court could recognize a wider margin for using the tool due to 'efficiency' it provides in courts.[871]

---

[866] Buckley v. The United Kingdom, Rep. 1996-IV ¶ 76 (1996); Oršuš and Others v. Croatıa, Rep. Judgm. Decis. 2010 ¶ 157 (2010).
[867] *Belgian Linguistic Case (No. 2)*, *supra* note 507 at ¶ 10 of the "Law" part; ¶ 7 of the "Decision of the Court" part; Molla Sali v. Greece, *supra* note 856 at ¶ 135; Burden v. The United Kingdom, *supra* note 434 at ¶ 60; Fabris v. France, *supra* note 808 at ¶ 56.
[868] Rasmussen v. Denmark, *supra* note 427 at ¶ 40; Petrovic v. Austria, Rep. 1998-II ¶ 38 (1998); Stec and Others v. The United Kingdom, Rep. Judgm. Decis. 2006-VI ¶ 51-52 (2006); Carson and Others v. the United Kingdom, *supra* note 426 at ¶ 61; Fabris v. France, *supra* note 808 at ¶ 56; Molla Sali v. Greece, *supra* note 856 at ¶ 136.
[869] Stec and Others v. The United Kingdom, *supra* note 868 at ¶ 52; Burden v. The United Kingdom, *supra* note 434 at 60; Carson and Others v. the United Kingdom, *supra* note 426 at 61.
[870] D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 176; Abdulaziz, Cabales and Balkandali v. The United Kingdom, A94 ¶ 78 (1985); Konstantin Markin v. Russia, *supra* note 865 at ¶ 127; Biao v. Denmark, *supra* note 434 at ¶ 114.
[871] *See* Chapter 2.1.3

Another point concerning the margin of appreciation is that given that there is a European consensus on a subject, the margin recognized for states could be reduced.[872] In the current artificial intelligence scene, it would not be possible to say that there is a European consensus on matters involving artificial intelligence. It could limit the margin of appreciation if the CoE states ever reaches a consensus on artificial intelligence and its relevant discriminatory risks. Such consensus could also be reached if legally binding rules specific to artificial intelligence were to be accepted in the Council of Europe.[873] The Court could then consider the CoE's works on artificial intelligence in artificial intelligence related cases. Given that the discriminatory risks involved in artificial intelligence are widely known, there should be a European consensus in this regard. However, for the ECtHR to accept such consensus, the states require to take more action toward artificial intelligence.

Proving the Discrimination Through Artificial Intelligence

In human rights law, the general standard of proof on which the burden is embarked on the person who claims applies.[874] All types of evidence are attainable, and the court makes an unrestrained evaluation for evidence that qualifies as *"sufficiently strong, clear and concordant inferences or similar unrebutted presumptions of fact."*[875] The weaknesses regarding artificial intelligence could be bringing such strong and clear evidence as obtaining such evidence in artificial intelligence related discriminations may not always be possible.[876] If available, artificial intelligence source codes, algorithms, training data, other similar instances, and statistics may all be considered as evidence. Yet, the black box of artificial intelligence and the opaqueness of the artificial intelligence based systems leaves a lacuna for proving and even demonstrating the difference in treatment.

Luckily, there is a specific situation for discrimination claims that could help overcome the particular evidentiary difficulties of discrimination through artificial

[872] Ünal Tekeli v. Turkey, Rep. Judgm. Decis. 2004-X Extr. ¶ 54 (2004); Stec and Others v. The United Kingdom, *supra* note 868 at ¶ 64; Weller v. Hungary, *supra* note 863 at ¶ 28; Konstantin Markin v. Russia, *supra* note 865 at ¶ 126.
[873] *See* Chapter 3.1.3
[874] Aktaş v. Turkey, Rep. Judgm. Decis. 2003-V Extr. ¶ 272 (2003); D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 179.
[875] Timishev v. Russia, Rep. Judgm. Decis. 2005-XII ¶ 39 (2005); Nachova v. Russia, Rep. Judgm. Decis. 2005-VII ¶ 147 (2005); D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 178.
[876] *See also* for an EU non-discrimination law perspective regarding the evidential requirements for discrimination through artificial intelligence Wachter, Mittelstadt, and Russell, *supra* note 19 at 15.

intelligence. Regarding discrimination, the burden of proof is on the state, with the condition that the applicants show a difference in treatment.[877] While this protection is essential for discrimination through artificial intelligence, there could still be problems with regard to proving it because even the difference in treatment could be challenging to ascertain as the examination is based on comparison with others in similar situations.

Exceptions to the standard burden of proof extend to situations where it is *"extremely difficult in practice for the applicant to prove."*[878] Thus, if there is no chance for the applicant to bring any evidence due to the opaqueness, secrecy, or any other reason of the artificial intelligence system, human rights bodies may shift the burden from the applicant to the state. In cases where the issue is wholly or primarily known only by the authorities, the burden of proof is on the state to provide a satisfactory explanation.[879] If the burden of proof is left on the applicant, proving discrimination through artificial intelligence could be highly challenging as it would require accessing data and algorithms of the AI-systems.[880] The CoE Resolution on Preventing Discrimination caused by artificial intelligence also mentions that if the burden of proof is on the claimant, difficulties may emerge in proving due to inadequacies in transparency and accountability.[881] In the context of the EU non-discrimination law, Hacker does not find establishing a *prima facie* discrimination and shifting of the burden of proof similar to the ECtHR enough for addressing discrimination through artificial intelligence.[882] However, if an opaque decision-making AI-system is used, the applicant would only require to show the different treatment but not prove the discrimination because proving would be extremely difficult for the applicant due to the opaqueness and the complexity of the AI-system. When met with a discrimination claim, the state may argue that differential treatment was justified.[883]

---

[877] Chassagnou and Others v. France, Rep. Judgm. Decis. 1999-III ¶¶ 91-92 (1999); Timishev v. Russia, *supra* note 875 at ¶ 39; D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 189; Nachova v. Russia, *supra* note 875 at ¶ 157.
[878] D.H. and Others v. The Czech Republic, *supra* note 434 at ¶¶ 179, 189; Hoogendijk v. The Netherlands, *supra* note 461 at ¶ 2 of the "Law" part.
[879] Salman v. Turkey, Rep. Judgm. Decis. 2000-VII ¶ 100 (2000); Anguelova v. Bulgaria, Rep. Judgm. Decis. 2002-IV ¶ 111 (2002).
[880] Hacker, *supra* note 19 at 1169.
[881] Council of Europe Parliamentary Assembly, *supra* note 848 at para 5.
[882] Hacker, *supra* note 19 at 1168–1169.
[883] Chassagnou and Others v. France, *supra* note 877 at ¶¶ 91-92; Timishev v. Russia, *supra* note 875 at ¶ 57; D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 177; Biao v. Denmark, *supra* note 434 at ¶ 92; Khamtokhu and Aksenchik v. Russia, *supra* note 865 at ¶ 65.

Proving indirect discrimination can be tricky, even when artificial intelligence is not used. For such cases, the court may resort to statistical evidence, which the courts consider a tool for offering a presumption of discrimination, although insufficient by itself.[884] Statistics may also offer help in proving discrimination through artificial intelligence.[885] Concerns for the burden of proof in the context of using artificial intelligence related cases are underlined by mentioning that the decisions should be possible to contest in the CoE Report of the Committee on Equality and Non-Discrimination on "Preventing Discrimination Caused by the Use of Artificial Intelligence".[886]

Case Law Relevant to Discrimination Through Artificial Intelligence

There are no established or pending cases for discrimination through artificial intelligence under the ECtHR at the time of this study, except several cases relating to use of algorithms, which are about the right to respect for private and family life under Article 8; the freedom of expression under Article 10, and may also be relevant on a certain level to the right to equality and non-discrimination under Article 14 of the ECHR.[887] Although the cases noted below are not about discrimination, because of the use of algorithms, they include aspects that went through court review that may be similar or applied in cases of discrimination through artificial intelligence. Accordingly, only the parts that concern the said elements are explored hereunder.

The case *Sigurður Einarsson and Others v. Iceland* deals with the use of an e-discovery system in a criminal proceeding, which was operated by putting keywords, and

---

[884] D.H. and Others v. The Czech Republic, *supra* note 434 at ¶ 180; Hugh Jordan v. The United Kingdom, *supra* note 461 at ¶ 154; Hoogendijk v. The Netherlands, *supra* note 461 at ¶ 2 of the "Law" part; Zarb Adami v. Malta, *supra* note 461 at ¶¶ 76-77.

[885] See for more explanation on statistical evidence and discrimination through artificial intelligence, with a focus on EU non-discrimination law Wachter, Mittelstadt, and Russell, *supra* note 19 at 32–35.

[886] CHRISTOPHE LACROIX, *supra* note 13 at 10.

[887] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 18; Magyar Kétfarkú Kutya Párt v. Hungary concerns fair elections and freedom of expression regarding a mobile application produced by a political party to share anonymous photos of voter ballots. Magyar Kétfarkú Kutya Párt v. Hungary, ¶¶ 18-21 (2020), https://hudoc.echr.coe.int/fre?i=001-200657 (last visited Jun 11, 2023); Delfi AS v. Estonia concerns the need for pre-publishment content moderation by automatic word-filtering regarding anonymous defamatory comments on a news company's website. Delfi AS v. Estonia, ¶¶ 11-15, 157 (2015), https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22002-8960%22]}; Sigurður Einarsson and Others v. Iceland concerns the use of an e-Discovery system and the denial of the applicants' lawyers to complete information thereof in terms of the right to a fair trial. Sigurður Einarsson and Others v. Iceland, *supra* note 573 at ¶¶ 16-21; Big Brother Watch and Others v. the United Kingdom and Centrum för rättvisa v. Sweden concern secret surveillance regimes regarding the respect for private life and freedom of expression. Big Brother Watch and Others v. the United Kingdom, (2021), https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22002-13278%22]}; Centrum för rättvisa v. Sweden, (2021), https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22002-13279%22]}.

in return, a cluster of documents was provided by the program along with "tagging" of the documents being possible.[888] The applicants argued that they were denied from using the e-Discovery system, thus the equality of arms was hindered because the prosecution was able to access the system and select whatever proof was going to be used from the collected documents.[889]

The ECtHR has acknowledged that it is permissible to eradicate some of the vast amounts of data to consider what is relevant and what is not to identify the relevant and handle it in proportion to the defense should be able to involve in defining what is relevant.[890] Regarding the data that is "tagged" at the end of the e-Discovery tool, which was reviewed by the investigators manually and operating the tool on what to include in the investigation, the Court held that not involving the defense in this process raised issues for the right to a fair trial.[891] However, since the applicants, in that case, have not formally requested a court order to access the "full collection of data," or for additional searching or suggested searching for the system using their keywords, the Court has considered the right to a fair trial was not violated.[892] The problem, in this case, was not due to the categorization made by an AI-system; instead, it resulted from not having access to the system and being excluded from the "filtering" process, which could also violate the equality of arms.[893]

The ECtHR in the *Delfi AS v. Estonia* case regarding an automatic word-based filter used by a news company, considered that it failed to filter out hate speech and violence provoking speeches. The *Delfi AS v. Estonia* case, even though it was not about discrimination, could be helpful because it concerns hate speech.[894] In the case at hand, the applicant news company has set up human content moderators to remove inappropriate comments on the website.[895] It should be noted that the court has considered that an obligation to take adequate measures against hate speech does not mean "private censorship".[896]

---

[888] Sigurður Einarsson and Others v. Iceland, *supra* note 573 at ¶ 16.
[889] *Id.* at ¶¶ 66-69, 89.
[890] *Id.* at ¶ 90.
[891] *Id.* at ¶ 91.
[892] *Id.* at ¶¶92-93.
[893] *Id.* at ¶¶ 69-72.
[894] Delfi AS v. Estonia, *supra* note 887 at ¶ 156.
[895] *Id.* at ¶ 157.
[896] *Id.*

The court considered that the interests of others and society as a whole might entitle states to impose liability on internet news portals if they fail to remove unlawful comments without delay, even if there is no notice from the alleged victim.[897] Therefore, there is an expectation of constant active control on the internet in such situations, which could be compared to artificial intelligence. There could be a violation if fully automated content moderation fails to prevent hate speech and discriminatory comments. On the other hand, it could also interfere that if there is no active control over the content, there could still be a violation.

The cases of *Big Brother Watch and Others v. The United Kingdom* and the *Centrum för Rättvisa v. Sweden* were about bulk interception processes in secret intelligence services that consisted of *"interception and initial retention of communications"* and relevant traffic data, *"application of specific selectors"* to the retained data; *"examination of selected communications data by analysts"* and *"subsequent retention of data"* and use of the data including sharing with third parties.[898]

The Grand Chamber of the ECtHR has held that in the bulk interception, communications of many individuals are not necessarily related to the interest of intelligence services, and some of the communications may be 'filtered out' in the first step.[899] The initial search in the bulk interception, which could start targeting individuals, is mainly automated by different types of "selectors," including "strong selectors," such as email addresses, and in the third stage, analysts examine the intercepted material.[900] In the last step of the bulk interception, the intercepted data is used, and it may be shared with foreign intelligence services.[901] The Court assessed the bulk interception concerning the right to private life under Article 8 and freedom of expression under Article 10. Regarding the first right, the ECtHR mentioned that for a bulk interception regime, there should be safeguards concerning "sufficient clarity" regarding the grounds for the bulk interception and the situations when it is acceptable to intercept an individual's communications, requiring a duration of interception, data storage, and usage procedures, precautions for third party sharing and conditions for data removal or destroyal.[902] In

[897] *Id.* at ¶ 159.
[898] Big Brother Watch and Others v. the United Kingdom, *supra* note 887 at para.325; Centrum för rättvisa v. Sweden, *supra* note 887 at 239.
[899] Big Brother Watch and Others v. the United Kingdom, *supra* note 887 at ¶ 326.
[900] *Id.* at ¶¶ 327-328.
[901] *Id.* at ¶ 329.
[902] *Id.* at ¶ 330.

addition, the ECtHR held that "end-to-end safeguards" are required for the necessity and proportionality of the measures taken, such as independent authorization for the bulk interception, defined object and scope, *ex post facto review* for the operation.[903]

Using selectors could relate to discrimination in the Big Brother case, in which the ECtHR has said where the targeting of particular individuals could begin, the United Kingdom government asserted that explaining or substantiating selectors or search criteria before authorization would restrict the effectiveness of the bulk interception.[904] In the Centrum för Rättvisa, Sweden has set for prior authorization of selector categories.[905] The ECtHR considered while the inclusion of all selectors in the authorization may not be possible, at least selector types or categories should be identified in the authorization and that when "strong selectors" are being used, justification of each selector is required.[906]

The ECtHR considered that every step in the bulk interception, including the authorization and renewals, choosing and applying selectors and query terms, and storage, transmission, or deletion, should be subject to the supervision by an independent authority to confirm it is necessary for a democratic society, by measuring necessity and proportionality; the Court has considered that since the selector choices and query terms defined the scope of communications examined by an analyst, it is essential for at least the category selectors to be identified in authorization and for strong selectors that are linked to identifiable individuals to be held by a prior authorization based on independent and objective verifications.[907] Furthermore, the Court mentioned an effective remedy available for everyone who suspects their communications being intercepted to ensure there is no abuse of surveillance powers.[908]

In *Big Brother Watch v. the UK*, the court had established that there were no "end-to-end" safeguards for preventing arbitrariness and abuse risk, due to the absence of independent authorization, not including selector categories when applying for a warrant, and due to that, the court had considered there had been a violation of Article 8 as it was

[903] *Id.* at ¶ 350.
[904] *Id.* at ¶ 353.
[905] Centrum för rättvisa v. Sweden, *supra* note 887 at ¶ 300; Big Brother Watch and Others v. the United Kingdom, *supra* note 887 at ¶ 353.
[906] Big Brother Watch and Others v. the United Kingdom, *supra* note 887 at ¶ 354-355.
[907] *Id.* at ¶¶ 356, 382.
[908] *Id.* at ¶ 357.

not capable of being necessary for a democratic society.[909] Regarding the freedom of expression in the *Big Brother Watch v. the UK*, the court has considered that selectors can be either intentionally or unintentionally connected to confidential journalistic material, which was the concern in the case, and also found a violation of Article 10 following its same assessment for Article 8.[910] The Court considered in *Centrum för Rattvisa* that even though the interception was within the limits of being necessary in a democratic society, it has passed the margin of appreciation because of the inadequacies in the safeguards for arbitrariness and the risk of abuse and, therefore, violation of Article 8.[911] The decisions above not only showed the court's view of cases involving algorithmic processes to date but also established a context for the potential problems mentioned so far in the thesis.

Visualizing the Case Law for Discrimination Through Artificial Intelligence

When turned to the vast collection of case law under the ECtHR for discrimination claims, the cases that can work as a basis for comparison for future applications involving artificial intelligence are the ones that deal with discrimination claims resulting from a general or specific rule or policy or well-established practices. *Di Trizio v. Switzerland* case is such a case that has no relation to artificial intelligence yet can be relevant for artificial intelligence systems. It is not a pioneer case for discrimination law, yet it has the potential to be highly noteworthy from the perspective of artificial intelligence. In *Di Trizio*, there is no machine learning or other artificial intelligence methods, but a formula is utilized to decide on disability pensions. The applicant argued that the combined method used for calculating the degree of disability has led to her refusal from eligibility for disability benefit.[912] Since the problem in this application had resulted from the use of the calculation method, it can be compared to artificial intelligence tools. The applicant claimed that the method was made on a presumption that only one partner of a couple, primarily men, was the breadwinner while the other partner, usually female, managed the household and children. According to the applicant, if the couple were to prefer shared roles, there was a risk of losing entitlements to disability benefits.[913]

In this application, it is unclear whether the formula calculates automatically or not. However, whether a human follows this formula or a computer or an artificial intelligence

[909] *Id.* at ¶¶ 425-427.
[910] *Id.* at ¶ 447-458.
[911] Centrum för rättvisa v. Sweden, *supra* note 887 at ¶ 372-374.
[912] Di Trizio v. Switzerland, *supra* note 467 at ¶ 48.
[913] *Id.*

system does, the outcome of the formula is discriminatory. The formula here construes as an algorithm. There is also a case of redundant encoding here, where the formula is discreetly biased towards women with children, although it does not aim to do so. This case is also an instance of intersectional discrimination. The formula would not yield discrimination if it were to be applied to a man in a similar condition.

Di Trizio claimed that since the combined method would not be applied, if she were not to perform any paid work, her disability percentage would be 44%, and she would be entitled to disability benefits, passing the minimum threshold of 40%.[914] She also claimed that she was discriminated when compared to persons who are not caregivers to children and do not manage households, as they could work full-time. She said that such a person in her situation would be given %55 percent disability.[915]

The applicant finally claimed discrimination on two grounds, first, disability, by explaining that the combined method discourages disabled persons from joining the workforce through part-time work as it may cost loss of the disability benefit, and secondly, on the grounds of gender, because the legal arrangements affected mostly women after childbirth.[916]

The Court considered that the statistics of the combined method usually were applied to women wanting to work part-timely after having children and brought up the recognition of the Swiss Federal Court that the combined method sometimes led to a loss of disability benefits, especially for women working part-time after giving birth.[917] The Court considered that the application of the combined method was driving the applicant and her partner to reconsider their share of roles in their relationship, which could impact their family and professional life.[918] After giving regard to these, the Court considered the application to fall within the ambit of family life under Article 8 and to private life because these affect a person's choices regarding dividing their private life between work, household, and children as well.[919]

The Government had acknowledged the criticism of the Federal Court's case law about the combined method that it mainly affected women whose working hours dropped

[914] *Id.* at ¶ 49.
[915] *Id.*
[916] *Id.*
[917] *Id.* at ¶ 62.
[918] *Id.*
[919] *Id.* at ¶ 63-65.

after childbirth but argued that this is a societal phenomenon not related to health, and accordingly, it should not be covered by disability insurance.[920] The Government had argued that the Swiss Federal Court also ruled that the assessment method was not based on gender or any other discrimination ground; instead, on the loss of capacity to perform routine tasks.[921]

The Government explained that changing the assessment method for disability of part-time working injured persons is discussed at the political level, yet ended up with no result, indicating that there is no other sustainable alternative for the method.[922]

The Court, in its assessment, underlined that if a general policy or a measure disproportionately affects a particular group, even if the policy does not particularly target the group, it may still be discriminatory.[923] ECtHR also accepts that discrimination not only results from a legislative measure, but can also stem from a *de facto* situation.[924] The Court expects the applicant to show evidence of disproportionately harmful effects on a particular group indicating a presumption of indirect discrimination.[925] After the applicant shows this presumption, the Court expects the State to disprove the presumption of indirect discrimination by clarifying that the difference in treatment was due to objective factors, not the ones claimed by the applicant.[926]

As mentioned earlier, statistics in themselves are not interpreted as disclosing a discriminatory practice.[927] But the Court considers the statistics, particularly in discrimination claims relating to general measures or *de facto* situations, to uncover a difference in treatment between two groups in similar situations.[928]

In *the Di Trizio* case, the Court first examined the existence of a presumption of indirect discrimination. To do that, first gave regard the number of applications of the combined method, which showed among the disability applications, the 97% of the cases concerned women in the relevant year.[929] Secondly, the Court had scrutinized that the

[920] *Id.* at ¶ 76.
[921] *Id.* at ¶ 77.
[922] *Id.* at ¶ 78.
[923] *Id.* at ¶ 80; Hugh Jordan v. The United Kingdom, *supra* note 461 at ¶ 154.
[924] Di Trizio v. Switzerland, *supra* note 467 at ¶ 80; Zarb Adami v. Malta, *supra* note 461 at ¶ 76.
[925] Di Trizio v. Switzerland, *supra* note 467 at ¶ 84.
[926] *Id.*; D.H. and Others v. The Czech Republic, *supra* note 434 at ¶¶189, 195.
[927] Di Trizio v. Switzerland, *supra* note 467 at ¶ 85; Hugh Jordan v. The United Kingdom, *supra* note 461 at ¶ 154.
[928] Di Trizio v. Switzerland, *supra* note 467 at ¶ 85; Hugh Jordan v. The United Kingdom, *supra* note 461 at ¶ 180; Zarb Adami v. Malta, *supra* note 461 at ¶ 77; Hoogendijk v. The Netherlands, *supra* note 461.
[929] Di Trizio v. Switzerland, *supra* note 467 at 88.

Swiss Federal Court itself admitted that most cases for combined method application were women whose working hours dropped after childbirth.[930] The Court also reckoned on the government's acceptance of the combined method's impact on women and the Swiss Federal Council's report recognizing that 98% of the cases related to the combined method had women claimants.[931] The Court found these satisfactory for deducing a presumption of indirect discrimination.[932] For an AI-based decision-making tool, the Court may give regard to tool and its impact.

After concluding a presumption of indirect discrimination, the Court secondly explored the presence of an objective and reasonable justification for the different treatment in two stages, first, if there is a legitimate aim or if there is not a reasonable relation of proportionality between the means employed and the aim.[933] The Government described the difference in treatment for women after childbirth by addressing the aim of the disability insurance, which was covering the risk of becoming unable to engage in paid work or perform routine tasks due to disability which they could do if they were in good health.[934] The Court here found that aim satisfactory and moved on to the examination of whether the difference is reasonable and proportionate.[935] The Court may find the simple yet rational aims mentioned in the areas of use part of this chapter as legitimate for using artificial intelligence.

The Court emphasized that the applicant worked full-time as a shop assistant before becoming disabled, and she initially received a 50% disability benefit following the disability.[936] Based on an assumption in her words to the Disability Insurance Office that if she were not to become disabled, she would have reduced her working hours after giving birth, her disability benefit was terminated after giving birth to twins due to the combined method.[937]

The government claimed that the combined method was not based on gender and only took into account the loss of the capacity of the individual to work or perform routine

[930] *Id.* at ¶ 89.
[931] *Id.*
[932] *Id.* at ¶ 90.
[933] *Id.* at ¶ 91.
[934] *Id.* at ¶ 92.
[935] *Id.* at ¶ 93.
[936] *Id.* at ¶ 94.
[937] *Id.*

tasks or both due to disability.[938] The government insisted that the loss of income experienced by individuals who reduced their work hours after childbirth was an independent matter because family considerations could apply to non-disabled persons besides the disabled.[939] In a case related to artificial intelligence, the parties may build their arguments on the principles of which the artificial intelligence tool or system uses in decision-making.

According to the Court, the pursuit of disability insurance is to cover individuals against the risk of becoming unable to work or perform routine tasks due to disability is in line with the essence and the constraints of such an insurance scheme that has limited resources and acts by the principle of control of expenses.[940] Yet, the principle of equality -equality of the sexes in this case- should not be forgotten, which restricts the margin of appreciation because different treatment based on sex and gender demands very weighty reasons.[941] It is natural to expect that cases involving gender or racial discrimination through artificial intelligence claims to be examined more strictly.

The Court noted that the applicant would likely have received a partial disability benefit if she had worked full-time or altogether stayed at home. She previously worked full-time and was granted such benefits until giving birth to children. Thus, the refusal to entitlement to benefits was based on her statement that she desired to reduce work hours to take care of her home and children. The Court had decided that the combined method is a source of discrimination for the majority of women wanting to work part-time after delivery.[942] This approach of the Court suggests that in an artificial intelligence based discrimination claim, the Court may likely compare what would the applicant's situation be under different circumstances.

Previously, the Swiss Federal Court considered that finding a solution to this problem was for the legislature than the courts to find a way to handle sociological conditions.[943] Swiss Federal Council also recognized in a report that the combined method could result in undervaluing the degree of disability, and potential indirect discrimination could thus

[938] *Id.* at ¶ 95.
[939] *Id.*
[940] *Id.* at ¶ 96.
[941] *Id.*
[942] *Id.* at ¶ 97.
[943] *Id.* at ¶ 99.

arise.[944] The ECtHR found that these indicated the combined methods are not in line with gender equality requirements.[945]

If there are more suitable ways, then justification will not be possible. Accordingly, the Court referred to different calculation methods for disability under Swiss law and stated that it is possible to utilize a method that pays attention to women's choice of work after childbirth, thus enabling better gender equality without endangering the objective of the disability insurance.[946] There may be difficulties in the justification of discrimination through artificial intelligence.[947]

In a discrimination through artificial intelligence application, the Court may also inquire the prospects if the artificial intelligence tools are not used for that same process. If the results of the two differs in discrimination, then utilizing such a tool may not be allowed.

The Court also noted in *Di Trizio* that refusing even a partial benefit impacts the applicant significantly, even if it is assumed she could work part-time.[948] Based on these observations, the Court concluded that the different treatment of the applicant did not have a reasonable justification, which revealed a violation of Article 8 of the ECHR.[949] By following the path of this case, similar decisions could be imagined for discrimination through artificial intelligence applications.

2.2.2 Further Considerations from the Human Rights Law Perspective

Discrimination through artificial intelligence may appear in different ways. First, it may occur through the target variables and class; second, the labeling; thirdly, through the collection and the training data; fourthly, through feature selection; fifthly, through proxies and lastly, with the purpose of discriminating.[950] This includes direct discrimination with an intent to discriminate.[951]

The areas of use, the examples, and the discrimination grounds may be broader than in this chapter. It should be again reminded that cases of intersectional discrimination,

---

[944] *Id.* at ¶ 100.
[945] *Id.*
[946] *Id.* at ¶ 101.
[947] Hacker, *supra* note 19 at 1146.
[948] Di Trizio v. Switzerland, *supra* note 467 at ¶ 102.
[949] *Id.* at ¶ 103.
[950] ZUIDERVEEN BORGESIUS, *supra* note 19 at 23.
[951] Barocas and Selbst, *supra* note 213 at 692; ZUİDERVEEN BORGESİUS, *supra* note 19 at 22.

especially in artificial intelligence, should also be considered. Moreover, not every different treatment is considered discrimination. Individuals come across different treatments daily, which do not convey discriminatory behavior. But some of such different treatment may lead to discreet discrimination that could be overlooked because of not being in the scope of non-discrimination law.[952] This issue will be discussed further in the third chapter, why artificial intelligence should be specifically addressed and subject to regulation within the context of discrimination law.

If there is an intent to discriminate, training data may be intentionally selected to produce discriminatory outcomes during data mining, or proxies may be selected for protected classes, making it more difficult to detect discrimination.[953] The risk of discrimination associated with features that correspond to discrimination grounds may be mitigated if these features, which could lead to discriminatory results, are not included in the system as input data.[954]

Discrimination Through Artificial Intelligence and Social Rights

Many of the examples of discrimination explored under this chapter have connections to social rights. The body of social rights is vulnerable to discrimination through the use of artificial intelligence. In the employment sector, for example, artificial intelligence could endanger the equal treatment of workers with a performance measurement tool or equal opportunities for work with biased recruitment tools. [955] Likewise, recruitment applications could hinder equality, especially gender equality.[956] Even though the scope of the thesis is limited with ECtHR in terms of examining existing human rights law framework, other human rights law instruments and mechanisms, including those for social and economic rights such as the ISECR and the ESC, should also be considered when handling the right to equality and non-discrimination.[957]

Social and economic rights are emphasized similarly by the CAHAI regarding work life, concerns for workers' rights, and potential impacts on workers' social security when AI-systems are used for social security and social welfare decisions.[958] Caution for using

[952] ZUIDERVEEN BORGESIUS, *supra* note 19 at 67.
[953] Barocas and Selbst, *supra* note 213 at 692; Joshua A Kroll et al., *Accountable Algorithms*, 165 UNIV. PA. LAW REV. 74, 682 (2017); ZUIDERVEEN BORGESIUS, *supra* note 19 at 22.
[954] ZUIDERVEEN BORGESIUS, *supra* note 19 at 53.
[955] Muller and The CAHAI Secretariat, *supra* note 90 at 28–29.
[956] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 10.
[957] Muller and The CAHAI Secretariat, *supra* note 90 at 28–29.
[958] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 10.

artificial intelligence within the welfare state is mentioned in the context of the discriminatory potential of artificial intelligence in the CoE "Report of the Committee on Equality and Non-Discrimination on Preventing Discrimination Caused by the Use of Artificial Intelligence."[959]

Furthermore, a declaration on the risks of artificial intelligence based decision-making in the social safety net has been published by the Committee of Ministers of the CoE.[960] The Committee considers that using artificial intelligence in the social safety nets could help to speed up and improve services for individuals while reducing costs and preventing fraud or corruption.[961] However, in deciding the distribution of social goods with artificial intelligence, there exists a risk that individuals' access to social rights and benefits may be reduced or impeded.[962] Artificial intelligence may even worsen the vulnerabilities of socially disadvantaged individuals and groups and may reproduce and reinforce bias and discrimination towards them.[963]

The immediate adoption of such tools by public authorities without thorough scrutiny may yield harms that are challenging to correct on a larger scale, even if individual instances are remedied.[964] The declaration by the Committee of Ministers recalls the ESC and previous recommendations of the Committee of Ministers, noting such use of artificial intelligence may be contrary to these standards. [965] The Committee of Ministers underlines that not only social rights but civil and political rights may also be affected by artificial intelligence.[966]

The Committee of Ministers has chosen to use the expression 'draw attention' and reminds member states of the potential human rights risks associated with artificial intelligence based decisions for social services in the absence of adequate supervision of

[959] CHRISTOPHE LACROIX, *supra* note 13 at 10.
[960] Council of Europe Committee of Ministers, *Declaration by the Committee of Ministers on the Risks of Computer-Assisted or Artificial-Intelligence-Enabled Decision Making in the Field of the Social Safety Net*, Decl(17/03/2021)2 (2021), https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a1cb98 (last visited Aug 4, 2022).
[961] *Id.*
[962] *Id.*
[963] *Id.*
[964] *Id.*
[965] *Id.*
[966] *Id.*

"management, attribution, revocation of entitlements, assistance, and related benefits".[967]

The Committee of Ministers emphasized the principles of "legal certainty, legality, data quality, non-discrimination, and transparency", as well as the need for user knowledge and human oversight.[968] The Committee of Ministers also underlines the need for 'effective arrangements' to protect vulnerable individuals from serious harm, such as extreme poverty and homelessness. This may be deemed as a reference to positive obligations.[969] The Committee of Ministers has also remarked on the significance of effective responsibility and accountability during the design, development, deployment, and evaluation phases of artificial intelligence systems.[970] The Committee of Ministers also underlined the need for a proactive approach to remedy seeking and acquiring the information required to challenge artificial intelligence based decisions.[971] Accordingly, public authorities should be mindful of these principles when employing artificial intelligence systems. Another relevant issue to consider is fair access to the benefits of artificial intelligence, which can directly impact social rights and open up a new arena for discussion.[972] The scope of the ECHR does not extend to social rights as long as their protection scope coincides with a right in the convention, but when looking at artificial intelligence from the perspective of human rights, social rights should not be overlooked.[973]

Keeping Private Actors Responsible for Discrimination Through Artificial Intelligence

The main actors in the artificial intelligence field is from the private sector and there could be challenges regarding keeping the private actors responsible in artificial intelligence.[974] While the human rights obligations are encumbered by states for the violations that stem from public authorities as well as actions of private individuals due to the horizontal effect, involving businesses directly to the responsibility realm could be

[967] *Id.*
[968] *Id.*
[969] *Id.*
[970] *Id.*
[971] *Id.*
[972] Marcello Ienca & Effy Vayena, *AI Ethics Guidelines: European and Global Perspectives*, *in* TOWARDS REGULATION OF AI SYSTEMS: GLOBAL PERSPECTIVES ON THE DEVELOPMENT OF A LEGAL FRAMEWORK ON ARTIFICIAL INTELLIGENCE SYSTEMS BASED ON THE COUNCIL OF EUROPE'S STANDARDS ON HUMAN RIGHTS, DEMOCRACY AND THE RULE OF LAW 38, 51 (2020), https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a.
[973] Smuha, *supra* note 19 at 597.
[974] McGregor, Murray, and Ng, *supra* note 1 at 313.

helpful in terms of artificial intelligence, for example to handle the global impact of artificial intelligence in different jurisdictions.[975] The business and human rights field presents a relevant opportunity in this regard to enable rapid addressing of the human rights violations that may occur through artificial intelligence.[976] Furthermore, the non-binding documents that will be explored in the next chapter can to a degree be accepted as tools to support the requirements under the UN Guiding Principles on Business and Human Rights and could bridge traditional human rights law and business and human rights. The CoE Commissioner for Human Rights also supports human rights due diligence and applying business and human rights instruments to artificial intelligence, mentioning that the states should implement them.[977] In the UN Guiding Principles on Business and Human Rights, human rights due diligence processes have been mandated on businesses for identifying, mitigating and accounting for their impact on human rights.[978] Discussions on keeping private actors responsible for human rights has an increasing tendency in human rights law through expanding it to businesses.[979] Finding responsibility for the artificial intelligence actors in the private sector, such as artificial intelligence developers and companies, for human rights violations could be made possible through business and human rights by bringing obligations for private actors.[980] The private actors should aim to provide remedies and ensure access to justice for deterring and mitigating human rights impacts of artificial intelligence.[981]

In conclusion, it should be noted that existing human rights law itself may not provide a complete solution to handling the right to equality and other human rights; it may require support from other areas of law, such as data protection law or new regulations specific to artificial intelligence could be required, which is explored in the third chapter.[982] However, human rights law provides an essential framework for addressing human rights

[975] *Id.* at 313.
[976] *Id.* at 312–313.
[977] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 9.
[978] UNİTED NATİONS GENERAL ASSEMBLY, *UN Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework*, 15(b) (2011); McGregor, Murray, and Ng, *supra* note 1 at 312.
[979] UNİTED NATİONS GENERAL ASSEMBLY, *supra* note 978; Lorna McGregor, *Accountability for Governance Choices in Artificial Intelligence: Afterword to Eyal Benvenisti's Foreword*, 29 EUR. J. INT. LAW 1079, 313 (2019), https://doi.org/10.1093/ejil/chy086 (last visited Mar 9, 2022).
[980] McGregor, *supra* note 979 at 327; SPECIAL RAPPORTEUR ON FREEDOM OF OPINION AND EXPRESSION, *supra* note 794 at 17.
[981] McGregor, Murray, and Ng, *supra* note 1 at 329.
[982] *Id.* at 313.

related challenges in artificial intelligence.[983] After analyzing discrimination through artificial intelligence from the standpoint of present human rights law, what can be done to improve artificial intelligence with new regulations will be examined under Chapter 3. Some of these regulations are of legal nature, and some are non-binding.

[983] *Id.*; Yeung, Howes, and Pogrebna, *supra* note 529 at 86; Smuha, *supra* note 19 at 100.

Chapter 3:
REGULATING ARTIFICIAL INTELLIGENCE FOR THE RIGHT TO EQUALITY AND NON-DISCRIMINATION

3.1 Approaches for Regulating Artificial Intelligence

When thinking about bringing rules for artificial intelligence, science fiction author Isaac Asimov's "three laws of robotics" comes to mind.[984] Three laws of robotics still inspire and are mentioned in the discussions for regulating artificial intelligence, although being of vague nature and not able to address legal problems.[985] Today, regulation of artificial intelligence is a vast field with various approaches, as the problems stemming from it can be quite different. In this direction, the issue of regulating artificial intelligence should be studied with a focus on different circumstances that may come up with artificial intelligence and application areas, one being this study's scope, the right to equality and non-discrimination in the context of artificial intelligence as a legal problem from the human rights law perspective. Hence, this chapter aims to examine and explore the regulation of artificial intelligence for the right to equality and non-discrimination.

The previous chapter has explained the existing international human rights law on the right to equality and non-discrimination and aim at integrating it with artificial intelligence. International human rights law can be applied to discrimination through artificial intelligence, yet as will be explored in this chapter, specific regulations for artificial intelligence can help to bring better protection for the right to equality and non-discrimination where international human rights law may come short of addressing problems therein. While, on the other hand, international human rights law is applicable

---

[984] Dena Dervanović, *I, Inhuman Lawyer: Developing Artificial Intelligence in the Legal Profession*, *in* ROBOTICS, AI AND THE FUTURE OF LAW 209, 221 (Marcelo Corrales, Mark Fenwick, & Nikolaus Forgó eds., 2018), https://doi.org/10.1007/978-981-13-2874-9_9 (last visited Apr 12, 2023); Suzanne Rab, *Telecoms and Connectivity*, *in* ARTIFICIAL INTELLIGENCE: LAW AND REGULATION 355, 361 (2022), https://www.elgaronline.com/view/edcoll/9781800371712/9781800371712.00031.xml; CHESTERMAN, *supra* note 57 at 173; Isaac Asimov, *Runaround*, (1942), https://web.williams.edu/Mathematics/sjmiller/public_html/105Sp10/handouts/Runaround.html (last visited Apr 12, 2023).

[985] Rab, *supra* note 984 at 361; Oren Etzioni, *Opinion | How to Regulate Artificial Intelligence*, THE NEW YORK TIMES, Sep. 2, 2017, https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html (last visited Aug 1, 2022); CHESTERMAN, *supra* note 57 at 174.

as it is, it could adapt to artificial intelligence by giving rise to new considerations, such as distinct positive and procedural obligations which can be strengthened through specific regulations for artificial intelligence as the CoE aims to achieve with its work on artificial intelligence regulation that will be explored in this chapter.

It should be underlined that the regulatory frameworks discussed in this chapter do not aim to examine how artificial intelligence should be regulated in general. Regulating artificial intelligence and discussions thereof are not limited to human rights, and private and public law concerns beyond human rights law may as well happen. Artificial intelligence, while directly impacting human rights and giving rise to situations that need to be regulated -with law or non-binding documents-, there are also other legal situations regarding artificial intelligence during its entire lifecycle as various legal problems may appear through artificial intelligence. Furthermore, within the scope of human rights, various human rights may deliver risks and violations other than the right to equality and non-discrimination, which call for their respective examinations. The regulatory documents to be studied in this chapter will specifically focus on ensuring the right to equality and non-discrimination. Accordingly, how these documents approach the right to equality and non-discrimination by looking different ways it can be addressed through the principles set for artificial intelligence will be examined. The legally binding and non-binding documents and principles thereunder are explored and discussed within these limits.[986] The comprehensive investigation on regulating artificial intelligence is beyond the scope of this study since the aim of this chapter is not to focus entirely on how to regulate artificial intelligence.

This chapter analyses the existing regulatory documents on artificial intelligence, not only with their provisions directly addressing equality and discrimination but also with other provisions throughout these documents that can be relevant to the right to equality and non-discrimination. The nature of the right to equality and non-discrimination, being a principle for human rights as well as being a right, allows for such insights.

---

[986] EUROPEAN COMMİSSİON, *Commission Staff Working Document Impact Assessment Accompanying the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*, 96 (2021), https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence; ANDREA RENDA ET AL., *Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe: Final Report.*, (2021), https://data.europa.eu/doi/10.2759/523404 (last visited May 21, 2023).

There are, of course, other instruments that can be used to address artificial intelligence related problems, such as data protection laws, other international agreements, and even civil liability laws, both on national, international, and supranational levels, yet since the aim of this study is not to evaluate the regulation of artificial intelligence as a whole but rather look at how to address the right to equality and non-discrimination, such legal documents are left beyond this study.[987] Internationally legally binding documents regarding artificial intelligence are still in development at the time of this thesis, and they have yet to come into force. If legally binding documents were in force with established application history of conflicts and cases, it would have been more convenient to examine solely one of such documents. Yet, since that is not the case at the time of this study, this chapter proceeds with delving into two primary potential legal documents, one from the EU, and one from the CoE, inquiring about the right to equality and non-discrimination through the principles establish(-ed/-ing) for artificial intelligence, to discover how these documents approach to the right to equality and non-discrimination and to make suggestions regarding them. Furthermore, the principles for artificial intelligence are traced in various non-binding documents, which the study finds important, such as from the UNESCO and the OECD and again from the EU and CoE. Likewise, relevant works from the private sector, non-profit and non-governmental organizations are mentioned throughout the chapter.

This chapter will begin by examining two approaches to regulating artificial intelligence. The first approach is regulation through non-binding documents, while the second is regulation through legally binding documents. Subsequently, the non-binding documents and the upcoming binding documents will be explained, specifically the EU AI Act Proposal and the CoE Draft Convention on AI. The chapter will then delve into the main principles of artificial intelligence, providing a path for analyzing the right to equality and non-discrimination within the landscape for artificial intelligence regulations. In this section, the non-binding and binding documents will be discussed together under the principles.

[987] *See* for a more in-depth and extended analysis of regulating artificial intelligence from the EU perspective EUROPEAN COMMISSION, *supra* note 986; See RENDA ET AL., *supra* note 986.

3.1.1 Non-Binding Documents for Regulating Artificial Intelligence

As mentioned above, addressing the problems that artificial intelligence, including the subject matter of this thesis, the right to equality and non-discrimination, among other issues in the realm of artificial intelligence, has led to a need to handle in need for managing these problems. Efforts toward addressing artificial intelligence related problems brought forward certain principles including accountability, transparency, safety, and fairness leading to what is called in the artificial intelligence industry as "ethical artificial intelligence," followed by a whole domain that is called "AI-ethics."[988] The non-binding documents set out principles for artificial intelligence to be willingly adhered to, by addressing different actors in the artificial intelligence field, also relating to concepts of corporate social responsibility and human rights due diligence. [989]

Accordingly, there is a corpus of non-binding artificial intelligence documents depicting the state of affairs in regulating artificial intelligence.[990] Addressing artificial intelligence and its potential problems is an issue on the agendas of different actors in the artificial intelligence field; artificial intelligence developers from the private sector introduced "self-regulations;" non-governmental or non-profit organizations and research institutes laid out non-binding "AI-ethics" documents; and a selection of prominent non-binding documents come from international organizations such as the UNESCO and the OECD, and even actors such as the EU and the CoE has besides their work on legally binding regulatory preparations had brought forward non-binding documents for artificial intelligence, resulting in an accumulation of such documents prepared for artificial intelligence both in non-governmental and governmental agents.[991] The non-binding

[988] Zuiderveen Borgesius, *supra* note 25 at 1582.
[989] Karen Yeung, *Recommendation of the Council on Artificial Intelligence (OECD)*, 59 INT. LEG. MATER. 27, 27 (2020), https://www.cambridge.org/core/journals/international-legal-materials/article/abs/recommendation-of-the-council-on-artificial-intelligence-oecd/EC74B60333EEB276393DB53307519B19; YEUNG, *supra* note 518 at 49.
[990] *See* Anna Jobin, Marcello Ienca & Effy Vayena, *The global landscape of AI ethics guidelines*, 1 NAT. MACH. INTELL. 389 (2019). JESSICA FJELD ET AL., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, (2020), https://papers.ssrn.com/abstract=3518482 (last visited Oct 21, 2021)
[991] Zuiderveen Borgesius, *supra* note 25 at 1582; ZUIDERVEEN BORGESIUS, *supra* note 19 at 49; CHESTERMAN, *supra* note 57 at 174–175; Google AI, *Our Principles*, GOOGLE AI, https://ai.google/principles/ (last visited Apr 10, 2023); Future of Life Institute, *The Asilomar AI Principles*, FUTURE OF LIFE INSTITUTE (2017), https://futureoflife.org/open-letter/ai-principles/; OECD, *OECD Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449 (2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (last visited Apr 10, 2023); HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12; Council of Europe Committee of Ministers,

documents, especially those from international and supranational organizations such as UNESCO, OECD, the EU, and the CoE, can be considered soft law documents.[992] The interest in regulating artificial intelligence is also seen in states, yet most of the works at the time of this study are also soft law instruments.[993] Many of the non-binding documents for artificial intelligence emanate from Asia, Europe, and North America, in parallel to competition for artificial intelligence, also implying the insufficient participation from the developing economies in the said competition and hence its regulation.[994] As seen, many non-binding ethics documents have been published to bring rules for overseeing artificial intelligence.[995] That renders the systematic examination of these documents challenging.[996] The recipients of the non-binding documents are different actors that have a role in developing and using an AI-system.[997]

The EU has produced a variety of non-binding documents on artificial intelligence, the most significant being the "White Paper on Artificial Intelligence" and

Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems, CM/Rec(2020)1 (2020), https://rm.coe.int/09000016809e1154 (last visited Apr 10, 2023); Nicholas Diakopoulos et al., *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms :: FAT ML*, https://www.fatml.org/resources/principles-for-accountable-algorithms; IEEE, THE IEEE GLOBAL INITIATIVE ON ETHICS OF AUTONOMOUS AND INTELLIGENT SYSTEMS. ETHICALLY ALIGNED DESIGN: A VISION FOR PRIORITIZING HUMAN WELL-BEING WITH AUTONOMOUS AND INTELLIGENT SYSTEMS (First Edition ed. 2019), https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/ autonomous-systems.html; Reiling, *supra* note 567 at 6; Ienca and Vayena, *supra* note 972 at 38–39; Urs Gasser & Carolyn Schmitt, *The Role of Professional Norms in the Governance of Artificial Intelligence*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 141, 145–146 (Markus D. Dubber, Frank Pasquale, & Sunit Das eds., 2020), https://doi.org/10.1093/oxfordhb/9780190067397.013.8 (last visited May 15, 2023).

[992] Lance Eliot, *AI Ethics And Legal AI Are Flustered By Deceptive Pretenses Known As AI Ethics-Washing Which Are False Claims Of Adhering To Ethical AI, Including For Autonomous Self-Driving Cars*, FORBES, Jun. 2022, https://www.forbes.com/sites/lanceeliot/2022/06/09/ai-ethics-and-legal-ai-are-flustered-by-deceptive-pretenses-known-as-ai-ethics-washing-which-are-false-claims-of-adhering-to-ethical-ai-including-for-autonomous-self-driving-cars/ (last visited Jun 9, 2023).

[993] CHESTERMAN, *supra* note 57 at 175; Australian Government Department of Industry, Science and Resources, *Australia's AI Ethics Principles | Australia's Artificial Intelligence Ethics Framework | Department of Industry, Science and Resources*, HTTPS://WWW.INDUSTRY.GOV.AU/NODE/91877 (2022), https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles (last visited Apr 13, 2023); New Zealand Government, *Algorithm Charter For Aotearoa New Zealand*, (2020), https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf (last visited Apr 13, 2023); INFO-COMMUNICATIONS MEDIA DEVELOPMENT AUTHORITY (IMDA) & PERSONAL DATA PROTECTION COMMISSION (PDPC), *Model AI Governance Framework (Second Edition)*, 70 (2020), https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf.

[994] Ienca and Vayena, *supra* note 972 at 39.

[995] Reiling, *supra* note 567 at 6.

[996] *Id.*

[997] Gasser and Schmitt, *supra* note 991 at 146.

"the Ethics Guidelines for Trustworthy AI," issued with a general perspective.[998] There are also context-specific documents, such as the "Ethical Guidelines on the Use of Artificial Intelligence and Data in Teaching and Learning for Educators".[999] The EU has also published an assessment list for measurement involving a series of questions that helps to contextualize and materialize the principles developed under the Ethics Guidelines for Trustworthy AI.[1000] Likewise, the CoE has issued numerous recommendations, resolutions, reports that are mentioned throughout the study, including the CEPEJ Charter regarding the judicial use of artificial intelligence.[1001]

OECD Recommendation of the Council on Artificial Intelligence (The OECD AI Recommendation) contains provisions to address artificial intelligence better, corresponding to the EU Ethics Guidelines for Trustworthy AI.[1002] The OECD AI Recommendation has significance regarding a political commitment to global partnership in artificial intelligence recognizing the dangers and benefits of artificial intelligence.[1003]

The UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO AI Recommendation) relates to fields concerning its scope, including education, science, and culture. It aims to guide states regarding the regulation of artificial intelligence and other actors in the artificial intelligence field to incorporate ethics throughout the AI-

[998] EUROPEAN COMMISSION, *supra* note 519; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12.
[999] European Commission Directorate General for Education, Youth, Sport and Culture, *supra* note 650.
[1000] High-Level Expert Group on AI (AI HLEG), *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment*, (2020), https://data.europa.eu/doi/10.2759/791819.
[1001] See e.g. Council of Europe Committee of Ministers, *supra* note 991; Council of Europe Parliamentary Assembly, *supra* note 832; Council of Europe Parliamentary Assembly, *supra* note 848; Council of Europe Parliamentary Assembly, *supra* note 528; Council of Europe Parliamentary Assembly, *Recommendation 2182 (2020) of the Parliamentary Assembly on Justice by Algorithm – The Role of Artificial Intelligence in Policing and Criminal Justice Systems*, Recommendation 2182 (2020) (2020), https://pace.coe.int/pdf/bf427be8505b6833866638cc75e52970b11091824f077b5caaf5f581f21f61c2/rec.%202182.pdf (last visited Aug 4, 2022); Council of Europe Parliamentary Assembly, *supra* note 560; Council of Europe Parliamentary Assembly, *supra* note 638; Council of Europe Parliamentary Assembly, *supra* note 633; Council of Europe Parliamentary Assembly, *supra* note 657; Council of Europe Parliamentary Assembly, *Recommendation 2181 (2020) of the Parliamentary Assembly on the Need for Democratic Governance of Artificial Intelligence*, Recommendation 2181 (2020) (2020), https://pace.coe.int/pdf/114e7524b1e69823f01fc2cb4080f5401933355f1c631e797e23f45fcf041f95/rec.%202181.pdf (last visited Aug 4, 2022); Council of Europe Parliamentary Assembly, *Recommendation 2102 (2017) of the Parliamentary Assembly on Technological Convergence, Artificial Intelligence and Human Rights*, Recommendation 2102 (2017) (2017), https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726 (last visited Aug 4, 2022); Committee of Ministers of the Council of Europe, *supra* note 593; European Commission for the Efficiency of Justice (CEPEJ), *supra* note 590.
[1002] Yeung, *supra* note 989 at 28.
[1003] *Id.*

lifecycle.[1004] It refers to promoting and respecting human rights and freedoms with a specific mention of equality.[1005]

Furthermore, a notable non-binding document from non-governmental organizations that provides guidance on the discriminatory impacts of artificial intelligence is the "Toronto Declaration Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems," (Toronto AI Declaration) published in 2018 against discrimination through algorithms and artificial intelligence.[1006] Moving on from international human rights law, the Toronto AI Declaration addresses the prevention of discrimination through artificial intelligence and the protection of individuals and groups, the responsibilities of states, the use of artificial intelligence by the state, and responsibilities in terms of the private sector, and promotes equality, diversity and inclusion, spreading to a comprehensive coverage under essential principles.[1007]

Regarding the private sector actors that introduced AI-ethics documents, Microsoft, IBM, Google and Facebook's self-regulatory stand out as they are among the forerunners of the artificial intelligence industry. For example, Microsoft mentions advancing artificial intelligence with human-centric ethical principles under its "Responsible AI Standards".[1008] Similarly, Google asserts that artificial intelligence technologies that cause or are likely to cause harm will not be designed or deployed unless Google considers the benefits to surpass the risks.[1009] It should be noted that evaluation of benefits and risks entirely by self-regulation poses concerns about what constitutes a benefit. In addition, Facebook/Meta declares that by taking inspiration from the OECD AI Recommendation and the EU Ethics Guidelines for Trustworthy AI, they organized their "Responsible AI" standards under five principles that encompasses transparency and

---

[1004] UNESCO, *UNESCO Recommendation on the Ethics of Artificial Intelligence*, 21 I.3 (2021), https://unesdoc.unesco.org/ark:/48223/pf0000380455.locale=en.
[1005] *Id.* at para II.8.
[1006] Amnesty International & Access Now, *The Toronto Declaration - Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems*, (2018), https://www.torontodeclaration.org/declaration-text/english/.
[1007] *Id.* at Preamble 2.
[1008] Responsible AI principles from Microsoft, MICROSOFT, https://www.microsoft.com/en-us/ai/responsible-ai (last visited Jun 10, 2023); News has been published about Microsoft that its team has been laid off. Rebecca Bellan, *Microsoft Lays off an Ethical AI Team as It Doubles down on OpenAI | TechCrunch*, (2023), https://techcrunch.com/2023/03/13/microsoft-lays-off-an-ethical-ai-team-as-it-doubles-down-on-openai/ (last visited Mar 15, 2023).
[1009] Google AI Principles, GOOGLE AI, https://ai.google/responsibility/principles/ (last visited Jun 10, 2023).

control, accountability and governance, robustness and safety, fairness and inclusion and privacy and security.[1010]

The private sector tends to handle artificial intelligence related matters by bringing solutions mostly on technical level.[1011] The self-regulation trend in the artificial intelligence sector is seemingly positive, as it brings a certain level of management over artificial intelligence, which is preferred over no regulation in the field.[1012] The problems that may come up in an AI-system have the chance to be addressed at its source and sometimes directly by the designers themselves.[1013]

The principles systemized for artificial intelligence are not uniform and limited in numbers, while some, such as transparency, accountability, and privacy are common.[1014] These non-binding documents remark on legal implications too. For example, many documents underline promoting human rights for the design, development, and deployment of the AI-systems.[1015] The documents on artificial intelligence are aimed at artificial intelligence actors, including the private actors and the states, bringing obligations to be fulfilled by the stakeholders. The CoE, for example, addresses both the states and private actors in the "Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems," and imposes obligations to both.[1016] Various criteria, and rules are set thereunder for AI-systems and artificial intelligence actors to adhere to. Some of the non-binding documents include tools for assessing the meeting of these criteria through the usage of AI-systems.[1017] However, these do not have power to produce legal consequences. In addition, addressing artificial intelligence only through non-binding documents and keeping the issue out of the legal realm may lead to problems significant for human rights

[1010] Facebook's five pillars of Responsible AI, META AI (2021), https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/ (last visited Jun 10, 2023).
[1011] Ienca and Vayena, *supra* note 972 at 51.
[1012] ZUIDERVEEN BORGESIUS, *supra* note 19 at 49.
[1013] *Id.*
[1014] Ienca and Vayena, *supra* note 972 at 39.
[1015] *Id.*
[1016] Council of Europe Committee of Ministers, *supra* note 991 See the section B with the title *"Obligations of States with respect to the protection and promotion of human rights and fundamental freedoms in the context of algorithmic systems" and section C with the title "Responsibilities of private sector actors with respect to human rights and fundamental freedoms in the context of algorithmic systems."*
[1017] *See* for the EU's Self-Assesment Tool HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) FOR SELF ASSESSMENT. (2020), https://data.europa.eu/doi/10.2759/791819 (last visited Aug 6, 2022).

if there is no legal consequences.[1018] The non-binding regulations do not have legal attributes to handle the risks and matters existent in AI-systems and processes as a legally binding document would do.[1019]

Another worry with non-binding documents is the risk of "ethics-washing", which connotates that artificial intelligence actors may use such documents as back doors while keeping clear their 'conscience' through basic rules without facing legal consequences in case of a problem is realized.[1020] The next title of this chapter addresses more discussion on ethics-washing as the concerns surrounding it also constitute a basis for the need to address artificial intelligence with legally binding documents. The non-binding documents should not disregard that inadequate ethics and design would result in harmful outcomes.[1021] Likewise, since the economic benefits of artificial intelligence are fruitful, such motivations may supersede adherence to keeping with ethical principles.[1022] While being necessary, valuable, and useful, the non-binding instruments do not remove the need for regulating artificial intelligence through legally binding rules.[1023]

On the other hand, the non-binding documents on artificial intelligence surely have the potential to navigate and influence legally binding documents.[1024] Human rights related provisions exist in non-binding documents, which to a degree, focus attention on the protection of human rights law, calling for an examination of human rights protection in AI-systems.[1025] The law should adapt to artificial intelligence to prevent claims of denouncing responsibility from artificial intelligence.[1026] The state has obligations to its subjects, such as in criminal law, administrative law, and human rights related aspects,

---

[1018] ZUIDERVEEN BORGESIUS, *supra* note 19 at 50.
[1019] *Id.*
[1020] Ben Wagner, *Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?*, *in* BEING PROFILED: COGITAS ERGO SUM. 10 YEARS OF 'PROFILING THE EUROPEAN CITIZEN 84, 84 (Emre Bayamlioglu et al. eds., 2018), https://doi.org/10.1515/9789048550180-016; ZUIDERVEEN BORGESIUS, *supra* note 19 at 50; See Eliot, *supra* note 992; Patricia Shaw & Charles Kerrigan, *Ethics*, *in* ARTIFICIAL INTELLIGENCE: LAW AND REGULATION 398, 407 (2022), https://www.elgaronline.com/display/edcoll/9781800371712/9781800371712.00034.xml; *See also* Gijs van Maanen, *AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics*, 1 DIGIT. SOC. 9 (2022), https://doi.org/10.1007/s44206-022-00013-3 (last visited May 16, 2023).
[1021] Greene, Hoffmann, and Stark, *supra* note 259 at 2129.
[1022] Thilo Hagendorff, *The Ethics of AI Ethics -- An Evaluation of Guidelines*, 30 MINDS MACH. 99, 10 (2020), http://arxiv.org/abs/1903.03425 (last visited May 16, 2023); International AI ethics panel must be independent, 572 NATURE 415 (2019), https://www.nature.com/articles/d41586-019-02491-x (last visited Jun 3, 2023); Hacker, *supra* note 19 at 1153.
[1023] ZUIDERVEEN BORGESIUS, *supra* note 19 at 50.
[1024] *Id.* at 49.
[1025] Ienca and Vayena, *supra* note 972 at 38.
[1026] Etzioni, *supra* note 985.

and the use of artificial intelligence demands further obligations for the state.[1027] This calls for a closer inspection on the need to address artificial intelligence with legally binding documents.

3.1.2 Why legally binding documents are required to address the right to equality and non-discrimination in artificial intelligence?

The public interest in using artificial intelligence is among the reasons for the need for regulating artificial intelligence with legally binding documents.[1028] As mentioned in the second chapter, existing human rights law framework will help to the specific artificial intelligence regulations.[1029] Existing human rights frameworks have gaps in certain aspects of artificial intelligence, such as the requirement for human control, explainability, and transparency, which could be necessary for addressing the right to equality and non-discrimination, as explained below in detail.[1030] If there is a legal framework fitted for specificities of artificial intelligence, uncertainties regarding the human rights risks of artificial intelligence could be better addressed.[1031]

Where human rights law falls insufficient for discrimination through artificial intelligence, that could be addressed with new legislative works.[1032] Besides the human rights law, existing legally binding documents, such as data protection law, consumer protection law, competition law, and constitutional law, may benefit in ensuring the accountability of artificial intelligence and facilitating equality.[1033] However, the principles for artificial intelligence, such as human oversight, transparency, and technical safety, are not always ensured in the existing legal corpora in a materialized and particular way suited for artificial intelligence, which leads to legal gaps that need to be addressed with a specific regulation.[1034] The lack of protection for the particularities of artificial intelligence, especially for these principles in the existing legal instruments are mentioned

[1027] ZUIDERVEEN BORGESIUS, *supra* note 19 at 54.
[1028] Smuha et al., *supra* note 106 at 2.
[1029] Council of Europe Parliamentary Assembly, *supra* note 832.
[1030] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 23–24.
[1031] *Id.* at 24.
[1032] Smuha, *supra* note 19 at 100.
[1033] YEUNG, *supra* note 518 at 47.
[1034] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 23–24.

in the EU's White Paper on AI and the CAHAI's Feasibility Study.[1035] In order to ensure the up-to-date interpretation of the principles under human rights law for artificial intelligence, constant assessments and where relevant new rules may be regulated.[1036] Yeung suggests that the inadequacies in the non-binding documents indicate a need for an approach complying with human rights norms that integrates tools and techniques and external oversight with public and stakeholder consultation.[1037]

Bringing specific legal regulations regarding artificial intelligence may help overcome some of the challenges in existing human rights law, such as getting protection for the rights that the protection of equality and discrimination may fail to address due to the subsidiary nature of the right in many human rights law documents. For example, the ECHR may come short of addressing discrimination in conjunction with social and cultural rights. However, since human rights violations in the context of artificial intelligence for the right to equality and non-discrimination can appear in various ways related to social rights, advancing 'artificial-intelligence-specific' protections under human rights law may be necessary.

As emphasized throughout the thesis, compliance with human rights law is required when legally addressing artificial intelligence risks.[1038] The approach of the EU Ethics Guidelines for Trustworthy AI is to differentiate between the two meanings of human rights and opt for the ethical interpretation by referencing the ethical norms recognized as human rights.[1039] In a way, human rights can be translated into the realm of artificial intelligence, as the second chapter of this study aimed to accomplish.[1040] The CAHAI Feasibility Study has evaluated possible options for addressing artificial intelligence by amending existing legal instruments, such as an additional protocol to ECHR, to go beyond the existing framework by bringing more legal certainty and clarity for artificial intelligence related rights.[1041] Another suggestion for incorporating human rights into

[1035] EUROPEAN COMMISSION, *supra* note 519 at 9; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 23; See for an overview of existing legally binding documents that can be used for artificial intelligence Mantelero, *supra* note 557.
[1036] Smuha, *supra* note 19 at 101.
[1037] Yeung, Howes, and Pogrebna, *supra* note 529 at 85.
[1038] Smuha et al., *supra* note 106 at 5.
[1039] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 707; Smuha et al., *supra* note 106 at 7.
[1040] Muller and The CAHAI Secretariat, *supra* note 90 at 32.
[1041] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 45.

artificial intelligence is through a framework convention.[1042] The CoE has currently opted for this method by introducing the CoE Draft Convention on AI.[1043]

On the other hand, including the EU Ethics Guidelines for Trustworthy AI, non-binding documents leave a cavity for providing the "lawfulness" of artificial intelligence.[1044] Legally binding documents are thus required for ensuring a "lawful" artificial intelligence.[1045] The EU Ethics Guidelines for Trustworthy AI refer to the rule of law and democracy as preconditions for reaching a "trustworthy AI".[1046] Although lawfulness is mentioned as a component of a "trustworthy AI" that should exist throughout the entire lifecycle of an AI-system, the EU Ethics Guidelines for Trustworthy AI falls short of mentioning the need for legal measures to ensure the lawfulness. [1047] Moving from the notion of "trustworthy AI", Smuha et al. suggest a "legally trustworthy" AI.[1048] Because the authors rightly suggest that merely having legal norms on artificial intelligence does not always ensure lawfulness in addressing it, it is crucial for the legal norms to address the associated risks adequately.[1049] For an AI-system to be "legally trustworthy," firstly, it is essential to appropriately assigning and distributing responsibility for wrongs and harms, extending to effective protection for human rights.[1050] Secondly, the need for a legal framework to prevent the harms that may result from artificial intelligence and, in the event of their occurrence, to ensure responsibility is advocated when discussing a "legally trustworthy AI".[1051] The realization of lawful artificial intelligence is closely related to legally binding regulations, which the EU AI Act Proposal that is going to be discussed in detail below, aims to achieve.[1052] EU AI Act Proposal acknowledge that non-binding documents do not bring enough protection mechanisms for problems may emerge from artificial intelligence.[1053] The EU AI Act Proposal is such a move in that direction to regulate artificial intelligence and fill the

[1042] Muller and The CAHAI Secretariat, *supra* note 90 at 32.
[1043] *See* Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, *supra* note 13.
[1044] Smuha et al., *supra* note 106 at 4.
[1045] *Id.*
[1046] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 9; Smuha et al., *supra* note 106 at 5.
[1047] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 5; Smuha et al., *supra* note 106 at 4.
[1048] Smuha et al., *supra* note 106 at 1, 4–9.
[1049] *Id.* at 5.
[1050] *Id.* at 6.
[1051] *Id.* at 7.
[1052] *Id.* at 5.
[1053] *Id.* at 2.

lawfulness cavity.[1054] International legally binding instruments for regulating artificial intelligence instead of non-binding documents would help to overcome the risk of "regulatory capture" in the artificial intelligence field that could arise in the case of non-binding documents.[1055]

The discussions on the regulation of artificial intelligence reflect a tension between utilizing the benefits and protecting human values, human rights, and other ethical and legal norms while adopting artificial intelligence technologies.[1056]

The hesitations towards regulating artificial intelligence with an EU-wide regulation include a potential chilling effect on competition.[1057] The EU AI Act Proposal's potential cost implications are noteworthy as there are concerns about the expensiveness of it.[1058] It is estimated that the EU AI Act will cost 10.9 billion euros per year by 2025, and until then, it will cost 31 billion euros, causing a reduce in benefits at 40% for an EU business with 10 million euros in turnovers if it deploys a high-risk AI-system.[1059] Criticisms toward the EU's approach to regulating artificial intelligence includes falling behind the United States and China for artificial intelligence advancements, arguing that introducing additional rules and restrictions through regulation could further lead the EU stay back in the race for artificial intelligence.[1060] In fact, concerns for market competition and economics are also noticed in the EU's approach to regulating artificial intelligence, as reflected in the White Paper for Artificial Intelligence.[1061] Not limiting intellectual property while regulating artificial intelligence is another concern reflected in the EU AI Act Proposal.[1062]

Similar worries are also present in the CoE's approach to regulating artificial intelligence; for example, the CAHAI's Feasibility Study underlines when discussing the

---

[1054] *Id.* at 4.
[1055] CHESTERMAN, *supra* note 57 at 203; Shaw and Kerrigan, *supra* note 1020 at 407.
[1056] AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517 at 13; Smuha et al., *supra* note 106 at 10.
[1057] BENJAMIN MUELLER, *How Much Will the Artificial Intelligence Act Cost Europe?*, 16 12 (2021), https://www2.datainnovation.org/2021-aia-costs.pdf (last visited Feb 26, 2023).
[1058] *Id.* at 4.
[1059] *Id.* at 4.
[1060] *Id.* at 11; DANIEL CASTRO & MICHAEL MCLAUGHLIN, *Who Is Winning the AI Race: China, the EU, or the United States? — 2021 Update*, 49 1 (2021), https://www2.datainnovation.org/2021-china-eu-us-ai.pdf.
[1061] European Commission, *supra* note 12 at Explanatory Memorandum 3; See Smuha et al., *supra* note 106 at 12; EUROPEAN COMMISSION, *supra* note 519 at 3–4; See MCKINSEY & COMPANY, *Shaping the Digital Transformation in Europe - Working Paper: Economic Potential*, 3 (2020), https://digital-strategy.ec.europa.eu/en/library/shaping-digital-transformation-europe-working-paper-economic-potential.
[1062] European Commission, *supra* note 12 at 11; See Smuha et al., *supra* note 106 at 48.

transparency principle that 'technical burdens' should not prevent the use and development of artificial intelligence systems and market opportunities.[1063] This study upholds that although profits and technological innovation are indeed substantial, they should not come at the expense of human rights and hinder equality in a society. The reasons for regulating artificial intelligence include contributing to adequate competition in the artificial intelligence field and establishing legal limits on problematic uses of artificial intelligence.[1064] Even if one considers accepting technological innovation as a legitimate justification for restricting human rights, it should not lead to completely overriding them.[1065] It is natural for legally binding documents not to have such a restrictive approach in regulating artificial intelligence to refrain from anti-technology measures. However, if there are going to be legal documents for artificial intelligence, they should provide implementable measures and legal consequences so that these do not turn into "law-washing". After all, what differentiates legally binding rules from ethics rules is their legal applicability.

As explored above, the non-binding documents aim to incorporate technical measures for artificial intelligence. While taking technical measures by the artificial intelligence industry is plausible, the entirety of solutions should not rely on them.[1066] The non-binding documents do not impose legally enforceable obligations.[1067] Those produced by the private sector for their own may lack independent outside evaluation mechanisms on their compliance.[1068] These factors contribute to the concerns of "ethics-washing" about the non-binding documents.[1069] The non-binding documents could become a tool to bypass concerns in the public regarding artificial intelligence by bringing about ethics and postponing the legal and regulatory means for addressing artificial intelligence related challenges.[1070] In addition to deliberate ethics-washing, it might also

[1063] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 34.
[1064] *Id.* at 13; Pekka Ala-Pietilä & Nathalie A. Smuha, *A Framework for Global Cooperation on Artificial Intelligence and Its Governance*, *in* REFLECTIONS ON ARTIFICIAL INTELLIGENCE FOR HUMANITY 237, 7 (Bertrand Braunschweig & Malik Ghallab eds., 2021), https://doi.org/10.1007/978-3-030-69128-8_15 (last visited Jun 14, 2023).
[1065] Smuha et al., *supra* note 106 at 12.
[1066] YEUNG, *supra* note 518 at 70.
[1067] *Id.* at 51; Shaw and Kerrigan, *supra* note 1020 at 407.
[1068] YEUNG, *supra* note 518 at 51.
[1069] *Id.* at 51; Wagner, *supra* note 1020 at 87; Thomas Metzinger, *EU guidelines: Ethics washing made in Europe*, DER TAGESSPIEGEL ONLINE, Aug. 4, 2019, https://www.tagesspiegel.de/politik/ethics-washing-made-in-europe-5937028.html (last visited May 16, 2023).
[1070] Metzinger, *supra* note 1069.

occur unintentionally due to illiteracy and negligence in ethics and law.[1071] Another form of "ethics-washing" is choosing from a variety of ethics principles that is most appropriate for the relevant artificial intelligence actor, to suggest their compliance with ethics that is called "ethics-shopping."[1072] Sometimes non-binding documents may be tools for marketing more than tools for compliance.[1073] The "ethics-washing" could also bring the discussions in the first chapter on fairness, in that direction, the non-binding documents could lead to "fairwashing", which indicates a presence of seemingly fair AI-system when, in reality, it may not be the case.[1074] If the non-binding documents are backed with institutional mechanisms that provide independent and impartial oversight and with the law, they would have more consequential protection to control the risks of artificial intelligence.[1075] Yeung suggests a roadmap for AI-systems to observe human rights and not end up in ethics-washing based on four principles: "*design and deliberation; assessment, testing, and evaluation; independent oversight, investigation, and sanction; and traceability, evidence, and proof.*"[1076]

The non-binding documents could be discretionary or pointless instead of functional and substantive tools to handle artificial intelligence if the concerns about "ethics-washing" or "ethics-shopping" are not appropriately addressed.[1077] In parallel, regulating artificial intelligence only with a variety of non-binding documents could yield giving preference to different ethical values in different jurisdictions, whereas the right to equality and non-discrimination in the context of artificial intelligence would require an international approach.[1078] The differences between the non-binding documents, even the ones referring to human rights such as the OECD AI Recommendation and the EU AI Ethics Guidelines, in terms of being of legally binding nature and enforceability is emphasized by some authors.[1079]

[1071] Eliot, *supra* note 992.
[1072] *Id.*; Wagner, *supra* note 1020 at 86.
[1073] CHESTERMAN, *supra* note 57 at 202; Hagendorff, *supra* note 1022 at 10; YEUNG, *supra* note 518 at 101.
[1074] Eliot, *supra* note 992; Ulrich Aivodji et al., *Fairwashing: The Risk of Rationalization*, *in* PROCEEDİNGS OF THE 36TH INTERNATİONAL CONFERENCE ON MACHİNE LEARNİNG 161, 2 (2019), https://proceedings.mlr.press/v97/aivodji19a.html.
[1075] YEUNG, *supra* note 518 at 51.
[1076] Yeung, Howes, and Pogrebna, *supra* note 529 at 86–89.
[1077] Wagner, *supra* note 1020 at 87.
[1078] Shaw and Kerrigan, *supra* note 1020 at 407.
[1079] *Id.* at 408.

The ethics or non-binding documents should not be considered a substitute for human rights obligations but support them by emphasizing the importance of human rights in the context of artificial intelligence, and support improving human rights.[1080] As Wagner emphasizes, if *"ethics are seen as an alternative to regulation or as a substitute for fundamental rights, both ethics, rights and technology suffer."*[1081] Since artificial intelligence may produce outcomes as decisions and applies them and thus impacts society, some authors suggest applying the rule of law requirements to artificial intelligence.[1082] In that direction, artificial intelligence systems would have to go through an assessment to evaluate legitimate processes abiding by human rights and democracy.[1083] The lack of enforceability and effective governance frameworks indicates the failure of ethical approach to artificial intelligence.[1084] Approaching different techniques and methods from different disciplines that are brought together to comply with human rights could help in a comprehensive way encompassing design and governance for addressing artificial intelligence. [1085] All being said, mentioning the concerns for "ethics-washing" and going beyond ethical handling of artificial intelligence should not correspond to "ethics-bashing", which means undermining the meaning and devaluing of ethics of artificial intelligence.[1086] The need for legal enforceability does not need to set aside the importance of philosophical methods in the artificial intelligence field.[1087]

When the approaches to regulation of artificial intelligence in the legal realm is examined, the forthcoming legally binding documents for regulating artificial intelligence, both from the CoE and the EU, picture a risk-based approach to artificial intelligence systems, their underlying reason being ensuring the promotion of artificial intelligence innovation and benefits while having the capability to address the risks and

[1080] SPECIAL RAPPORTEUR ON FREEDOM OF OPINION AND EXPRESSION, *supra* note 794 at 17.
[1081] Wagner, *supra* note 1020 at 87.
[1082] Paul Nemitz, *Constitutional Democracy and Technology in the Age of Artificial Intelligence*, 376 PHILOS. TRANSACT. A MATH. PHYS. ENG. SCI. 20180089, 13 (2018).
[1083] *Id.*
[1084] Yeung, Howes, and Pogrebna, *supra* note 529 at 101.
[1085] *Id.*
[1086] Elettra Bietti, *From Ethics Washing to Ethics Bashing: A View on Tech Ethics from within Moral Philosophy*, *in* PROCEEDINGS OF THE 2020 CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 210, 211 (2020), https://dl.acm.org/doi/10.1145/3351095.3372860; Eliot, *supra* note 992.
[1087] Bietti, *supra* note 1086 at 218.

harms that may result through artificial intelligence.[1088] The risk-based approach is preferred because artificial intelligence's risks for human rights come in a different range, depending on factors including the application context, related technology, and stakeholders.[1089] The CAHAI suggests that the risk-based approach protects societal well-being, which may relate to protecting equality on a societal level.[1090] Because, artificial intelligence innovation is also aimed at benefiting society, and the risk-based approach is argued by saying that categorizing the risks of artificial intelligence helps to enjoy the benefits of artificial intelligence while getting ahead of the risks thereof to ensure societal well-being.[1091] However, it should be noted that CoE Draft Convention on AI differs in its approach to the risk-based system; it does not explicitly mention categories of risks and instead mentions systems used in contexts involving human rights in Article 4.[1092] On the other hand, the risk-based approach is criticized for being susceptible to undermining artificial intelligence's potential risks and impacts to risks that are technical, mathematically measurable, and administrative, which could yield human rights protection in an on-the-surface and technocratic way.[1093] The human rights interference regime under the EU law may face challenges since the EU AI Act Proposal is highly affected by economic and innovation concerns and because of the adoption of the risk-based approach to AI Systems.[1094] The risks of being sufficient with the risk-based approach really could be in two ways. First, more than the solutions of the risk-based approach may be needed to address the problems that may emerge.[1095] Secondly, human rights issues could emanate from AI systems that are not classified as high-risk AI systems but still carry risks against human rights; in such cases, the risk-based approach could result in overlooking those and neglecting comprehensive protection for human rights.[1096] In such a case, since there are no obligations as the high-risk ones under the

[1088] AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517 at 12–13; EUROPEAN COMMİSSİON, *supra* note 519 at 17.
[1089] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 12.
[1090] *Id.* at 12–13.
[1091] *Id.* at 12 The societal well-being is essential in ensuring equality in artificial intelligence related contexts and discussed more in detail under the heading 3.2.5 of this study.
[1092] Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, *supra* note 13 at Article 4.
[1093] Smuha et al., *supra* note 106 at 11.
[1094] *Id.* at 10–11.
[1095] *Id.* at 11.
[1096] *Id.* at 13.

EU AI Act Proposal for the second type of AI systems, there could be more hardships in addressing human rights issues.[1097]

A legal framework for artificial intelligence would also uphold the rule of law as it would enforce legal rights and obligations and ensure clarity and coherence in the law.[1098] The rule of law requires coherency and consistency in legislation, judicial protection, and efficient enforcement of laws.[1099] The rule of law is closely related to legal clarity, ensuring that what is legal and not is known and understanding the potential consequences of these actions; the principle of no punishment without the law is such an example.[1100] "Legally trustworthy AI," thus, calls for such a regulatory framework with enforcement mechanisms, providing procedural rights.[1101] Legal regulations bring binding norms along with enforcement and monitoring mechanisms.[1102] Approaching artificial intelligence from a legal standpoint also demands regulatory documents to reflect a democratic commitment through principles such as transparency and accountability. [1103] This indicates a conjunction between the rule of law requirements and the principles developed for overseeing artificial intelligence. Therefore, adherence to these principles also contributes to keeping the rule of law. The rule of law requires principles such as proportionality and justification when intervening with human rights.[1104] One should also remember that equality is also a principle for upholding the rule of law.[1105] Artificial intelligence may while potentially contributing to rule of law, may also damage legal authority.[1106] Addressing artificial intelligence through legally binding documents is then required as a natural consequence of the rule of law. The relationship between artificial intelligence and the rule of law is increasingly being addressed by organizations such as the EU and the CoE.[1107] Authors suggest that to uphold the rule of law in artificial intelligence, human rights should be brought into the artificial intelligence context.[1108]

[1097] *Id.*
[1098] *Id.* at 6.
[1099] *Id.* at 7.
[1100] *Id.* at 8.
[1101] *Id.*
[1102] *Id.* at 5.
[1103] *Id.* at 6.
[1104] Muller and The CAHAI Secretariat, *supra* note 90 at 31.
[1105] *Id.*
[1106] *Id.*
[1107] *Id.* at 32; EUROPEAN COMMISSION, *supra* note 519 at 1.
[1108] Muller and The CAHAI Secretariat, *supra* note 90 at 32.

The AI-ethics perspective would look at problems that can be arisen from the lack of the principles for artificial intelligence and the problems deriving from that as ethical problems interfering with the moral life calling to be addressed in ethics codes and standards, whereas the legal perspective would look at those as interfering with the legally protected human rights and address from the legally binding documents.[1109] As explained, governing artificial intelligence with non-binding regulations, although a positive step, is not enough. The main reasons are the risk of "ethics-washing" with the non-binding documents and, secondly, the rule of law requiring to deal with artificial intelligence with legal rules specific to artificial intelligence. In particular, this is essential for protecting human rights and the right to equality and non-discrimination, both an overarching principle and a right to be duly ensured in artificial intelligence. Using legally binding requirements for artificial intelligence that include fairness, accountability, and transparency would support states and motivate the private sector to develop and use AI-systems conforming to the rule of law.[1110] As mentioned recurrently, human rights law can still be applied to artificial intelligence as it is. However, the question here is whether there is a need to legally reorganize artificial intelligence in terms of the right to equality and non-discrimination rather than whether there is a need to regulate artificial intelligence in general through legally binding rules, and the answer to that is yes. Now, the most prominent forthcoming legally binding documents will be examined.

3.1.3 Forthcoming Legal Documents for Regulating Artificial Intelligence

European Union Regulation Preparations

Given that the EU is a pioneer in regulating different domains with global impact, such as competition and data protection, it is expected that the EU will be a pioneer in regulating artificial intelligence as well. The EU's journey to regulating artificial intelligence is a lengthy one.[1111] When the EU AI Act Proposal was finally published, it

[1109] Paula Boddington, *Normative Modes: Codes and Standards*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 125, 134 (Markus D. Dubber, Frank Pasquale, & Sunit Das eds., 2020), https://doi.org/10.1093/oxfordhb/9780190067397.013.7 (last visited May 15, 2023).
[1110] Muller and The CAHAI Secretariat, *supra* note 90 at 34.
[1111] European Parliament, *Proposal for a Regulation on a European Approach for Artificial Intelligence | Legislative Train Schedule*, EUROPEAN PARLIAMENT (2023), https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence (last visited May 26, 2023).

had met with both applause and criticism.[1112] While the EU AI Act Proposal will be examined in the second section of this chapter through the principles, it is important to highlight certain points initially. The EU AI Act Proposal is accompanied by a 15-pages Explanatory Memorandum that outlines the reasons behind the necessity for regulation, and a lengthy preamble section of around 20-pages. The EU AI Act Proposal contains a total of thirteen titles. The Proposal starts with definitions, including the definition of AI, which is mentioned in the first chapter of this study. Title II outlines prohibited AI practices, while Title III focuses on high-risk AI-systems. Title IV introduces certain transparency obligations, and Title V addresses innovative measures. Titles VI to VIII are reserved for governance and implementation. Title IX lays out the requirements for codes of conduct. Finally, Titles X to XIII contains the final provisions, with Title X handling confidentiality. The remaining two titles are dedicated to the Proposal's working methods and further implementation details. The EU AI Act Proposal aims to complement the existing legislative framework of the EU, for assuring consistency among the relevant union-wide regulations.[1113] The explanatory memorandum further details the legal basis for the proposal, highlighting that Article 114 of the TFEU provides the authority to adopt measures for the effective functioning of the EU internal market.[1114] The Explanatory Memorandum of the proposal contains a subheading reserved to fundamental rights, which references articles in the EU Fundamental Rights Charter, encompassing the right to equality and non-discrimination, among other rights.[1115] It further explains the principle of subsidiarity and why regulating artificial intelligence is required on a union-wide level instead of a national level by explaining that nationwide policies may lead to legal uncertainty.[1116]

Proportionality is the primary approach in the EU AI Act Proposal that supposedly aims to interfere in AI-systems when there is a risk to human rights, which righteously

[1112] *See* for a detailed discussion on the EU's AI Act Proposal Smuha et al., *supra* note 106; *See also* for the collection of feedbacks offered for the EU AI Act Proposal Artificial intelligence – ethical and legal requirements: Feedback and statistics: Proposal for a regulation, EUROPEAN COMMISSION (2021), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback_en?p_id=24212003 (last visited May 20, 2023).
[1113] European Commission, *supra* note 12 at 4–5.
[1114] *Id.* at 6.
[1115] *Id.* at 11.
[1116] *Id.* at 6.

receives criticism.[1117] Although one of the aims behind this regulation is to *"ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values"* according to the Explanatory Memorandum, the EU AI Act Proposal do not have a systematic human rights approach.[1118] The EU AI Act Proposal's handling of human rights requires having the necessary standpoint of human rights law instead of approaching it technically by almost negating it to technical standards.[1119] The EU AI Act Proposal is righteously criticized for neglecting the unique nature of human rights.[1120] That the EU AI Act Proposal does not apply to AI-systems that are *"developed or used exclusively for military purposes"* according to Article 2(4) is another provision that could potentially impact the level of protection for human rights with the EU AI Act Proposal.[1121] This study supports the opinion that the EU AI Act Proposal's offering for protecting human rights is insufficient and needs to be revised.[1122] Under the EU law, interference with human rights is subject to criteria similar to those outlined in the ECHR. Limitations on fundamental rights should meet the requirements outlined in Article 52 of the EU Fundamental Rights Charter.[1123] Thus, prioritizing concerns for the pursuit of economic and innovative development may lead to the failure of an adequate balance for human rights.[1124] Smuha et al. advocate for legally enforceable standards that protect against the abuse of power due to artificial intelligence, that may lead to positive obligations.[1125]

The EU AI Act Proposal offers a regime of risk-based categorization to AI-systems under Article 6.[1126] Article 6 of the EU AI Act Proposal mentions high-risk AI-systems

---

[1117] *Id.* at 7.
[1118] *Id.* at 3.
[1119] Smuha et al., *supra* note 106 at 9.
[1120] *Id.* at 9–10.
[1121] See *Id.* at 18.
[1122] See *Id.* at 3.
[1123] The EU Charter of Fundamental Rights of the European Union, *supra* note 354 at 52 Article 52.1 thereof states: *"Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others."* ; Smuha et al., *supra* note 106 at 10.
[1124] Smuha et al., *supra* note 106 at 10.
[1125] *Id.*
[1126] European Commission, *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, Article 6 (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (last visited Aug 1, 2022) The first scenario for being categorized as high-risk AI system according to Article 6.1 is: _*"Irrespective of whether an AI system is*

classification under two cases existing together, one being listed under Annex II and being a safety component of a product, or is a product, and the second being:

> "the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product under the Union harmonization legislation listed in Annex II."

Another possibility for being considered a high-risk AI-system is to be listed in Annex III. Article 7 allows for an amendment to Annex III, which aims to keep up with technological developments and state of the art. The high-risk AI-systems mentioned in Annex III are the ones that may the violations of the right to equality and non-discrimination may occur the most as these extend to areas including law enforcement, employment and education.[1127] The list in Annex III is open for a yearly assessment for a need to amendment according to Article 84 of the EU AI Act Proposal.[1128] As mentioned earlier, critics argue that risk-based approach under the EU AI Act Proposal falls short of accordance with the human rights law.[1129] It is pointed out as a failure of the EU AI Act in recognizing the uniqueness of human rights, which could lead to undermining human rights instead of protecting them.[1130] More explanations based on the articles envisioned in the EU AI Act Proposal will be dealt through the principles.

---

placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled: _(a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II; _(b) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II". ; Annex II is a selected list of Union Harmonization Legislation European Commission, *Annexes to the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, Annex II (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (last visited Aug 1, 2022); The second scenario to be categorized as high risk AI system according to Article 6.2 is: *"(...) AI systems referred to in Annex III (...)". Id.* at Annex III

[1127] European Commission, *Annexes to the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, Annex III (2021), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (last visited Aug 1, 2022) Annex III lists high-risk AI systems according to different areas of use. These are *"1. Biometric identification and categorisation of natural persons; 2. Management and operation of critical infrastructure; 3. Education and vocational training; 4. Employment, workers management and access to self-employment; 5. Access to and enjoyment of essential private services and public services and benefits; 6. Law enforcement; 7. Migration, asylum and border control management; 8. Administration of justice and democratic processes".*

[1128] European Commission, *supra* note 12 at Article 84.

[1129] Smuha et al., *supra* note 106 at 12.

[1130] *Id.* at 10.

The May 2023 report of the MEPs suggesting amendments to the EU AI Act Proposal seems to address some of the concerns, such as the expansion of high-risk classification under Annex III to include systems presenting *"harm to people's health, safety, fundamental rights or the environment,"* and systems that could influence voters in democratic processes with political campaigns and social media recommendations.[1131]

The necessity of artificial intelligence systems to be designed and used to comply with human rights standards requires examining artificial intelligence from the human rights law perspective. [1132]

Council of Europe Framework Convention Preparations

The CAHAI has emphasized the necessity for a legally binding document pertaining to artificial intelligence.[1133] The Committee's successor, the CAI, has published its initial draft for a framework convention, the CoE Draft Convention on AI.[1134] The CoE Draft Convention on AI differs from the EU AI Act Proposal due to approaching the subject with a focus specifically on human rights aspects of artificial intelligence.

The CoE Draft Convention is at an early stage at the time of this thesis, meaning that it is potentially subject to future changes before reaching its finalized form. It comprises 38 articles spread into eight chapters. Chapter I therein handles general provisions: definitions, and scope and introduces non-discrimination as a superseding principle for the convention. Chapter II brings obligations and requirements for human rights, respect for democratic institutions, and the rule of law, dealing with the *"application of artificial intelligence systems by public authorities."* In contrast, Chapter III conveys obligations to both public and private actors with regard to AI-systems *"in the provision of goods, facilities, and services."* Chapter IV presents the *"fundamental principles of design, development, and application of artificial intelligence systems,"* which are addressed in

[1131] AI Act: a step closer to the first rules on Artificial Intelligence | News | European Parliament, (2023), https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence (last visited May 11, 2023); BRANDO BENİFEİ & IOAN-DRAGOŞ TUDORACHE, *Draft Compromise Amendments on the Draft Report - Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, 117, 125 (2023), https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf.
[1132] Yeung, Howes, and Pogrebna, *supra* note 529 at 89.
[1133] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *Possible Elements of a Legal Framework on Artificial Intelligence, Based on the Council of Europe's Standards on Human Rights, Democracy and the Rule of Law*, 13 7 (2021), https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d (last visited Aug 4, 2022).
[1134] See Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, *supra* note 13.

the second section of this chapter of the present study. Chapter V specifically addresses measures and safeguards to ensure accountability and redress. The CoE Draft Convention also foresees a regime for managing the risks and adverse impacts through assessment and mitigation in Chapter VI, succeeded by a follow-up mechanism laid out in Chapter VII. Lastly, Chapter VIII concerns other relevant provisions, such as the effects of the convention, amendments, or accession.

Article 1(1) of the CoE Draft Convention sets the convention's purpose and aim that is *"ensuring that design, development, and application of artificial intelligence systems are fully consistent with respect for human rights, the functioning of democracy and the observance of rule of law."* Article 1(2) mentions a positive obligation to states to take necessary measures in national legislation to give effect to the principles, rules, and rights set out in the convention. The reasoning behind this may be leaving room for states to implement the convention following their needs. However, this may be one of the reasons why the convention is not too strict. Article 2 includes relevant definitions for artificial intelligence systems, which were explored in the first chapter of this study.

Within the context of the CoE Draft Convention, there are three main sides when using an AI-system, the provider, the user, and the subject. An artificial intelligence provider can be a natural or legal person, a public authority, or another entity involved in developing an artificial intelligence system intending to put it into service or commission it, as defined under Article 2(c). Therefrom, it is apparent that an artificial intelligence provider's meaning is limited in developing AI-systems. Secondly, an artificial intelligence user can also be a natural or legal person, a public authority, or another body as defined under Article 2(d). According to the said article, the key criterion for being classified as an artificial intelligence user is utilizing the AI-system in one's name and under their authority. There are no specific limitations in the article for purposes such as commercial purposes for using AI-systems. That means anyone using an AI-system in their daily life may potentially be deemed an artificial intelligence user. Thirdly, the artificial intelligence subject may be natural and legal persons whose human rights impacted through artificial intelligence according to Article 2(e). Since this definition of an artificial intelligence subject requires an impact on the rights or freedoms of a person through the use of artificial intelligence, it implies that if someone's rights or freedoms are not affected, that person will not qualify as an artificial intelligence subject. That may

either be a conscious choice to limit the scope of the CoE Draft Convention on AI with human rights or stem from the idea that artificial intelligence will affect or impact human rights in all cases.

Article 4 of the CoE Draft Convention on AI delivers the scope of the convention. The subparagraph one hereunder adopts a contextual approach for its scope, that being *"a context involving issues relating to the respect for human rights, the functioning of democracy and the observance of the rule of law (...)"*[1135] Subparagraph two mentions that this article applies to both public and private actors. Subparagraph three completely excludes applicability of the convention to situations involving national defense. This is possibly put for the states not to hesitate when signing the convention due to national security concerns, though complete exclusion of national defense systems is quite broad since there could be many human rights problems, particularly for discrimination when using AI-systems in national defense.[1136]

Article 22 of the CoE Draft Convention on AI states that it does not limit or derogate the rights and obligations outlined in other human rights instruments. Additionally, Article 23 of the CoE Draft Convention on AI allows parties to adopt broader measures they see fit. Therefore, states may take or implement better mechanisms for protecting equality in the context of artificial intelligence.

Article 30(2) of the CoE Draft Convention on AI is noteworthy as it makes a reference to the EU and declares that the EU Member parties of the CoE Draft Convention on AI should apply the EU law where applicable, *"without prejudice to the object and purpose of the present Convention and without prejudice to its full application with other Parties."* This provision reveals the cooperation of the CoE and the EU for artificial intelligence, as it is apparent that the CoE Draft Convention on AI aims to work together and adapt to the EU regulations.[1137]

That could be worthwhile to handle the lack of human rights specific provisions in the EU AI Act Proposal, considering all of the EU Members are members of the CoE. One question regarding the CoE Draft Convention on AI is the absence of specifications

---

[1135] See Ian Barber, *First Thoughts on the Revised Zero Draft of the Council of Europe's AI Treaty | Global Partners Digital*, GLOBAL PARTNERS DIGITAL, Apr. 2023, https://www.gp-digital.org/first-thoughts-on-the-revised-zero-draft-of-the-council-of-europes-ai-treaty/ (last visited May 10, 2023).
[1136] See *Id.*
[1137] The EU also refers to the CoE's efforts in artificial intelligence in the EU White Paper on AI. EUROPEAN COMMISSION, *supra* note 519 at 11.

on which human rights it protects. Instead, the CoE Draft Convention on AI leaves that to states' national laws, whichever human rights they accept and implement. Thereto may be different reasons, as the CoE Draft Convention on AI is firstly a treaty under the CoE, although being open for signatures from other states and the EU. First, most of the signatories will come from the CoE, of which all members are parties to the ECHR. Further, the EU is also a potential signatory to this convention, where all 27 member states are, as stated before, parties to the ECHR due to being a member of the CoE.

A second reason could be increasing the number of parties to the CoE Draft Convention on AI as much as possible because if this treaty were to present a body of human rights relevant to artificial intelligence, the focus could have been shifted from artificial intelligence, as that could lead to hesitation about some of the rights. A third reason could be to reach a flexible and adaptative convention in order to avoid leaving out a human right that may not seem relevant at the time but later become so. Another reason could be the aim of having a more inclusive approach by addressing the discussion of human rights around the national laws of different jurisdictions, leaving states room to apply their set of human rights. The CoE Draft Convention on AI has the potential to serve as a mechanism for adapting current human rights law in order to provide protection for human rights in the context of artificial intelligence.

At this point, various principles necessary to oversee and manage artificial intelligence stand out in both non-binding and binding regulations, which are necessary to oversee and manage artificial intelligence. In the second section of this chapter, how the right to equality and non-discrimination can be governed in the context of artificial intelligence will be examined through these principles. In this respect, when discussing the theoretical meanings and scopes of the principles, approaches to it in the non-binding documents are valuable and give critical insights. For this reason, while the principles are being discussed, how these principles are handled in the legally binding and non-binding documents will be presented together.

3.2 Pursuing the Right to Equality and Non-Discrimination through the Principles in Artificial Intelligence Regulations

Exploration of the principles for governing artificial intelligence, firstly demands mentioning some points. As mentioned earlier, analysis of the regulation of artificial

intelligence is limited by the scope of the thesis, the right to equality, and non-discrimination. The principles developed for governing artificial intelligence are not limited in numbers and can be organized in various ways. The principles hereunder are prominent in academia, and the non-binding and binding regulatory works on artificial intelligence. For protecting human rights -including the right to equality and non-discrimination in artificial intelligence, principles of transparency, explainability and accountability are essential.[1138] These principles should exit throughout the lifecycle of artificial intelligence.[1139] The requirements of a principle may function as conditions for the realization of other principles.[1140] As will be explored, equality, fairness, non-discrimination, and related notions come up as separate principles in these documents.

It is essential to recognize that all the other principles are also steps towards facilitating the right to equality and non-discrimination. These principles are not limited to the human rights dimension of artificial intelligence, they could also be relevant for example, in private law related matters. However, applying these principles also come to the fore when evaluating a problem in artificial intelligence relevant to human rights law. The content and promises of the principles could reflect requirements for an AI-system to be non-discriminatory.[1141] For example, if an AI-system lacks transparency or safety, revealing discrimination for a court when examining discrimination claims regarding artificial intelligence could be impossible. [1142] Likewise, if there is no human oversight over an AI-system, it could result in uncontrollably biased outcomes.[1143] To realize accountability, the explainability of the decisions is required, meaning that different principles complement each other.[1144] In addition, measures under the accountability principle could support accountability under human rights law.[1145] If the requirements under transparency, such as auditability and traceability, or safety, such as accuracy in data, which are explained below, are fulfilled, then deciding on discrimination could be

[1138] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 1133 at 7.
[1139] *Id.*
[1140] FJELD ET AL., *supra* note 990 at 42; UNI GLOBAL UNION, *Top 10 Principles for Ethical Artificial Intelligence*, 10 7 (2017), http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (last visited May 22, 2023).
[1141] Smuha, *supra* note 19 at 101.
[1142] *Id.*; See Tanna and Dunning, *supra* note 128 at 436; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 24.
[1143] EUROPEAN COMMISSION, *supra* note 519 at 21; McGregor, Murray, and Ng, *supra* note 1 at 334.
[1144] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13.
[1145] McGregor, Murray, and Ng, *supra* note 1 at 311, 325.

possible in a court.[1146] Therefore, these principles could render positive procedural obligations for human rights, including the right to equality and non-discrimination.[1147]

In addition, if these principles support human rights, moving beyond ethical protection comprising safeguards and obligations can be possible.[1148] These principles can make the obligations of the states and help prevent the risk of private actors escaping from being accountable by defining the expectations to be fulfilled by the private actors with these principles.[1149] When these principles are fulfilled throughout the AI-lifecycle, better protection of human rights in the context of artificial intelligence becomes attainable.[1150] As these principles are not apt to provide complete protection for human rights by themselves, there should be an understanding of supporting and realizing them with a lens of human rights law.[1151] In this regard, human rights impact assessments could be relevant for ensuring these principles are fulfilled throughout the AI-lifecycle.[1152] For example, detecting indirect discrimination could be possible with an impact assessment conducted prior to start using an AI-system.[1153] Technical fulfillment of these principles with metrics in the AI-systems would also contribute to closing the gap between the technical and legal understanding of equality, fairness, and discrimination.[1154] Adequate usage of artificial intelligence in different areas explained in the previous chapter, such as health and education, requires the principles that will be discussed to be fulfilled.[1155] Human rights protection should contribute to these principles by ensuring their accountability, and these principles should also contribute to human rights protection by providing a map for obligations.[1156]

[1146] Smuha, *supra* note 19 at 101; Tanna and Dunning, *supra* note 128 at 436; EUROPEAN COMMISSION, *supra* note 519 at 19; *See also* McGregor, Murray, and Ng, *supra* note 1 at 338.
[1147] Smuha, *supra* note 19 at 101.
[1148] *Id.* at 102.
[1149] McGregor, Murray, and Ng, *supra* note 1 at 314.
[1150] *Id.*
[1151] *Id.* at 320.
[1152] *Id.* at 330.
[1153] *Id.* at 334.
[1154] Wachter, Mittelstadt, and Russell, *supra* note 19 at 66.
[1155] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 32–33.
[1156] McGregor, Murray, and Ng, *supra* note 1 at 311, 326, 342; Smuha, *supra* note 19 at 101.

3.2.1 Transparency and Explainability

Transparency and explainability are two principles that need to be interpreted concurrently, as they reflect two different aspects of the same idea.[1157] The transparency principle has particular importance for human rights because of the inability to examine the AI-systems and algorithms; examining human rights risks in them could also be challenging.[1158] Transparency requires the process of artificial intelligence from development to deployment to be traced and explained to the persons affected directly or indirectly by the artificial intelligence system, making oversight of the AI-system possible.[1159] Explainability refers to understandability and providing a perception on AI-systems.[1160] Transparency and explainability principles are found in most of the documents on artificial intelligence.[1161] This could strengthen their importance and suggest that these would be expected as requirements from the courts when evaluating a case with a discriminatory claim involving an AI-system to assess whether the system's transparency exists or how transparent it is. For states, it could give rise to positive obligations to verify whether the state has made efforts to ensure transparency, even when private actors are the users of the AI-systems.

Bias and instances of discrimination discussed in the first and second chapters may appear due to the common usage of algorithms in artificial intelligence systems without understanding the causative connections.[1162] Transparency does not mean having complete knowledge about every element of an AI-system, as that may not align with the objective of this principle.[1163] Here, explainability is a key that enables a description of the AI-system's outcome or decision, taking into account the relevant contexts.[1164] When the impact of the AI-system is significant, as with instances of discrimination, it is

[1157] See FJELD ET AL., *supra* note 990 at 41.
[1158] McGregor, Murray, and Ng, *supra* note 1 at 321.
[1159] VIRGINIA DIGNUM, CATELIJNE MULLER & ANDREAS THEODOROU, *Final Analysis of the EU Whitepaper on AI*, 16 9 (2020), https://allai.nl/wp-content/uploads/2020/06/ALLAI-Final-Analysis-of-the-EU-Whitepaper-on-AI-consultation.pdf; FJELD ET AL., *supra* note 990 at 42; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13, 18.
[1160] UNESCO, *supra* note 1004 at 9 para. 40.
[1161] FJELD ET AL., *supra* note 990 at 41; Ienca and Vayena, *supra* note 972 at 39, 50.
[1162] O'NEIL, *supra* note 21 at 162; WAGNER, *supra* note 19 at 37.
[1163] WAGNER, *supra* note 19 at 38; Joanna J Bryson & Andreas Theodorou, *How Society Can Maintain Human-Centric Artificial Intelligence*, *in* HUMAN-CENTERED DIGITALIZATION AND SERVICES 305, 12 (Marja Toivonen & Eveliina Saari eds., 2019).
[1164] WAGNER, *supra* note 19 at 38; UNESCO, *supra* note 1004 at 9 para. 40.

essential to provide proper and appropriate explanations to those impacted.[1165] Thus, for the right to equality and non-discrimination, transparency should be comprehended together with explainability rather than in isolation. Explainability can also be considered with the discussions on the discrimination test in the previous chapter. The underlying rationale of explainability is that since human decision-making is delegated to machines, humans require explanations for the decisions from which they are impacted.[1166]

The relation between transparency and explainability with contestability is highlighted by the EU, UNESCO, and CoE in their relevant documents.[1167] For example, if one is to claim that they have been exposed to discrimination by an AI-system, in case of no transparency, their claim could be held back, whereas a transparent AI-system would reinforce the possibility of a remedy in case of a discrimination.[1168] Since not knowing whether the decision that one is being discriminated against or another right is violated is not desirable, notifying an artificial intelligence based decision in human rights related circumstances is a must.[1169] Individuals should be able to know how to handle an artificial intelligence based decision.[1170] The EU Ethics Guidelines for Trustworthy AI outlines the transparency requirement with these three aspects and relates that with the principle of explicability.[1171] It calls for transparent data, systems, and business models.[1172]

Among the non-binding documents, the transparency principle under the OECD AI Recommendation provides an overview encompassing the requirements for transparency and explainability by emphasizing the right to meaningful information by the context and

[1165] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 18.
[1166] Boddington, *supra* note 1109 at 134; David Gunning & David Aha, *DARPA's Explainable Artificial Intelligence (XAI) Program*, 40 AI MAG. 44, 45 (2019), https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2850 (last visited May 22, 2023).
[1167] HİGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13; AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517 at 33; UNESCO, *supra* note 1004 at 9 para. 37; Council of Europe Committee of Ministers, *supra* note 991 at Principle B.4.3 and Principle C.4.3 under the "Guidelines on addressing the human rights impacts of algorithmic systems."
[1168] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13; DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 9; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 33; UNESCO, *supra* note 1004 at 9 para. 37.
[1169] FJELD ET AL., *supra* note 990 at 45; ACCESS NOW AND ANDERSEN, *supra* note 8 at 33.
[1170] FJELD ET AL., *supra* note 990 at 45; EUROPEAN COMMİSSİON, *supra* note 96 at 16.
[1171] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13, 18.
[1172] *Id.* at 18; FJELD ET AL., *supra* note 990 at 42.

technological state, promoting understanding of the AI-systems, raising awareness for all stakeholders, aiding to comprehension of the outcomes, and allowing contestability.[1173]

Traceability and Understandability

Traceability and understandability are components of explainability.[1174] Traceability allows for transparency from the data collection to decisions, clarifying what is wrong with the decision.[1175] Tracing is of critical importance for a case of discrimination through artificial intelligence, especially proof-wise. As mentioned in the second chapter, in a discrimination claim, if the claimant shows the difference in treatment, then the burden of proof shifts to the other party; traceability would allow for both parties in such a claim to exhibit their claims with proof. Tracing is also crucial for preventing future discriminatory situations as it would show the root cause of the discrimination.[1176] Traceability through logging and documenting the AI-system processes allows for revealing the AI-system's aim, capacity, and limits.[1177] Communication openly about the AI-system, such as notification when an AI-system is being employed for decision-making or when facing an AI-system, is also required for transparency.[1178] These principles make it possible to pinpoint the impact of an automated or artificial intelligence based decision on human rights.[1179] Transparency in this regard, is also a step for a democratic approach to artificial intelligence, as transparency will contribute to the public debate on AI.[1180] Transparency also helps to achieve substantive equality, contributing to inclusivity in societies as it would help to detect and prevent the discriminatory impacts on human rights that may arise from using (or not using) AI.[1181]

Since the reasoning, the ways in which an AI-system produces an outcome, and when and for what purposes artificial intelligence is used are only sometimes possible to explain

[1173] OECD, *supra* note 991 at Principle 1.3.
[1174] UNESCO, *supra* note 1004 at 9 para. 40.
[1175] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 18.
[1176] *Id.*
[1177] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 33; EUROPEAN COMMISSION, *supra* note 519 at 20.
[1178] DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 9; FJELD ET AL., *supra* note 990 at 41; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 18; Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 10.
[1179] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 9.
[1180] Smuha et al., *supra* note 106 at 8.
[1181] UNESCO, *supra* note 1004 at 9 para. 39.

due to opaque and complex systems that give rise to the black-box effect, achieving transparency may be challenging.[1182] In such situations, sharing the capabilities of an AI-system, assuring auditability and traceability deters the risks that may emerge due to the black-box effect.[1183] Nevertheless, a vital point to consider is why an AI-system is complex to such a degree that it is not available to be reviewed by humans.[1184] If there is a risk that transparency and accountability not be achieved, refraining from using artificial intelligence could be considered in circumstances where the risk is high.[1185] This consideration could also extend to specific areas of use; for example, if there is a risk of using artificial intelligence to result in discriminatory decisions when the absence of transparency in the AI-system is known, refraining from using artificial intelligence may be considered. Therefore, as stated in the EU Ethics Guidelines for Trustworthy AI, the level of "explicability" depends on the context and the gravity of the potential outcomes for which artificial intelligence is used.[1186] The UNESCO AI Recommendation stresses the importance of explainability and transparency together.[1187] The UNESCO AI Recommendation mention these principles in line with the protection and promotion of human rights.[1188] Explainability is considered a substantive right in the CAHAI's Feasibility study, which involves the right to meaningful explanation.[1189] States may be obligated to set traceability requirements to ensure transparency and explainability.[1190] At least a level of knowledge about the algorithm or the AI-system would be required to understand the training data, the objectives of the AI-system, and type and quantity of data concerning the right to equality and non-discrimination.[1191] As the Asilomar AI Principles asserts, explaining an AI-system is necessary for areas of artificial intelligence

[1182] DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 9; FJELD ET AL., *supra* note 990 at 41; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13.
[1183] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13; DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 9.
[1184] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 10.
[1185] Amnesty International and Access Now, *supra* note 1006 at Principle 32c; FJELD ET AL., *supra* note 990 at 42; Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 10.
[1186] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 13.
[1187] UNESCO, *supra* note 1004 at 9 para. 37.
[1188] *Id.* para. 37.
[1189] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 34.
[1190] *Id.*
[1191] WAGNER, *supra* note 19 at 38.

use, such as judicial systems, so that humans can overview them, making transparency and explainability also connect to the principle of human oversight.[1192]

Transparency Approaches in the Forthcoming Legal Documents

When the forthcoming legal documents are examined, transparency principle is mentioned together with oversight principle in the CoE Draft Convention on AI under Chapter II in Article 15 – "Principle of Transparency and Oversight" stating that:

> "Each Party shall, within its jurisdiction and in accordance with its domestic law, ensure that adequate oversight mechanisms as well as transparency and auditability requirements tailored to the specific risks arising from the context in which the artificial intelligence systems are applied are in place."

The CoE Draft Convention on AI obliges the states to ensure transparency, integrating it with oversight and auditability requirements for AI-systems within their jurisdiction. In this regard, the context in which the AI-system is being used is critical. A potential impact on the right to equality and non-discrimination is a significant factor in determining the context. Furthermore, considering the areas of use for AI-systems explored in the second chapter would be also helpful in determining the context.

On the other hand, the EU AI Act Proposal addresses transparency requirement in various provisions, including a direct provision on transparency and others relating thereto. Firstly, the transparency principle is regulated under Article 13 of the EU AI Act Proposal. This article envisages, *"high-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately."* Article 60 under Title VII of the EU AI Act Proposal reads, "EU database for stand-alone AI high-risk systems," which envisions a database containing information open to the public with maintaining the high-risk systems listed in Annex III. Such a database would be beneficial in maintaining the right to equality and non-discrimination by providing transparency and traceability and contributing to accountability, as the "standalone high-risk systems" listed under Annex

[1192] Future of Life Institute, *supra* note 991 at Principle 8; FJELD ET AL., *supra* note 990 at 43.

III consist of AI-systems that are most open to discrimination to transpire, overlapping with some areas of use for AI-systems.[1193]

Achieving Explainability

Explainability also makes examining the actual impacts of artificial intelligence on individuals available.[1194] For example, the biases impacting individuals can be disclosed and minimized if the AI-system is explainable.[1195] In addition to technical explainability, humans should also explain why artificial intelligence is being used in a particular area/matter.[1196] Accordingly, Ienca and Vayena diagnose two facets of transparency requirements in the artificial intelligence regulatory documents, one being for the transparency of the algorithms and data processing methods and secondly, transparency in the human practice throughout the AI-lifecycle.[1197] Transparency and explainability are of particular importance in cases with human rights risks.[1198] "Explainable Artificial Intelligence" is an entire field within artificial intelligence devoted to achieving transparency and explainability.[1199] When a decision is made that affects individuals, mainly if it results in unequal treatment, there is a need for explanations to be provided.[1200] The level of explanations may depend on the context in which the AI-system is used.[1201] Technical transparency aim to get ahead of black-box methods, while explainability aims to ensure that the technicalities and decisions of AI-systems are comprehensible.[1202] While technical transparency is crucial as it helps interpret the AI process and decisions and contributes to explainability, it is not enough; as noted by Ienca and Vayena, it is

[1193] The high-risk AI systems under Annex III of the EU AI Act Proposal consist of *"biometric identification and categorization of natural persons, management and operation of critical infrastructure; education and vocational training; employment, workers management and access to self-employment, access to and enjoyment of essential private services and public services and benefits; Law enforcement; migration, asylum, and border control management; administration of justice and democratic processes." See* for details European Commission, *supra* note 104 at Annex III.
[1194] Amnesty International and Access Now, *supra* note 1006 at Principle 32; FJELD ET AL., *supra* note 990 at 43.
[1195] FJELD ET AL., *supra* note 990 at 43; EUROPEAN COMMİSSİON, *supra* note 96 at 14.
[1196] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 18.
[1197] Ienca and Vayena, *supra* note 972 at 51.
[1198] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 33.
[1199] See for an overview of approaches to the exlainable artificial intelligence ('XAI') Gunning and Aha, *supra* note 1166.
[1200] Boddington, *supra* note 1109 at 134; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 34.
[1201] Boddington, *supra* note 1109 at 134; AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517 at 34; OECD, *supra* note 991 at Principle 1.3.
[1202] FJELD ET AL., *supra* note 990 at 42–43; Ienca and Vayena, *supra* note 972 at 51.

more of an approach in the private sector, and as discussed above, connecting transparency to accountability through oversight mechanisms is more of a legally binding approach.[1203] There is ambiguity in how to ensure transparency on the technical level, whether through releasing the databases or source codes of the AI-system.[1204] For example, the private sector, IBM's transparency principle accepts the weakness of the bias risk not being able to be removed entirely but mentions the effort to test the systems and find new data sets to align the outputs of the artificial intelligence with human values and expectations.[1205] While such a proposition may seemingly contribute to substantive equality, it lacks legal applicability and fails to provide a sufficient basis for accountability, despite noting explainability. Open source data and algorithms and open government procurement are additional recommendations for achieving transparency.[1206] Disclosing the source codes or algorithms of the AI-systems can make auditing of these available.[1207] However, publicizing algorithms may be impossible if left to the private sector due to intellectual property protections.[1208] Therefore, the protection provided by intellectual property rules may sometimes impede transparency and accountability.[1209] The CoE's "Resolution 2343 (2020) on Preventing Discrimination Caused by the Use of Artificial Intelligence" underlines that intellectual property can make proving discrimination challenging in the burden of proof-related matters as well as the presence of automation bias referring to assuming the correctness of artificial decisions, can be puzzling to overcome.[1210] Nevertheless, the CoE again considers intellectual property among national security to be balanced with the transparency and explainability principles; these two are seen as 'legitimate interests.'[1211] Intellectual property and business secrecy should not go beyond balancing other legitimate interests.[1212] If these principles become later legal principles, legitimate interests could extend in such a way, in line with the legitimate aim under the discrimination test. The EU AI Act Proposal in

[1203] Ienca and Vayena, *supra* note 972 at 51.
[1204] *Id.* at 39.
[1205] IBM, *IBM's Principles for Trust and Transparency*, IBM POLICY (2018), https://www.ibm.com/policy/trust-principles/; IBM, *supra* note 216 at 28.
[1206] FJELD ET AL., *supra* note 990 at 41–42.
[1207] *Id.* at 44; IEEE, *supra* note 991 at 26 Principle 5 Recommendation.
[1208] WAGNER, *supra* note 19 at 38.
[1209] Council of Europe Parliamentary Assembly, *supra* note 848 at 1 para. 5.
[1210] *Id.* para. 5.
[1211] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 33.
[1212] *Id.* at 34.

Article 70 has a measure that aims to maintain confidentiality when assessing AI-systems, extending to intellectual property, business, and trade secrets.[1213] Other reasons listed for confidentiality in a non-limited list includes effective implementation of the regulation; public and national security interests and integrity of the criminal or administrative proceedings.[1214] Confidentiality, not-sharing private data unless there is an explicit consent from the subject, may help to maintain human control over artificial intelligence.[1215]

Accordingly, sharing datasets and codes are limited with "serious threats" to human rights in some documents, as seen in UNESCO AI Recommendation, yet what may constitute a severe threat is unclear, which indicates the ambiguity in non-binding regulatory documents on artificial intelligence.[1216] The EU AI Act Proposal Article 11 is about technical documentation that must be done before the market placement and later kept up to date, enabling compliance assessments. This can also factor in reaching better transparency in AI-systems. The EU AI Act Proposal Article 12 is on record keeping which declares high-risk systems to be designed and developed so that automated logging is possible. This allows for traceability in the AI-lifecycle, appropriate to the intended purpose of the AI-system, as stated in Article 12(2).

Transparency further extends to the public and accessible use of AI-systems.[1217] Individuals should be able to fathom how a decision is made and confirmed.[1218] Where public authorities use artificial intelligence, the need for transparency may intensify.[1219] In public uses of AI, such as credit scoring systems and chatbots, as well as in the workplace, healthcare, judicial systems, and welfare systems, individuals should be aware when interacting with an AI.[1220] In such cases, the assistance of an AI-system with a human may also be required.[1221] However, the priority of transparency reveals itself even

[1213] European Commission, *supra* note 12 at Article 70.
[1214] *Id.* at 70.
[1215] Etzioni, *supra* note 985.
[1216] UNESCO, *supra* note 1004 at 9 para. 39.
[1217] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 9.
[1218] *Id.*
[1219] ZUIDERVEEN BORGESIUS, *supra* note 19 at 54.
[1220] FJELD ET AL., *supra* note 990 at 45–46; OECD, *supra* note 991 at Principle 1.3.ii; Amnesty International and Access Now, *supra* note 1006 at Principle 32a; Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 10.
[1221] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 34.

more when artificial intelligence is used in judiciary-related matters and criminal justice.[1222] The right to a fair trial necessitates a transparent court procedure, thus, if artificial intelligence is used in judicial decision-making, transparency is required in the artificial intelligence based assessment.[1223] The absence of transparency in artificial intelligence based decision-making leads to illegitimate and autocratic decisions. If such artificial intelligence based tools are used in criminal justice and judicial systems, the right to equality and non-discrimination may be infringed, among other human rights.[1224]

Transparency, explainability, and accountability are extremely important when using artificial intelligence in judicial systems.[1225] For instance, the Asilomar AI Principles distinguish between two aspects of transparency: failure transparency, which demands explanations in case of any harm caused, and judicial transparency, which calls for human-made audits when AI-systems are utilized in judicial systems.[1226] Yet, some authors criticize non-binding documents for addressing judicial transparency for being an "AI-ethics overreach" on the ground that judicial-related measures should be addressed in law with legal procedural rules.[1227] Transparency allows such decisions to be open for judicial reexamination.[1228] One should note that the principle of transparency in artificial intelligence has become an established case law in the Netherlands.[1229] The SyRI software used by the Dutch state was subjected to judicial evaluation by the Court of the Hague, which established its use was contrary to Article 8 of the ECHR and decided to terminate its use because it lacked transparency for welfare distribution.[1230] Transparency in artificial intelligence is closely affiliated to accountability and safety principles.[1231] As reflected in the CEPEJ Charter, transparency is paramount for protecting the right to equality and non-discrimination in AI-systems because the decisions made by artificial

[1222] Završnik, *supra* note 598 at 568.
[1223] Reiling, *supra* note 567 at 7.
[1224] Završnik, *supra* note 598 at 568.
[1225] Boddington, *supra* note 1109 at 134.
[1226] Future of Life Institute, *supra* note 991 at Principle 7 for "failure transparency" and Principle 8 for "judicial transparency"; Boddington, *supra* note 1109 at 134.
[1227] Boddington, *supra* note 1109 at 134.
[1228] Reiling, *supra* note 567 at 6.
[1229] *Id.*
[1230] Amos Toh, *Dutch Ruling a Victory for Rights of the Poor*, HUMAN RIGHTS WATCH (Feb. 6, 2020), https://www.hrw.org/news/2020/02/06/dutch-ruling-victory-rights-poor; ECLI:NL:RBDHA:2020:1878, Rechtbank Den Haag, C-09-550982-HA ZA 18-388 (English), *supra* note 804 at paras. 6.86-6.91.
[1231] FJELD ET AL., *supra* note 990 at 42.

intelligence can directly impact individuals' lives, making it vital to reveal the grounds used for decision-making transparently.[1232]

Transparency should involve open data processing methods and external audits.[1233] The CoE's suggestion that states should establish procedures for "independent and effective" audits of AI-systems has the potential to become a specific procedural obligation for artificial intelligence concerning states, when human rights claims involving artificial intelligence are brought before human rights bodies.[1234] The choices, data, and assumptions that decisions are made on should be public through these audits.[1235] Audits on AI-systems are also specifically necessary where the reason why an AI-system made a decision cannot be precisely known.[1236] Regular reporting on the AI-system can also be used to achieve transparency, as it demands AI-system operators to systematically share information about the use of AI, including how the AI-system produces outcomes and the actions taken by the artificial intelligence operator.[1237] As mentioned in the OECD AI Recommendation, reporting can be conducted by states through the facilitating international metrics for the principles.[1238]

"I'm a Robot": The Right to Information in Transparency

Transparency is also related to and required for the right to information regarding the use of artificial intelligence.[1239] The right to information is mentioned as a substantive right for individuals and as an obligation for states to require developers and deployers of artificial intelligence to provide information on the use of artificial intelligence in CAHAI's Feasibility Study.[1240] The scope of the right to information may vary depending on the potential outcomes of artificial intelligence use; for instance, according to the UNESCO AI Recommendation, when a decision made through artificial intelligence impacts human rights, individuals should be thoroughly informed by the artificial intelligence actor or the public sector.[1241] The reasons for a decision made by or with the

[1232] *Id.*; European Commission for the Efficiency of Justice (CEPEJ), *supra* note 590 at Principle 4.
[1233] Reiling, *supra* note 567 at 6.
[1234] AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517 at 33.
[1235] Reiling, *supra* note 567 at 6.
[1236] AD HOC COMMİTTEE ON ARTİFİCİAL INTELLIGENCE (CAHAI), *supra* note 517 at 34.
[1237] FJELD ET AL., *supra* note 990 at 41, 46; ACCESS NOW AND ANDERSEN, *supra* note 8 at 33.
[1238] FJELD ET AL., *supra* note 990 at 46; OECD, *supra* note 991 at Principle 2.5.d.
[1239] FJELD ET AL., *supra* note 990 at 41, 44–45; *See also* Amnesty International and Access Now, *supra* note 1006 at Principle 32a; EUROPEAN COMMISSION, *supra* note 519 at 20; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38.
[1240] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 34.
[1241] UNESCO, *supra* note 1004 at 9 para. 38.

help of an AI-system should be accessible.[1242] Individuals should be able to resort to human review regarding the decision.[1243] Information concerning the AI-system or product should be provided swiftly and effectively.[1244] Likewise, the EU AI Act Proposal Article 13(2) recognizes a right to information that is *"concise, complete, correct and clear"*; that is "*relevant, accessible and comprehensible to users,*" which should specify issues named hereunder, bringing noteworthy explainability requirements for transparency. The EU AI Act Proposal Article 13(3) details the right to information within the context of transparency in high-risk systems.[1245] Additionally, Article 52 of the EU AI Act Proposal is titled "transparency obligations for certain AI-systems." According to this provision:

> "Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use."

The obligation thereto does not apply to the cases where AI-systems are being used to *"to detect, prevent, investigate and prosecute criminal offences"* with the authorization of law unless such systems are open to the public for reporting a crime. Article 52(2) of the EU AI Act Proposal underlines that when using emotion recognition or biometric categorization systems, individuals exposed to such systems should be informed. That, too, excludes lawful use in criminal investigations. While understandable, excluding such obligations from criminal law processes as long as the laws recognize them can be open

---

[1242] *Id.* para. 38.
[1243] *Id.* para. 38.
[1244] *Id.* para. 38.
[1245] European Commission, *supra* note 12 at Article 13 reads as *"(a) the identity and the contact details of the provider and, where applicable, of its authorised representative;" "(b) the characteristics, capabilities and limitations of performance of the high-risk AI system, including:" (i) its intended purpose; (ii) the level of accuracy, robustness and cybersecurity referred to in Article 15 against which the high-risk AI system has been tested and validated and which can be expected, and any known and foreseeable circumstances that may have an impact on that expected level of accuracy, robustness and cybersecurity; (iii) any known or foreseeable circumstance, related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, which may lead to risks to the health and safety or fundamental rights; (iv) its performance as regards the persons or groups of persons on which the system is intended to be used; (v) when appropriate, specifications for the input data, or any other relevant information in terms of the training, validation and testing data sets used, taking into account the intended purpose of the AI system. (c) the changes to the high-risk AI system and its performance which have been pre-determined by the provider at the moment of the initial conformity assessment, if any; (d) the human oversight measures referred to in Article 14, including the technical measures put in place to facilitate the interpretation of the outputs of AI systems by the users; (e) the expected lifetime of the high-risk AI system and any necessary maintenance and care measures to ensure the proper functioning of that AI system, including as regards software updates.".*

to discriminatory risks because it is a broad criterion, as these may as well be interfering with -and potentially violating human rights.

Deep fakes[1246] are particularly relevant to transparency because it relates to political and public discussions as well as the freedom of expression; for instance, if deep fakes micro-target voters with inauthentic content and accounts, public debate and democracy could be hindered, leading to polarization of the society and thereby impacting equality and due to effecting people's opinions; causing discrimination.[1247] Deep fakes are also mentioned as a specific concern with regard to transparency in Article 52(3) of the EU AI Act Proposal because they are generated or manipulated images, videos, or audio that resemble real objects, persons, or other things that may appear true or real, accordingly, its use must be informed that it is generated or manipulated: *"that the content has been artificially generated or manipulated."*[1248] Deep fakes based on generative intelligence carry even more gravity regarding the disinformation potential.[1249] In parallel, the risks of expansion in generative intelligence have caught the eye of the EU. The proposed amendments by the MEPs to the EU AI Act Proposal which extend transparency measures for generative artificial intelligence tools to ensure the protection of human rights, democracy, and the rule of law, health, and safety and the environment, aiming to avert producing illegal content.[1250] The CoE Draft Convention on AI mentions the right to information with regard to redress mechanisms, transparency in the context of contestability, connecting it to accountability. Subparagraph (b) of Article 19 thereto states that *"guaranteeing that such communication contains sufficient information for an*

[1246] Deep fakes are photos, images, or videos produced by deep learning methods that appear so authentic that it may not be possible to differ from reality despite being "fake" and inaccurate. University of Virginia, *What the Heck Is a Deepfake? | Information Security at UVA, U.Va.*, INFORMATION SECURITY AT UVA, https://security.virginia.edu/deepfakes (last visited Jun 14, 2023).

[1247] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 9,11; *See also* Reuters, *Deepfakes Are Biggest AI Concern, Says Microsoft President*, THE GUARDIAN, May 25, 2023, https://www.theguardian.com/technology/2023/may/25/deepfakes-ai-concern-microsoft-brad-smith (last visited Jun 9, 2023).

[1248] Article 52(3) of the EU AI Act Proposal excludes the use of deepfakes if:
"authorized by law to detect, prevent, investigate and prosecute criminal offenses or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties."

[1249] Alex Pasternack, *GPT-Powered Deepfakes Are a 'Powder Keg,'* FAST COMPANY (2023), https://www.fastcompany.com/90853542/deepfakes-getting-smarter-thanks-to-gpt (last visited Jun 9, 2023).

[1250] AI Act, *supra* note 1131.

effective possibility of contesting the application of the system or challenging the decision(s) affecting the artificial intelligence subject's rights and freedoms insofar as the use of such system is concerned." Being aware of being faced with an AI-system could impose a procedural obligation as it would help in a demonstrating a discrimination through artificial intelligence claim.

Transparency for Accountability

The transparency principle also closely relates to accountability.[1251] Requirements for transparency are to be realized by the humans to be held accountable, allowing them to justify the ways of utilizing AI.[1252] Transparency also calls for auditability as a legal requirement for oversight and examination to ensure accountability for humans; therefore, the construction of the AI-systems should be fitted for independent audits and also public information to discover such systems' processes, effects on human rights and the measures taken for the associated risks and impacts.[1253] The developers, operators, and owners should document that their AI-system complies with human rights at all stages of the AI-lifecycle, including design, development operation, and implementation.[1254] This, of course, calls for designing the AI-system to be auditable by making available verification, validation logging, and recording of the systems suited for legal obligations to be realized, reminding the importance of multidisciplinary nature when working on AI.[1255] The Institute of Electrical and Electronics Engineers's (IEEE) effort in tying together the ethics of artificial intelligence with the artificial intelligence design in this regard is significant in providing elaborative suggestions and is one of the leading non-binding documents in addressing artificial intelligence related situations.[1256] Human rights due diligence is a related obligation that helps to fulfill the implementation of transparency and accountability measures to support and realize artificial intelligence principles throughout the all stages of AI-lifecycle.[1257] In parallel, academia, governmental institutions, and international organizations particularly emphasize the

[1251] YEUNG, *supra* note 518 at 46; UNESCO, *supra* note 1004 at 9 para. 41.
[1252] YEUNG, *supra* note 518 at 46.
[1253] Yeung, Howes, and Pogrebna, *supra* note 529 at 93; Bryson and Theodorou, *supra* note 1163 at 12; Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 10; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 33.
[1254] Yeung, Howes, and Pogrebna, *supra* note 529 at 93.
[1255] *Id.*
[1256] IEEE, *supra* note 991 at 27, Principle 5-Transparency. also see the "Law" Chapter 211-281.
[1257] Yeung, Howes, and Pogrebna, *supra* note 529 at 93.

importance of human oversight mechanisms in measuring transparency and accountability.[1258]

Transparency makes known the reasons for an action or decision through artificial intelligence and questions quality of the reasons and whether they are enough for justification.[1259] Justification hereto may remind the justification in the discrimination test. The application of these principles should also align with the requirements of the rule of law.[1260] Accountability mechanisms are related to governmental power use; thus, accountability and responsibility for human rights violations are essential.[1261] The CoE's "Guidelines on Addressing the Human Rights Impacts of Algorithmic Systems" enunciate transparency, accountability, and effective remedies under the same title.[1262] According to the said Guidelines, private actors are obliged to lay out terms of service regarding the AI-systems; ensure an accessible human reviewer for contestability; publicize information for transparency; ensure effective remedy mechanisms; engage in consultation with different stakeholders.[1263] On the other hand, public actors are obliged to establish appropriate transparency levels, ensure identifiability and traceability in the public and private sphere for artificial intelligence uses that may significantly impact human rights; afford contestability; ensure adequate oversight; make available effective remedies; seek to reduce barriers that could limit the effective remedies.[1264] Accountability is by itself a fundamental principle requiring its own examination.

3.2.2 Accountability and Responsibility

Accountability is another principle frequently recognized in non-binding and binding documents, extending to impact assessments, monitoring, evaluation, auditing methods, and responsibility.[1265] Accountability is required and possible throughout the AI-

[1258] Ienca and Vayena, *supra* note 972 at 51.
[1259] YEUNG, *supra* note 518 at 46–47; Karen Yeung & Adrian Weller, *How Is 'Transparency' Understood by Legal Scholars and the Machine Learning Community?*, *in* BEING PROFILED: COGITAS ERGO SUM. 10 YEARS OF 'PROFILING THE EUROPEAN CITIZEN 36, 37 (2018), https://mediarep.org/handle/doc/14193 (last visited May 22, 2023); See Monika Zalnieriute, Lyria Bennett Moses & George Williams, *The Rule of Law and Automation of Government Decision-Making*, 82 MOD. LAW REV. 425, 19 (2019), https://onlinelibrary.wiley.com/doi/10.1111/1468-2230.12412 (last visited May 22, 2023).
[1260] Zalnieriute, Moses, and Williams, *supra* note 1259 at 31.
[1261] YEUNG, *supra* note 518 at 47.
[1262] Council of Europe Committee of Ministers, *supra* note 991.
[1263] *Id.* at Principle C4 13-14.
[1264] *Id.* at Principle B4 9.
[1265] FJELD ET AL., *supra* note 990 at 28.

lifecycle, at the design stage before using an AI-system, at monitoring during the use of artificial intelligence, and after, if any harm occurs, by redress.[1266] Mechanisms are required for accountability that aim at bringing compliance to AI-systems and applications.[1267] Accountability helps to hold a person responsible, supposing that discrimination occurs.[1268] It can be realized differently and only sometimes refers to legal accountability. Likewise, legal accountability comprises different types, such as private accountability, liability, and human rights law accountability.[1269] A part of accountability discussions is around the accountability of the private actors in private law, which is beyond the scope of this study. Rather human rights accountability for equality and non-discrimination is the focus of this study.

To be held accountable, one must have a role and responsibility at varying levels toward the outcome.[1270] It is not always clear who has control and is responsible enough to be held accountable for artificial intelligence.[1271] Designers and deployers of an AI-system may be different, and neither the ones who design it may not predict or understand the system's potential effects, mainly if techniques of deep learning are used, nor the deployers may not know the potential effects or fully understand the system's working mechanism.[1272] The "black-box effect" challenges reaching accountability as it makes responsibility difficult.[1273] Accountability by design is advocated by authors because it helps to bring control without the risks, such as sharing of the data required for transparency.[1274] Accountability requires verifiability and replicability to ensure that AI-systems do not exhibit distinct conduct but consistently render similar outcomes in similar situations, thereby improving the reliability of these systems.[1275] However, there are also arguments against the possibility of reproducibility in all cases.[1276]

[1266] *Id.* at 29.
[1267] Gasser and Schmitt, *supra* note 991 at 153; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38.
[1268] WAGNER, *supra* note 19 at 39.
[1269] *Id.*
[1270] *Id.*
[1271] *Id.*; See IEEE, *supra* note 991 at 237–239.
[1272] WAGNER, *supra* note 19 at 39.
[1273] IEEE, *supra* note 991 at 237.
[1274] Joshua Kroll, *Accountable Algorithms (A Provocation)*, MEDIA@LSE (Feb. 10, 2016), https://blogs.lse.ac.uk/medialse/2016/02/10/accountable-algorithms-a-provocation/ (last visited May 24, 2023).
[1275] FJELD ET AL., *supra* note 990 at 29; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1276] Kroll, *supra* note 1274.

In order to understand the accountability principle, close inspection of the forthcoming legally binding documents is valuable. The CoE Draft Convention on AI constructs accountability, responsibility and legal liability together under Article 14 – "Principle of Accountability, Responsibility and Legal Liability":

> "Each Party shall, within its jurisdiction, take measures necessary to ensure accountability, responsibility and legal liability for any unlawful harm or damage to human rights and fundamental freedoms resulting from the application of artificial intelligence systems."

The article approaches accountability with respect to human rights by obliging them to ensure its realization. However, what unlawfulness represents in the provision needs to be clarified. In a legally binding document specific to artificial intelligence, at least describing the circumstances that constitute *"unlawful harm or damage to human rights and freedoms"* would be more appropriate.

Looking for the Accountable

An important question for accountability is whom to hold accountable for an AI-system's decision.[1277] International human rights law would answer this by holding the state as the ultimate accountable and the guarantor of accountability by taking measures and establishing accountability mechanisms; further business and human rights mechanisms bring a solution by incorporating the private actors.[1278] The recommendations presented by the CoE and the Draft Framework Convention on AI support this interpretation as explained below. In the first chapter, fairness discussions in artificial intelligence were explained, and connecting fairness in law and ethics was aimed. One element of fairness is accountability because to achieve fairness, responsibility, and being held accountable for the circumstances, mechanisms should be put in place for redress, and the strongest ones are the legal mechanisms.[1279] Therefore, redress mechanisms are a must for addressing accountability; without a remedy, accountability would lack its meaning.[1280] Regarding accountability, the CAHAI presented the right to an effective remedy from the ECHR within the context of developing or using AI, applicable when that leads to unjust harm or violations either by

[1277] WAGNER, *supra* note 19 at 40.
[1278] McGregor, Murray, and Ng, *supra* note 1 at 312–313.
[1279] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 19.
[1280] *Id.* at 20; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38.

private or public authorities and obliged them to ensure accessible redress mechanisms and other relevant measures for reminding the burden of proof reversal in discrimination.[1281] Effective and accessible remedies before a national authority for artificial intelligence related human rights violations are needed. [1282] As mentioned in the second chapter, showing the difference in treatment is sometimes enough when evaluating discrimination. Therefore, for accountability mechanisms to function correctly, caution for not falling into automation bias should be given, presenting evidence should not be challenging, and reversal of the burden of proof should be possible.[1283]

Human oversight and control principle is also related to accountability.[1284] Human control over artificial intelligence is paramount when artificial intelligence makes independent decisions without any human participation.[1285] That is important, as underlined by the CoE Commissioner for Human Rights, to hold a natural or legal person responsible when an adverse impact on human rights occurs, even if there is no human involvement in the decision itself.[1286] The responsible authorities should be established for human rights related matters to ensure accountability.[1287] The accountability principle is again related to democracy because it would allow different stakeholders to participate in the process of public debates on artificial intelligence.[1288]

The CoE's Resolution 2341 (2020) of the Parliamentary Assembly on the Need for Democratic Governance of Artificial Intelligence underlines that a legal framework is required to be put in place for providing independent and proactive oversight and algorithmic vigilance mechanisms where all stakeholders are present to ensure effective compliance throughout the lifecycle.[1289] There is also a mention of a high body capable

[1281] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 39.
[1282] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 13; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38.
[1283] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 14; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38.
[1284] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38; MICROSOFT CORPORATION, *Microsoft Responsible AI Standard v2 General Requirements*, 27 See Goal A5 (2022), https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl 8; AI at Google: our principles, GOOGLE Principle 4 (2018), https://blog.google/technology/ai/ai-principles/ (last visited May 19, 2023).
[1285] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 13.
[1286] *Id.*; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38.
[1287] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 13.
[1288] Smuha et al., *supra* note 106 at 8.
[1289] Council of Europe Parliamentary Assembly, *supra* note 657 at para 15; Mantelero, *supra* note 557 at 73; Council of Europe Directorate General of Human Rights and Rule of Law, *supra* note 142 at paras II.10, III.2-III.3.

of technically, legally, and ethically competent. [1290] Given that there are no mechanisms to hold the agents accountable at all stages of the artificial intelligence lifecycle, assigning accountability renders near impossible.[1291] If there are no mechanisms for involvement of responsibility and accountability, then such mechanisms cannot be trusted.[1292] Using such systems due to the dangers they hold, especially concerning equality and discrimination, could even be unlawful as these systems are used for decision-making.

This study supports the suggestion that considering the role of the private sector and their hesitation towards legal regulations, applying the existing human rights law framework, which is always applicable and should be the primary approach, another way to approach private actors is through business and human rights mechanisms. The UN Business and Human Rights principles can be applied to address human rights problems related to artificial intelligence.[1293] That could greatly contribute to increasing substantive equality, particularly where the application of legal norms may not be sufficient to uncover unseen inequalities. Private actors may feel compelled and motivated through such regulations to take on the primary responsibility.[1294] This could also be important for facilitating immediate and convenient action. All design stage processes, audits, and assessments can be a part of this application, but a connection needs to be established. Human rights due diligence tools may be helpful, especially for accountability, before adverse impacts such as discrimination occur. Further research is required in this regard.

Accountability Measures Before and During the Design and Deployment

Human oversight, impact assessments, risk assessments, audits, and due diligence mechanisms are essential in realizing the accountability principle.[1295] The non-binding regulatory documents on artificial intelligence involve different methods of supervision. These can be achieved through either technical methods, especially throughout the design and deployment, or with other methods, such as impact assessment and audits during the deployment of the AI-systems.[1296] When discussing transparency, accountability was mentioned already.

[1290] Council of Europe Parliamentary Assembly, *supra* note 657 at para 15.
[1291] IEEE, *supra* note 991 at 236.
[1292] *Id.*
[1293] McGregor, Murray, and Ng, *supra* note 1 at 313.
[1294] *Id.* at 312–313.
[1295] UNESCO, *supra* note 1004 at 9–10 para. 43.
[1296] *Id.* at Principle 43.

Traceability is also an essential aspect of accountability.[1297] Auditability is an integral part of accountability as it makes available assessing the data, algorithms, and design processes.[1298] Similar reflections on accountability is found in some of the leading private sector documents. For example, Microsoft's Responsible AI Standard introduces accountability as the first principle defined under five goals with different requirements: impact assessment, oversight mechanisms, fitness for purpose, data governance and management, and human oversight and control.[1299]

Accountability of AI-systems can be improved through internal and external audits; for example, equality and non-discrimination-specific audits to be conducted by independent auditors may be considered.[1300] Algorithmic impact assessments are essential again for accountability, at the steps of development and deployment and during the use of artificial intelligence.[1301] Reporting negative impacts is another requirement for accountability, especially to identify, evaluate, document, and analyze, thus minimizing the adverse outcomes of AI-systems.[1302] Reporting such matters is significant for the right to equality and non-discrimination because, occasionally, some discriminatory cases may not "qualify" as legal cases as such, or some may be indirectly affected, or equality, in general, is damaged.[1303] Reporting should be an obligation for states, in addition to the private actors.[1304] The accountability mechanisms are in line with the business and human rights realm.[1305] The CoE also supports implementing business and human rights in the private sector, referring to UN Guiding Principles and the CM/Rec(2016)3 of the Committee of Ministers to member states on human rights and business.[1306] Impact

[1297] *Id.*
[1298] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 19.
[1299] MICROSOFT CORPORATION, *supra* note 1284 at Goal A1 to Goal A5.
[1300] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 20.
[1301] *Id.*
[1302] *Id.*; AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 38.
[1303] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 20.
[1304] Mantelero, *supra* note 557 at 73.
[1305] See HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 20 citing the EU business and human rights under the remedy discussions; *See also* IMPROVING ACCESS TO REMEDY IN THE AREA OF BUSINESS AND HUMAN RIGHTS AT THE EU LEVEL - OPINION OF THE EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, 80 (2017), https://fra.europa.eu/sites/default/files/fra_uploads/fra-2017-opinion-01-2017-business-human-rights_en.pdf.
[1306] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 9; UNITED NATIONS GENERAL ASSEMBLY, *supra* note 978; Council of Europe, *Recommendation CM/Rec(2016)3 of the Committee of Ministers to Member States on Human Rights and Business*, Recommendation CM/Rec(2016)3 (2016), https://rm.coe.int/human-rights-and-business-recommendation-cm-rec-2016-3-of-the-committe/16806f2032 (last visited Aug 4, 2022).

assessments and reports are mentioned in private-sector documents.[1307] Additional human rights protection mechanisms are advised for artificial intelligence by the CoE Commissioner for Human Rights, such as through human rights due diligence, implying that conducting human rights due diligence has real potential for becoming a positive obligation for states with regard to artificial intelligence.[1308]

The CoE Recommendation CM/Rec(2020)1 refers to openly conducted human rights impact assessments.[1309] To achieve more substantively equal effects from the assessments, impacted groups and individuals should be a part of such assessments.[1310] Recommendation CM/Rec(2020)1 suggests that if high-risk systems are to be used, monitoring and review techniques and processes should be publicly available, and where secrecy is required, competent authorities should still be able to reach them.[1311] This would help not to neglect equality in the name of secrecy, while respecting the artificial intelligence actors.

The CoE Draft Convention on AI, after introducing accountability as a principle, dedicates a chapter on the *"measures and safeguards ensuring accountability and redress"* between Articles 19 to 23. Article 19 of the CoE Draft Convention on AI juxtaposes the measures for *"ensuring the availability of redress"* and mandates that the states provide available redressing mechanisms for harm to human rights resulting from the application of AI-systems. Subparagraph (a) hereunder obliges the states to set up a system to "*the relevant usage of the system is recorded and, where appropriate, communicated to the artificial intelligence subjects concerned."* This system is understood to be left to the states as long as they render possible redress specific to AI-systems. Subparagraph (b), as mentioned under transparency discussions, connects accountability with transparency requirements involving a right to information enabling contestability. Subparagraph (c) follows *"making available effective redress mechanisms"* directly bringing a procedural obligation for AI-systems. It is understood that the system where artificial intelligence use will be recorded should be specific for artificial intelligence use, but the redress mechanisms can be existing human rights mechanism as long as it is available for artificial intelligence. Article 21 of the CoE Draft

[1307] MICROSOFT CORPORATION, *supra* note 1284 at Goals A1-A5 4-8.
[1308] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 9.
[1309] Council of Europe Committee of Ministers, *supra* note 991 at para 5.3.
[1310] *Id.* at para 5.3.
[1311] *Id.* at 5.3.

Convention on AI provides restrictions to the measures provided in Articles 19 and 20, constructed in a similar way to a second paragraph of an article under the ECHR. Nevertheless, it needs to be clarified why the restriction regime is limited by Articles 19 and 20.

After examining the CoE Draft Convention, how accountability is dealt under EU documents is also critical. Data governance and management practices recognized under Article 10 of the EU AI Act Proposal involve design choices, collection of data, *"relevant data preparation processing operations"*, *"formulation of relevant assumptions"* such as what the data measures and what it represents; assessing the datasets with respect to availability, quantity, and suitability; bias examination and identifying data gaps and how can they be targeted. For example, the biases mentioned in the first chapter must be tested with regard to their discriminatory potential before using AI-systems in this article.

Article 64 of the EU AI Act Proposal concerns access to data, and subparagraph three thereof brings documentation requirements to national public authorities or bodies *"(…) which supervise or enforce the respect of obligations under Union law protecting fundamental rights"* concerning the use of standalone high-risk AI-systems under Annex III, bringing more protection to human rights contributing to transparency, accountability, and oversight. In case the documentation is insufficient, Article 64(5) mentions that national public authorities can request from the market surveillance authority to test with the involvement of the public authority the respective high-risk AI-system through technical means to establish a breach of obligations under EU law for fundamental rights.

Article 65 deals with the *"procedure for dealing with AI systems presenting a risk at national level"* and mentions that if the market surveillance authority determines an AI-system carries risks to fundamental rights, the market surveillance authority should inform the relevant national public authorities or bodies. The article further elaborates on subparagraph two that:

> "(…) the market surveillance authority finds that the AI system does not comply with the requirements and obligations laid down in this Regulation, it shall without delay require the relevant operator to take all appropriate corrective actions to bring the AI system into compliance, to withdraw the AI system from the market, or to recall it within a reasonable period, commensurate with the nature of the risk, as it may prescribe."

Such action is essential for addressing the potential violations of the right to equality and non-discrimination in taking corrective actions and compliance with that.

Article 66 under Title VIII of the EU AI Act Proposal, deal with safeguards procedures. Article 68 of the EU AI Act Proposal further details the measures to be taken by the market surveillance authority in case of non-compliance from the provider. Article 69 of the EU AI Act Proposal brings an obligation to member states to *"encourage and facilitate"* the preparations of codes of conduct with the participation of different stakeholders, minding diversity for voluntary application besides the requirements set for high-risk AI-systems *"on the basis of technical specifications and solutions that are appropriate means of ensuring compliance with such requirements in light of the intended purpose of the systems."*

Accountability Measures After the Design and Deployment

Accountability is crucial concerning due process rights and the rule of law discussions.[1312] Effective redress is required whenever someone is discriminated against through artificial intelligence.[1313] In that direction, it is possible to claim Article 14 of the ECHR in conjunction with the right to a fair trial.

Chapter VI of the CoE Draft Convention on AI is titled *"Assessment and Mitigation of Risks and Adverse Impacts."* The chapter reflects influences of a post-business and human rights world, but as it is a convention that is going to be ratified by the states, the obligations are imposed on states. Article 24 thereunder introduces a requirement for risk and impact management, which the states are obliged to *"provide effective guidance."* The scope of this guidance encompasses identifying, assessing, preventing, and mitigating risks and adverse impacts *"resulting from the application of an artificial intelligence system in relation to the enjoyment of human rights, the functioning of democracy and the observance of rule of law."* In this provision, there is no explicit obligation to conduct a risk and impact assessment by the state parties. The second subparagraph requires that and gives directions relating to risk and impact assessments, requiring the risk and impact assessments to be continually and promptly conducted throughout the lifecycle of an AI-system to involve human rights related requirements. Article 24 further demands guidance on the *"recording, and due consideration of adverse*

[1312] WAGNER, *supra* note 19 at 39.
[1313] *Id.*

impacts resulting from application of enhanced artificial intelligence systems in relation to the enjoyment of human rights, the functioning of democracy and the observance of rule of law," to document the impacts of AI-systems. However, addressing the *due consideration* is at least interesting because there is no clarity regarding whether it is enough for the adverse impacts to be duly considered solely and what will happen when adverse effects on human rights are identified.

Subparagraph 3 of Article 24 allows for a moratorium or ban of an AI-system where appropriate and necessary. Instead of limiting an AI-system with a ban, it could be better to include specific methods of artificial intelligence to be used in a specific area to the scope of a potential ban, which could help to bring a contextual approach.

Article 25 concerns the obligations of artificial intelligence providers and users, including private and public authorities. The obligation imposed on the states is to ensure that artificial intelligence users apply sufficient preventive and mitigating measures resulting from the risk and impact management framework, which means that the aim is to prevent them beforehand. Nevertheless, mitigating means reducing and lessening the effects, so the purpose is not to eliminate unlawful risks and impacts. Instead, the goal is "mitigating" them. There is a *realistic* approach and the acceptance that there will always be risks and impacts that cannot be eliminated when using AI. Article 26 of the CoE Draft Convention on AI envisions a training regarding the risk and impact management framework, presenting an important supportive measure.

Follow-up is also necessary to accomplish accountability, particularly for private actors, which involves a process of implementing review mechanisms such as algorithmic impact assessments and taking action to address risks and adverse impacts with fast and instantaneous control on the AI-lifecycle and rectifying any adverse effects, primarily to address feedback loops.[1313][1314]

Chapter VII of the CoE Draft Convention on AI is titled "Follow-up mechanism and cooperation". Article 27 thereunder is about consultation of the parties. The CoE Draft Convention on AI sets out a periodic consultation for parties to make a proposal about the use, implementation, and amendment of the convention, formulate and express opinions on proposals, and the interpretation and application of the convention. The consultation is to be convened by the secretary general of CoE whenever s/he finds it

[1314] Council of Europe Committee of Ministers, *supra* note 991 at para 5.4.

necessary and when the majority of the parties or the Committee of Ministers request the meeting. Article 28 of the CoE Draft Convention on AI addresses international cooperation, which may seem optimistic considering the competition for becoming leaders in artificial intelligence. However, international cooperation is necessary for addressing human rights issues effectively in artificial intelligence, especially considering the blurred borders regarding it. Dealing with human rights related issues necessitates the cooperation of states in the context of artificial intelligence. Article 29 of the CoE Draft Convention on AI is about national supervisory authorities aimed at *"overseeing and supervising compliance with the requirements of the risk and impact assessment of artificial intelligence systems."* The convention brings a procedural obligation on the states to establish or designate independent and impartial supervisory authorities at the national level.

On the other hand, the EU AI Act Proposal has dedicated Title VIII to *"post-market monitoring, information sharing, market surveillance."* Article 61(1) and 61(2) under Title VIII deals with post-market monitoring, which is encumbered on the artificial intelligence providers proportionate to the nature of the artificial intelligence technologies and the risk of AI-systems actively and systematically to collect the document and analyze relevant performance data throughout the lifetime of the AI-systems, allowing for compliance with the requirements such as transparency, human oversight listed under the Chapter 2 of the Title III. According to Article 62 under Title VIII, severe incidents and malfunctionings that violate the obligations under the EU law intended to protect human rights are due for reporting obligations to the market surveillance authorities. As the article mentions, for the protection of human rights and especially for the right to equality and non-discrimination, such a provision is essential for accountability and addressing these issues. Under Title VIII Chapter 3, an enforcement mechanism is implemented—Article 63 concerns market surveillance and controlling AI-systems in the EU market.

There is a penalty regime appointed in Articles 71 and 72 of the EU AI Act Proposal, which sets out administrative fines in different cases where the regulation is not followed, in the latter for union institutions and bodies if they fail to comply with the regulation.[1315]

[1315] European Commission, *supra* note 12 at Article 71-72.

It should be noted that the EU AI Act Proposal's provisions on accountability are considered as inadequate in ensuring accountability.[1316]

3.2.3 Human Control and Oversight

Human control is another main principle concerning artificial intelligence.[1317] Human review of automated decisions, the ability to opt-out of automated decisions, and general human control over technology are such outlooks of this principle.[1318] As mentioned under other principles, human control and oversight are related to transparency and explainability, safety and security, and accountability.[1319] Human control and oversight is also essential to ensuring fairness and protecting the right to equality and non-discrimination.[1320] Human oversight principle tells that the design of AI-systems should not result in a decision based only on the automated processing of data, without human opinions.[1321] Human control over the AI-systems is required.[1322] Humans should be able to not resort and deviate from an AI-system's decisions or recommendations.[1323] Humans should be able to make autonomous decisions when dealing with and for AI-systems.[1324] Humans should be able to understand and have the information on interacting with AI-systems, make self-evaluations, and come against the AI-system, where reasonable.[1325] Human autonomy should be essential, and forms of manipulation through artificial intelligence to influence humans should be eliminated.[1326] Thus, human autonomy should

[1316] *See* Smuha et al., *supra* note 106 at 48.
[1317] FJELD ET AL., *supra* note 990 at 53.
[1318] FJELD ET AL., *supra* note 990 at 53; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *Ethics Guidelines for Trustworthy AI*, 39 16 (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419; HLEG Ethics Guidelines refer to GDPR in this regard that mentions a right to *"...not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), OJ L 119 4.5.2016 Article 22 (2016), https://eur-lex.europa.eu/eli/reg/2016/679/oj (last visited Aug 10, 2022).
[1319] FJELD ET AL., *supra* note 990 at 53.
[1320] *Id.*
[1321] Mantelero, *supra* note 557 at 73.
[1322] *Id.*; UNESCO, *supra* note 1004 at 8–9 paras. 35-36; Council of Europe Directorate General of Human Rights and Rule of Law, *supra* note 142 at para II.8 important especially for its relation to personal data.
[1323] Mantelero, *supra* note 557 at 73; Council of Europe Directorate General of Human Rights and Rule of Law, *supra* note 142 at para III.4.
[1324] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16.
[1325] *Id.*
[1326] *Id.*; Muller and The CAHAI Secretariat, *supra* note 90 at 34.

be supported by AI-systems, which compel human oversight.[1327] This principle is mentioned with reference to the "common good" and with relation to benefit for all humanity instead of one group or a state, which can also be beyond the limits of narrow artificial intelligence and valid for the future of artificial intelligence.[1328] Incorporation of stakeholders and collaboration is also valuable for equality and non-discrimination because it allows for consultation, and it could help to ensure a more substantively similar, inclusive approach by bringing academics, affected groups, policymakers, and such together.[1329] It could be a significant contribution to addressing feedback loops and biases mentioned in the first chapter.[1330] One should note that discussions on human control on artificial intelligence also goes together with singularity discussions, which foresee a future without humans being the central actors in the world.[1331]

"User-controlled artificial intelligence" refers to circumstances where users should be able to control what artificial intelligence does, have choices, and act differently than artificial intelligence.[1332] Mechanisms and safeguards should be taken to protect "human values" and "fairness."[1333] As mentioned earlier, the right to ask for a human reviewer for an automated decision relates to human control, which is a remedial measure allowing an objection and second look at the decision and allowing for any automation bias to be removed.[1334] Human control allows for human intervention in decision-making.[1335] For objectives determined by humans, humans should be able to choose whether to conduct a task through artificial intelligence or whether to do it with the help of artificial intelligence.[1336] Accordingly, opting out of automated decisions is another opportunity for ensuring human control, to not to be subject to automated decisions and to leave room for human choice.[1337] EU Ethics Guidelines for Trustworthy AI mention a

[1327] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 15; EUROPEAN COMMISSION, *supra* note 519 at 21.
[1328] Future of Life Institute, *supra* note 991 at Principle 23; FJELD ET AL., *supra* note 990 at 58.
[1329] FJELD ET AL., *supra* note 990 at 58; IBM, *supra* note 216 at 24.
[1330] FJELD ET AL., *supra* note 990 at 59; IBM, *supra* note 216 at 30–31.
[1331] FJELD ET AL., *supra* note 990 at 53; See for different perspectives on legal singularity, and the future of law IS LAW COMPUTABLE?: CRITICAL PERSPECTIVES ON LAW AND ARTIFICIAL INTELLIGENCE, (Simon Deakin & Christopher Markou eds., 1 ed. 2020).
[1332] Reiling, *supra* note 567 at 6.
[1333] OECD, *supra* note 991 at Principle 1.2.b; FJELD ET AL., *supra* note 990 at 53.
[1334] FJELD ET AL., *supra* note 990 at 53.
[1335] *Id.* at 54.
[1336] Future of Life Institute, *supra* note 991 at Principle 16; FJELD ET AL., *supra* note 990 at 54.
[1337] FJELD ET AL., *supra* note 990 at 54; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 12.

"self-determination" for humans in their interactions with AI-systems to protect human autonomy.[1338] This study is of the opinion that where a public authority or a private actor uses an AI-system that the potential impact the right to equality and non-discrimination is widely known, the right to ask for a human reviewer should apply directly, without requiring to ask to use that right. In other cases, it could apply in a request-based way.

Artificial intelligence requires a "human-centric approach" stemming from the idea of human dignity, which is at the foundation of human rights.[1339] Human centric approach to artificial intelligence is also highlighted in EU documents.[1340] One strong form of human oversight principle is found in CEPEJ Charter.[1341] The CEPEJ Charter mentions that artificial intelligence use should not restrict user autonomy, rather, it should be increased.[1342] The CEPEJ Charter also notes judicial review in the context of human review.[1343] Considering the CEPEJ Charter is directly relevant for using artificial intelligence in judicial systems, this option is necessary to protect the ultimate decision maker from being a human. Judicial adoption of artificial intelligence underlines some of the considerations for oversight and human control, essentially the competency of human operators.[1344] In the *Loomis* case, human control was the central theme, where the use of AI-based assessment and the fair trial right was at stake.[1345] The judges have thereunder ruled that using such tools is acceptable if the judges give reasons on how they use the risk assessment tool meaning that human control should not be assigned to AI-systems entirely.[1346] In the UK, there is another example highlighting the importance of human oversight, where the financial capacity of ex-spouses is evaluated in maintenance proceedings; the tool makes the calculations. A mistake in calculation has led to false calculations in more than 3500 cases where some of them already completed.[1347] Judges, if they are using artificial intelligence tools and AI-based decisions, should provide the

[1338] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 12; FJELD ET AL., *supra* note 990 at 53.
[1339] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 10.
[1340] European Commission, *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions - Building Trust in Human-Centric Artificial Intelligence*, 11 (2019), https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence.
[1341] FJELD ET AL., *supra* note 990 at 54.
[1342] European Commission for the Efficiency of Justice (CEPEJ), *supra* note 590 at Principle 5 12 .
[1343] *Id.* 12; FJELD ET AL., *supra* note 990 at 54.
[1344] IEEE, *supra* note 991 at 231–235.
[1345] Reiling, *supra* note 567 at 7.
[1346] *Id.*
[1347] *Id.*

reasons for their use.[1348] Human control is required in all stages of AI-based decisions.[1349] These examples should render into procedural obligations. Furthermore, using artificial intelligence in government should be subject to oversight mechanisms, including court orders and ombudspersons for complaints.[1350] In a similar vein, intervention through public enforcement is also essential for human oversight.[1351] There could be cases where it is required from the governments to stop the use of an AI-system to end the inequalities resulting from its use.[1352] Besides the risk level, the degree of oversight may also depend on the intended use and the potential effects on individuals.[1353]

Professional responsibility also aligns with the discussions on the human control principle.[1354] Professional responsibility obliges the artificial intelligence professionals to provide accuracy in AI-systems when making decisions, categorizations, predictions, or recommendations.[1355] Responsible design is also related to professional responsibility, stemming from the aim of being responsible and aware when designing the AI-systems.[1356] This is important for ensuring equality and protection and preventing discrimination.

Where there is a potential risk for discrimination or other adverse impacts on human rights, human rights assessments should be conducted to ensure oversight beforehand.[1357] EU Ethics Guidelines for Trustworthy AI suggest that should be conducted before the deployment and also include an analysis on whether the *"risks can be reduced or justified as necessary in a democratic society in order to respect the rights and freedoms of others,"* mainly reminding the part of the balancing test[1358] used by human rights mechanisms.[1359] This could be applied to the right to non-discrimination. In that case, the potential discriminatory risks are determined, and in the impact assessment, whether risks can be justified or minimized is shown in the assessment to become ready for future

[1348] *Id.*
[1349] *Id.* at 8.
[1350] Muller and The CAHAI Secretariat, *supra* note 90 at 34.
[1351] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16.
[1352] Muller and The CAHAI Secretariat, *supra* note 90 at 34.
[1353] EUROPEAN COMMISSION, *supra* note 519 at 21.
[1354] FJELD ET AL., *supra* note 990 at 55; IBM, *supra* note 216 at 18.
[1355] FJELD ET AL., *supra* note 990 at 56; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17; AI at Google, *supra* note 1284 at Principle 1; See IEEE, *supra* note 991 at Principle 4 25.
[1356] FJELD ET AL., *supra* note 990 at 57.
[1357] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 15.
[1358] *See* title "Justifying the Discrimination" under Chapter 1.3.2 and "Uncovering Discrimination Through Artificial Intelligence: The Discrimination Test" under Chapter 2.2.1.
[1359] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 15.

discrimination cases. External feedback mechanisms can help contribute to equality, ensuring human control and oversight of human rights.[1360] Microsoft's AI Principles mention oversight with regard to accountability and at design stage measures, conducting evaluations and participation of stakeholders.[1361]

Human oversight is essential to ensure that the AI-system does not undermine human autonomy or result in discriminatory effects.[1362] There are different approaches to oversight, including human-in-command, referring to a decision being on humans when and how to use an AI-system in a situation; human-in-the-loop, referring to the ability to intervene in every step of the lifecycle and decision cycle in the AI-system, and human-on-the-loop, referring to intervention in design and monitoring during the deployment.[1363] Depending on the risk and application area, varying degrees of oversight mechanisms should be chosen.[1364] For the protection of human rights, and particularly the right to equality and non-discrimination, a human-in-the-loop approach should be preferred when deciding on any issue, and a human-in-command approach when using judicial systems or criminal justice or especially in public uses, or even there is no public use, affecting society and equality in general.

The EU considers in AI White Paper that in high-risk artificial intelligence applications, human involvement is required for a human-centric and trustworthy artificial intelligence.[1365] EU AI White Paper refers to GDPR for cases of data processing by an AI-system, mentioning that the rights thereof should not be hindered.[1366] The EU AI White Paper suggests measures for ensuring human oversight, such as a human review and validation required for the output of the AI-system to be effective, implying a human-in-command approach.[1367] Alternatively, another measure is the outputs being immediately effective but open to human intervention later on, as in the human-in-the-loop approach.[1368] The last measure is monitoring the AI-system during the operation and

---

[1360] *Id.*
[1361] MICROSOFT CORPORATION, *supra* note 1284 at Goal A5.
[1362] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16; Muller and The CAHAI Secretariat, *supra* note 90 at 34.
[1363] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16; Muller and The CAHAI Secretariat, *supra* note 90 at 34.
[1364] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16.
[1365] EUROPEAN COMMISSION, *supra* note 519 at 21.
[1366] *Id.*
[1367] *Id.*
[1368] *Id.*

the ability to intervene and deactivate, such as a "stop button" and putting impositions of operational measures on systems, implying a human-in-the-loop approach.[1369] The examples for discrimination can be as such; for example, when deciding on social security benefit assignment or criminal recidivism, humans would stay as the sole decision-makers; but an AI-system may decide on denying or accepting credit loan applications, but contesting that decision is possible.[1370] UNESCO's human oversight principle calls for the constant availability of ethical and legal responsibility and remedy at all steps of AI-lifecycle.[1371] UNESCO's recommendation looks at oversight of an assortment of human-in-command and human-in-the-loop.[1372]

The CoE Human Rights Commissioner suggests a legislative framework is required for establishing independent and effective oversight over the use of artificial intelligence both for public and private parties.[1373] The requirements to protect the right to equality and non-discrimination may ask for different approaches and measures.[1374] Legally structured oversight mechanisms should be built for AI-systems to comply with human rights.[1375] There can be different oversight mechanisms independent from public or private entities, monitoring and investigating human rights compliance and intervening in case of a human rights violation.[1376] There could be reporting obligations for ensuring oversight to such bodies for public and private authorities and applying the recommendations from the oversight body.[1377]

Toronto AI Declaration's relevant principles for human oversight have valuable considerations for the right to equality and non-discrimination in a manner of legal language.[1378] It firstly states that states should take steps for public officials to be aware and sensitive to discriminatory risks and mandates states to take proactive steps such as engaging in consultations and hiring diversely to ensure all of the steps in the AI-lifecycle have people from different backgrounds.[1379] The second requirement for states is to

---

[1369] *Id.*
[1370] *Id.*
[1371] UNESCO, *supra* note 1004 at 8 para. 35.
[1372] *Id.* at para. 36.
[1373] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 10.
[1374] See for more detail and examples AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 1133 at 12.
[1375] Yeung, Howes, and Pogrebna, *supra* note 529 at 91.
[1376] Dunja Mijatović, Council of Europe Commissioner for Human Rights, *supra* note 55 at 10.
[1377] *Id.* at 11.
[1378] Amnesty International and Access Now, *supra* note 1006 at paras. 33-35.
[1379] *Id.* at para. 33.

conduct human rights training and data analysis for officials who have a role in the lifecycle of AI-systems.[1380] Thirdly, independent oversight mechanisms, including judicial ones, are ordered for states.[1381] Finally, Paragraph 33 of the Toronto AI Declaration reminds the connection to due process rights and says that decisions through artificial intelligence should serve that.[1382] As mentioned in Paragraph 34 of the Toronto AI Declaration, AI's leading actors are from the private sector, and when public use of artificial intelligence is done through contracting to private parties, human rights mechanisms and obligations of states should not be forgone and forgotten.[1383] In cases where the public sector obtains AI-systems and applications from the private sector, control and oversight over the system should be maintained, and conduct human rights due diligence from third parties.[1384]

Another provision in the CoE Draft Convention on AI regarding human oversight is Article 15, which was mentioned when discussing transparency, requiring the placement of adequate oversight mechanisms. Furthermore, Article 20 of the CoE Draft Convention on AI provides procedural safeguards for accountability and human oversight. These include a right to human review *"where an artificial intelligence system substantially informs or takes decision(s) affecting human rights and fundamental freedoms";* a right to know interaction with an AI-system and option for human interaction where appropriate; and appropriate measures concerning the rights mentioned under Article 19 and 20 for persons with disabilities. The said provision approaches human oversight as a safeguard for ensuring accountability. The same applies to transparency; knowing when dealing with an AI-system would help to take better actions. The highlight of the safeguards extending to persons with disabilities shows a substantive equality approach in artificial intelligence-related matters. Finally, the CoE Draft Convention on AI Article 1(3) brings a follow-up mechanism that is detailed in Article 27 for effective implementation of the convention, as mentioned under the accountability principle.

In the EU AI Act Proposal, there are numerous provisions related to human oversight. Article 14 of the EU Act Proposal concerns human oversight. Article 14(1) mentions that AI-systems should be able to *"be effectively overseen by natural persons during the*

[1380] *Id.*
[1381] *Id.*
[1382] *Id.*
[1383] *Id.* at para. 34.
[1384] *Id.* at para. 35.

period in which the AI system is in use." Effective oversight during the use seems to be a human-in-the-loop approach. Regarding human oversight, Article 14(2) explicitly mentions human rights risks that may emerge from the use of a high-risk AI-system and states that the oversight should aim to prevent and minimize those.

Article 14(3) of the EU AI Act Proposal supports that, mentioning measures to ensure human oversight that:

> "(a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;
> (b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user."

Article 14(4) explains the obligations of the individuals assigned with oversight.[1385] Article 14(5) notes that for standalone high-risk AI-systems, *"no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons".* Article 16 of the EU AI Act Proposal brings obligations to high-risk AI-systems providers. Article 16 of the EU AI Act Proposal establishes obligations for providers of high-risk AI-systems. The primary requirement is to adhere to the requirements specified in Chapter 2 under Title III, including transparency, human oversight, data governance, and risk management systems. Additional obligations imposed on providers of high-risk AI-systems encompass establishing a quality management system, keeping technical documentation, maintaining automatic logs, conducting conformity assessments, adhering to registration obligations, implementing corrective actions in case of non-compliance with Chapter 2 requirements, informing national authorities upon deploying an AI-system, and, when applicable, taking appropriate corrective measures. When considered in the context of human rights, these obligations on providers of high-risk AI-systems may be sought to be realized through

[1385] European Commission, *supra* note 12 at Article 14(4) states that: *"The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances: (a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible; (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons; (c) be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available; (d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system; (e) be able to intervene on the operation of the high-risk AI"..*

human rights due diligence methods. These obligations would help to examine a discrimination through artificial intelligence claim.

In Articles 18 to 29, a myriad of obligations are set up for high-risk artificial intelligence providers. There seem to be different levels of oversight provisions making different available requirements to be ensured and distributed to different actors involved throughout the AI-lifecycle.

Chapter 3 under Title III of the EU AI Act Proposal is dedicated to notifying authorities and notified bodies that are supposed to be thought for human controlling purposes, making these bodies overseers, responsible for *"setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring"* as stated in Article 30(1). In addition, Articles 31 to 39 thereto regulate the notifying and notified bodies, conformity assessments, and relevant prospects.

Record keeping mandated under Article 12 should also allow for monitoring, especially in conformance with Article 61, titled "post-market monitoring by providers and post-market monitoring plan for high-risk AI-systems," and Article 65, titled "procedure for dealing with AI-systems presenting a risk at national level."

Title VI of the EU Act Proposal is dedicated to governance, which envisages a European artificial intelligence board and national competent authorities. The European Artificial intelligence board aims to provide effective cooperation between the national supervisory authorities and the EU Commission, coordination with the EU Commission and the national supervisory authorities, assisting them with guidance and analysis, and assisting in consistently applying the regulation. Such coordination authorities are essential to conveying stronger protection for cases specific to artificial intelligence before having to go to court.

Article 58 mentions the board's tasks, including collecting and sharing *"expertise and best practices among Member States,"* which may help to contribute to equality and non-discrimination by allowing for a "lesson" learned in one state to apply that expertise. Subparagraph (b) states contributing *"to uniform administrative practices in the Member States, including for the functioning of regulatory sandboxes referred to in Article 53,"* which can also be necessary for the violations of the right equality and non-discrimination stemming from administrative practices to be minimized through such

uniform practices. Subparagraph (c) mentions issuing *"opinions, recommendations or written contributions on matters related to the implementation of this Regulation."* In this direction, issuing opinions and recommendations specific to the right to equality and non-discrimination would be beneficial.

Under Article 59 of the EU Act Proposal, the national competent authorities are to be established and designated by each member state to ensure the application and implementation of the regulation. National competent authorities are required to assign a national supervisory authority which acts as the "notifying authority" and "market surveillance authority". However, states may designate more than one authority if they have adequate reasons, according to Article 59(2). Additionally, Article 59(4) of the EU Act Proposal states that national competent authorities should have the capacity with personnel expertise in artificial intelligence and human rights, and also in risks to health and safety and existing standards and legal requirements. Their status should be notified to the Commission with annual reports, according to Article 59(5). That is important for protecting human rights, as this study has highlighted throughout the study; artificial intelligence presents a legal problem that calls for expertise legally and technically, and human rights law calls for even more expertise. Article 59(7) of the EU Act Proposal states that national competent authorities will guide and advise on the regulation's implementation.

3.2.4 Security and Safety

Safety and security is another important principle that should exist for artificial intelligence.[1386] The safety and security of artificial intelligence are vital for equality and non-discrimination because they are fundamental for avoiding harm and preventing unintentional adverse impacts -potential violations of discrimination- and foreseeing them prior to their instances.[1387] In this regard, safety represents both technically and socially safe systems, even extending to their impact on society.[1388] For the safety and security principle, the regulatory documents refer to names such as "robustness,"

[1386] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 7.
[1387] *Id.*
[1388] *Id.* at 7; UNESCO, *supra* note 1004 at 7 para. 27; DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 8; FJELD ET AL., *supra* note 990 at 38; E.g. AI at Google, *supra* note 1284 at Principle 3.

"security," "safety," and "predictability" of the systems, usually together.[1389] When deciding to utilize artificial intelligence, it is essential to determine whether the potential risks, including those being present for the right to equality and non-discrimination, change by using artificial intelligence, compared to not using it, and whether its utilization exacerbates such risks.[1390]

AI-systems should be safe and secure throughout their lifecycle.[1391] The meaning of this principle is to anticipate the uses and misuses, adverse circumstances and impacts, preventing unreasonable risks for safety and ensure their proper functioning.[1392] Traceability, as explained when explaining transparency, is also required for establishing safe and secure AI-systems.[1393] Reliability relates to security and safety, which are usually mentioned together.[1394] Having reproducible outcomes provides an accurate analysis of the outcomes of the AI-systems.[1395] Safety is also related to accountability due to verifiability and evaluation mechanisms.[1396] Human control is again related to safety, considering the role of individuals or organizations in designing and maintaining AI-systems cautiously.[1397]

The safety principle usually refers to the functioning of AI-systems and preventing unintended harm while the security principle refers to threats outside the AI-system.[1398] There are, of course, many risks involving the security and safety of AI-systems. However, considering the scope of the thesis, this principle aims to address the risks such as biases, discrimination, and social risks that can endanger overall equality in society. With the safety and security principles, risks should be identified, prevented, and eliminated throughout the lifecycle of AI-systems if the risk is realized.[1399] For instance, the EU's White Paper on AI mentions that training data should extend to potentially

[1389] See e.g. HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12; OECD, *supra* note 991; UNESCO, *supra* note 1004; FJELD ET AL., *supra* note 990 at 37.
[1390] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 13.
[1391] FJELD ET AL., *supra* note 990 at 38; E.g. Future of Life Institute, *supra* note 991 at Principle 6.
[1392] OECD, *supra* note 991 at Principle 1.4.a.
[1393] *Id.* at Principle 1.4.b.
[1394] FJELD ET AL., *supra* note 990 at 37–38; See e.g. HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17; *See also* MICROSOFT CORPORATION, *supra* note 1284 at Goals RS1-RS3 "Reliability and Safety Goals" and Goals PS1-PS2 "Privacy and Security".
[1395] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1396] FJELD ET AL., *supra* note 990 at 37–38; AI at Google, *supra* note 1284 at See Principle 3; See Future of Life Institute, *supra* note 991 at Principle 6.
[1397] FJELD ET AL., *supra* note 990 at 38; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1398] FJELD ET AL., *supra* note 990 at 37.
[1399] UNESCO, *supra* note 1004 at 7 para. 27.

dangerous scenarios to address them if they occur, and reasonable measures considering different grounds, such as race, ethnicity, and gender, should be taken to prevent the outcomes of an AI-system does result in discrimination, and additional requirements should be taken for the protection of personal data.[1400] Since the functioning of the AI-systems mainly depends on the training data, for a non-discriminatory safe system, training data should not be contrary to human rights values.[1401] Safety in the context of the EU relates to the trustworthiness of an AI-system, and the application and improvement of other relevant safety legislation besides the existing human rights framework to artificial intelligence could be possible if and when appropriate.[1402]

The CoE "Resolution on Preventing Discrimination Caused by the Use of Artificial Intelligence" mentions that algorithm optimization for efficiency, profitability, or other reasons should consider equality and non-discrimination, otherwise, discriminatory outcomes may occur by reminding the potential discrimination grounds.[1403] The said Resolution states that if there is a potential impact on human rights, respect for equality and non-discrimination should be given regard from the starting their design and conducted tests before and after the deployment to ensure the rights are guaranteed.[1404] In this direction, the approach should aim at substantive equality.

Accuracy and Robustness for Safe and Secure Artificial Intelligence

Robustness is an essential part of a safe AI-system to effectively reach its purpose in the context of discrimination.[1405] If an AI-system is not robust, its accuracy and safety could fail, leading to outcomes; for instance, a healthcare system is not robust if it is incapable of considering all relevant details depending on gender for a disease, thereby resulting in discriminatory outcomes.[1406] Article 15 of the EU AI Act Proposal regulates "accuracy, robustness and cybersecurity," aiming at the safety and security of high-risk AI-systems. This article mentions designing and developing AI-systems to be at a certain level of accuracy and robustness. Accuracy demands transparency in disclosing the levels of accuracy and the relevant metrics for accuracy according to the second subparagraph

[1400] EUROPEAN COMMISSION, *supra* note 519 at 19.
[1401] *Id.*
[1402] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 22; EUROPEAN COMMISSION, *supra* note 519 at 10, 19.
[1403] Council of Europe Parliamentary Assembly, *supra* note 848 at para 4.
[1404] *Id.*
[1405] Tanna and Dunning, *supra* note 128 at 436.
[1406] *Id.*

under Article 15. Article 15(3) of the EU AI Act Proposal states that potential errors, faults, or inconsistencies due to the system or the environment should be able to be addressed by the high-risk systems, especially when they are interacting with humans or other systems. Robustness in AI-systems should be achieved through technical solutions, including "backup" or "failsafe" plans. Regarding feedback loops, Article 15(3) of the EU AI Act Proposal mentions that they should be addressed with mitigation measures to prevent AI-systems' biased outputs from being used as inputs later. Article 15(4) of the EU AI Act Proposal mainly addresses security, referring to the residence in systems for the attempts by third parties to alter the performance of the AI-systems by exploiting vulnerabilities, and notes relevant technical solutions.

The EU embarks on the relationship between the trustworthiness of a system and its safety.[1407] To minimize the risks of harm -in this context discrimination, reasonable measures should be taken.[1408] In this direction, there are requirements for ensuring robustness and accuracy, assuring reproducible outcomes, and ensuring adequate management of mistakes and discrepancies through the AI-lifecycle. [1409] Since technical safety is closely related to the prevention of harm, it is thus an essential requirement for the protection of the right to equality and non-discrimination in AI-systems.[1410] Fallback scenarios for safety-related risks should be done so that the problems are safeguarded.[1411] The AI-system should consult human operators before continuing an action, or decision-making could be conducted based on rules instead of through statistics.[1412]

Assessing and Managing the Safety and Security

Clarifying and assessing potential risks particular to AI-systems is needed for different artificial intelligence application areas.[1413] Risk assessments are thus a requirement for safe and secure AI-systems.[1414] At least the application areas mentioned in the second chapter are required for discriminatory risks. It is mentioned that the level of safety measures should depend on the gratitude of the risks concerning system

[1407] EUROPEAN COMMISSION, *supra* note 519 at 20; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16–17.
[1408] EUROPEAN COMMISSION, *supra* note 519 at 20.
[1409] *Id.*
[1410] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16.
[1411] *Id.*; FJELD ET AL., *supra* note 990 at 40.
[1412] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1413] *Id.*
[1414] FJELD ET AL., *supra* note 990 at 38; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17; OECD, *supra* note 991 at Principle 1.4.c.

capacity.[1415] Safety measures should be firm when there is an apparent risk of discrimination. For example, supposing that an AI-system is used in judicial courts or criminal justice, the measures should be the strongest considering the risks and the system's capability. Accuracy is even more critical when there is a direct impact on human rights, such as equality and non-discrimination.[1416] Safety-related principles are also common in non-binding documents.[1417]

The CoE Draft Convention on AI introduces the safety and security principle under Article 16:

> "Each Party shall, within its jurisdiction and in accordance with its domestic law, ensure that adequate safety, security, data quality, data integrity, data security, cybersecurity and robustness requirements are in place for design, development and application of artificial intelligence systems."

The safety and security principle is envisioned under the CoE Draft Convention on AI as an element of AI-systems that should be ensured throughout the AI-lifecycle, including data-related aspects.

In addition, the CoE Draft Convention on AI has a safety provision regarding health and the environment under Article 11, mandating parties to take necessary measures to protect those. Article 11 is relevant for preserving public health and the environment regarding AI-systems application. This can be relevant to the right to equality in connection to protecting the public health and the environment which are equally important for each individual. Another similar provision is found in the Article 67 of the EU AI Act Proposal which deals with cases where an AI-system complies with the regulation but carries risks to health and safety or EU fundamental rights and the national law on fundamental rights, mentioning that such systems should no longer present risks to fundamental rights after being deployed. This is important considering all of the EU members are also members of the CoE, which the ECHR also applies to all of the members of EU states, so the human rights under the ECHR are also implicitly placed in this provision.

Risk management methods for dealing with risks such as bias, privacy, and digital security should exist throughout the lifecycle of the AI-systems, which should work with

[1415] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1416] *Id.*
[1417] FJELD ET AL., *supra* note 990 at 37.

the technical developments suitable to the context.[1418] The types of biases mentioned in the first chapter should be addressed with such risk management mechanisms by the AI actors, and ensuring such mechanisms' existence can lead to procedural obligations to states for the right to equality and non-discrimination in the context of artificial intelligence.[1419] Risk assessments can be instrumental in determining the contextual benefits and risks of using artificial intelligence for the right to equality and non-discrimination.[1420]

In the EU AI Act Proposal, Chapter 2 of Title III is dedicated to the requirements for high-risk systems; Article 9 thereto envisions a risk management system for high-risk systems with an up-to-date process throughout the AI-lifecycle, it requires:

> "(a) identification and analysis of the known and foreseeable risks associated with each high-risk AI system;
> (b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used according to its intended purpose and under conditions of reasonably foreseeable misuse;
> (c) evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system referred to in Article 61;
> (d) adoption of suitable risk management measures (…)"

This article brings a human oversight, safety, and security mechanism for high-risk systems. Risk management measures should be taken accordingly. That is crucial for protecting the right to equality and non-discrimination to oversee the risks of discrimination. Article 9(4) strengthens this provision and mentions that risk management systems should be so robust that residual risks should be acceptable if the AI-system is used following the intended purpose or under reasonably predictable misusage. According to Article 9(4) of the EU AI Act Proposal, the risk management system should eliminate or reduce the risks "as far as possible" with design and development. The said subparagraph also includes a measure for transparency and explainability, requiring to be explained to users the residual risks. Meaning that design and development, as the first stages in the lifecycle of an AI-system, require the risks for equality and discrimination should be targeted at that stage before being used. It should be noted that "as far as possible" is an ambiguous term.[1421] Mitigation and control measures should be

[1418] OECD, *supra* note 991 at Principle 1.4.b.
[1419] OECD, *supra* note 991.
[1420] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI), *supra* note 517 at 13.
[1421] *See* Smuha et al., *supra* note 106 at 29–30.

implemented for those risks that cannot be eliminated, which implies a safety and oversight measure allowing for intervention in case risk is realized.

Achieving Safety and Security with Technical Measures

The first chapter of this study has explained the methods of artificial intelligence and machine learning and the importance of data. Quality data and better learning methods are essential for safe and secure AI-systems.[1422] That is especially important, considering that most discriminatory outcomes occur due to data and during learning. Where human agents have a role in the AI-system, potential changes in the system should also be considered when addressing safety and security.[1423] Testing and validating for predictability is one of the technical methods for ensuring safety and security by design.[1424] Proactive and continuous testing shall be conducted to prevent unintended and harmful situations.[1425] Therefore, to prevent discriminatory outcomes, testing for potential discriminatory cases should be done; in doing so, the examples such as discriminatory potential and cases can be chosen, and an assessment can be done. Of course, the intended harmful situations should be left entirely out. AI-systems that have predictive functions should predict accurately and decide on correct decisions.[1426] If there are cases where ensuring correct decisions and predictions is not possible, the acceptable inaccuracy level should be determined.[1427] That would help whenever a case of discrimination through artificial intelligence claim comes to a court, in which the judge may look at the level of accepted inaccuracies, which legal frameworks should at least determine. To ensure reliable systems, safety, and security is principle is required.[1428] Likewise, security by design, a specific type of ethics by design (X-by design), is relevant to security and safety to implement compliance with norms at the design stage; for example, addressing threats to attacks on data that could hinder equality and discrimination may be addressed through security by design.[1429]

[1422] UNESCO, *supra* note 1004 at 7 para. 27.
[1423] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 16; DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 8.
[1424] FJELD ET AL., *supra* note 990 at 40; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 22.
[1425] DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 8; See FJELD ET AL., *supra* note 990 at 38.
[1426] DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 8.
[1427] *Id.*
[1428] EUROPEAN COMMISSION, *supra* note 519 at 20; DIGNUM, MULLER, AND THEODOROU, *supra* note 1159 at 8; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1429] FJELD ET AL., *supra* note 990 at 40; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 21.

Technical knowledge and education are underlined concerning the elimination of risks in Article 9(4) of the EU AI Act Proposal, which can suggest the attempts to target risks through technical methods. Article 9(5) of the EU AI Act Proposal includes a provision for testing to identify the most appropriate measures for risk management. Testing should ensure that AI-systems work consistently, suit their intended purpose, and comply with the requirements.

Article 9(6) of the EU AI Act Proposal states that testing should not "go beyond what is necessary" to achieve the intended purpose of an AI-system. Article 9(7) of the EU AI Act Proposal asserts that testing for preliminarily defined metrics and probabilistic thresholds appropriate for the intended purpose of the AI-system should be done at any point of the development stage as appropriate, and in any event, before market placement or starting to service and for a discriminatory situation, considering the purpose of the AI-system. That could be exemplified in an area mentioned in the second chapter of this study, such as healthcare. In such a case, metrics and probabilistic thresholds specific to the purpose of the healthcare system, which are defined for potential discriminatory cases should be tested during the development stage and before the market placement.

Article 10(1) of the EU AI Act Proposal is about data and data governance in high-risk systems using training models with data, which involves criteria that should be met in training, validation, and testing. The EU AI Act Proposal Article 10(3) states that AI-systems' training, validation, and testing should be statistically appropriate, relevant, and representative of the persons for whom the AI-system will be used, hinting at the possibility of a contextual approach. This article promotes addressing the biases, especially representation bias. Geographical, behavioral, and functional settings of the AI-systems are again crucial for the testing, as Article 10(4) states. That also applies to discriminatory cases, where, for example, in specific geographies, discriminatory cases may differ in grounds or the impacted individuals. The EU AI Act Proposal Article 10(5) has a specific measure for bias monitoring, detection, and correction in high-risk systems, further noting personal data-related facets.

Article 40 of the EU AI Act Proposal is titled "standards, conformity assessment, certificates, registration," expressing harmonized standards for high-risk AI-systems. A specific title under the EU AI Act Proposal is assigned for measures supporting innovation. Article 53 under Title V of the EU AI Act Proposal is about regulatory

sandboxes. Regulatory sandboxes aim to provide *"a controlled environment that facilitates the development, testing, and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan*" as defined under Article 53(1). Competent authorities should guide and supervise the regulatory sandboxes to ensure compliance with the regulation. According to Article 53(3) of the EU AI Act Proposal, if risks to health and safety and human rights are diagnosed with these regulatory sandboxes, they should be mitigated. If that is failed, the development and the testing can be stopped. There is a liability provision for the participants of artificial intelligence regulatory sandboxes for harm to third parties due to experimentation in the sandboxes.

Data Protection and Privacy for Safe and Secure Artificial Intelligence

Data protection and privacy are relevant notions to safety and security principle.[1430] However, privacy is also considered as a separate principle for AI-systems.[1431] Privacy is an important aspect of preventing discrimination through artificial intelligence.[1432] In this direction data protection laws and the fulfillments under the GDPR and the CoE Convention No. 108 should not be forgotten for a secure AI-system in the context of the right to equality and non-discrimination.[1433] Article 13 of the CoE Draft Convention on AI introduces privacy and personal data protection as a principle. Likewise, provisions concerning data protection are mentioned throughout the EU's AI Act Proposal. Collected data for AI-systems should not cause discrimination.[1434] Article 54 of the EU AI Act Proposal concerns personal data processing in the regulatory sandboxes. Article 55 of the EU AI Act Proposal is an article that contributes to the equality of the artificial intelligence providers, supporting measures for them such as providing small-scale providers and start-ups access to regulatory sandboxes priority and awareness-raising activities for them. Ensuring privacy and data protection in the AI-systems may help to overcome the black-box effect to a certain degree as it also relates to transparency. [1435]

[1430] FJELD ET AL., *supra* note 990 at 21; OECD, *supra* note 991 at See Principles 1.2 and 1.4.
[1431] FJELD ET AL., *supra* note 990 at 21; UNESCO, *supra* note 1004 at 8 paras. 32-34; HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1432] FJELD ET AL., *supra* note 990 at 21.
[1433] *Id.*
[1434] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 17.
[1435] ALLEN QC AND MASTERS, *supra* note 804 at 56.

3.2.5 Connecting the Dots: Equality-Related Principles

The landscape for non-binding documents reveals a rush to "ethical artificial intelligence," although its meaning and requirements lack a consensus.[1436] Human rights are at the center of ethics -the non-binding documents-, and legal regulations, providing a bridge between the two, as the notion has two aspects, one being the moral and humanity, and the second one being the framework of human rights law.[1437] The references to human rights in non-binding documents do not always align with the legal framework of human rights.[1438] The realization of human rights depends on the legal documents and human rights mechanisms.[1439] However, even in instances when non-binding documents on artificial intelligence refer to human rights, this may be far from fulfilling the requirements in human rights law. As discussed in the first chapter on fairness in artificial intelligence, the understanding of human rights may also change depending on the perspective. Since novel outlooks of the right to equality and non-discrimination may arise in artificial intelligence, there may be a necessity for new ways of protecting human rights.[1440] In doing so, human rights, human dignity, societal well-being, fairness, equality, non-discrimination, and diversity and inclusion come up as principles under the regulatory works on artificial intelligence.

The UNESCO Recommendation on AI refers to respect for human rights mechanisms and frameworks during the whole of the AI-lifecycle. [1441] It reminds artificial intelligence actors and member states human rights should be protected and promoted in accordance with national and international human rights law.[1442] That demonstrates that non-binding documents refer to existing human rights law, and the protection of human rights should be accomplished through the co-working of the non-binding documents as complements of the existing legally binding documents, even in cases legally binding regulations specifically addressing artificial intelligence is absent.[1443] Such references to human rights law documents are crucial because even if the legally binding documents

[1436] Ienca and Vayena, *supra* note 972 at 38.
[1437] Smuha et al., *supra* note 106 at 7.
[1438] *Id.* at 6–7.
[1439] *Id.*
[1440] Amnesty International and Access Now, *supra* note 1006 at 15.
[1441] UNESCO, *supra* note 1004 at 6 para. 16.
[1442] *Id.* at Principle 42.
[1443] *Id.*

are valid and applicable even without such mentioning, reminding the private actors is essential. Non-binding documents refer to the ethical responsibility of the artificial intelligence actors wherever they have a role in the lifecycle of the AI-systems.[1444]

The documents referring to redress and remedy mechanisms bring protection for the right to equality and non-discrimination, considered in connection with the right to fair trial.[1445] Ienca and Vayena had diagnosed five approaches in non-binding documents on justice and fairness in artificial intelligence, which applies to the right to equality and non-discrimination: standard setting and recommended practices; awareness raising for existing rights and regulations; system testing, monitoring and auditing; promotion of the rule of law and implementation of redress and remedy mechanisms; systemic changes such as governmental oversight, interdisciplinary work, bringing different stakeholders together.[1446]

In this direction, human rights due diligence frameworks could help incorporate the private sector to take action for equality, as supported in the CoE "Recommendation on the Human Rights Impacts of Algorithmic Systems" if adopted by the private sector actors throughout the AI-lifecycle to ensure that all lifecycle steps do not result in discrimination.[1447] Toronto AI Declaration also addresses human rights due diligence as a responsibility of private actors.[1448] It consists of three steps: identifying potential discrimination, taking "effective action" to prevent and mitigate discrimination, and transparency in their efforts regarding the first two.[1449]

The CoE Draft Convention on AI has numerous provisions regarding equality. Article 5 of the CoE Draft Convention on AI sets forth obligations for public authorities:

> "Each Party shall, within its respective jurisdiction, ensure that:
> a. the application of an artificial intelligence system substantially informing decision-making by a public authority in the exercise of its function, or any private entity acting on its behalf, is fully compatible with its obligations to respect human rights and fundamental freedoms as guaranteed under its domestic law or under any relevant applicable international law;

[1444] *Id.*
[1445] Ienca and Vayena, *supra* note 972 at 51.
[1446] *Id.*
[1447] Council of Europe Committee of Ministers, *supra* note 991.
[1448] Amnesty International and Access Now, *supra* note 1006 at paras. 42-51.
[1449] *Id.*

> b. any interference with human rights and fundamental freedoms by a public authority or any private entity acting on its behalf resulting from such application of an artificial intelligence system is compatible with core values of democratic societies, in accordance with the law and necessary in a democratic society in pursuit of a legitimate public interest."

Article 5 obliges the states to ensure that either the public authority or the private entity is fully compatible with its obligations to respect human rights and freedoms under relevant domestic and international laws. Here an *obligation to respect* is imposed. Moreover, the article itself refers to other human rights treaties. This means these conventions' obligations may be interpreted if a case comes to an ECtHR. For example, the state may be examined whether it has taken the necessary measures to realize full compatibility. Subparagraph b of Article 5 asserts that any interference with human rights as a result of the application of AI-systems by both public authorities and any private entity should be compatible with *"values of democratic societies, in accordance with the law and necessary in a democratic society in pursuit of a legitimate public interest."* This article thus delivers an interference regime comparable to existing human rights law. In this direction, the legitimate aim would be "public interest". What constitutes a public interest is not detailed, leaving a lacuna.

The CoE Draft Convention on AI brings a requirement regarding human rights under Article 6, titled "Requirements regarding respect for human rights:"

> "Each Party shall take such measures as are aimed at minimising and, to the extent possible, preventing any unlawful harm or damage to human rights and fundamental freedoms, which could result from the inappropriate application of artificial intelligence systems by public authorities."

The approaches thereunder are determined as *"to the extent possible*," and *"minimizing"* is tricky regarding a provision directly addressing human rights. It leaves a lacuna because states will determine the application of this convention, and due to such a provision, states may say the measures they have taken are *"aimed at minimizing,"* where, in fact, they do not. Secondly, the acceptance of taking measures "*to the extent possible"* shows already an acceptance of a defeat from the technology field that artificial intelligence is inevitable, and it is the *status quo* leaving no room for questioning the booming of artificial intelligence. Thirdly, such an approach is not quite firm. The provision already starts with acknowledging that there can be cases where the failure to prevent unlawful harm or damage is valid and accepted. The wording of the

provision itself weakens the protection level of the convention. The right to equality and non-discrimination is the most vulnerable right to being affected by artificial intelligence. As explored thoroughly in the thesis, it can be affected in many ways. Accordingly, such a provision in a legal document does not bring enough protection for the right to equality. It could easily result in letting go of responsibility in case of human rights violations. For example, given that discrimination occurs in AI-systems, if a state has previously mandated each artificial intelligence developer to adopt an ethics guide and brings a rule to minimize human rights deficiencies in AI-systems by this provision, it could be said that the measures aimed at the prevention of human rights violations. In this direction, how to establish and which criteria to pick to determine the appropriateness of the measures to minimize and prevent unlawful harm or damage to human rights should be determined.

Article 7 of the CoE Draft Convention on AI is titled "Requirements regarding respect for democratic institutions and the rule of law" and declares:

> "Each Party shall take all necessary measures to preserve the integrity of democratic institutions and processes and ensure the respect for the rule of law and proper administration of justice in the context of application of an artificial intelligence system. To this end, a thorough assessment of the necessity, proportionality and potential risks shall be carried out by a competent domestic authority both prior and during the application of an artificial intelligence system to the elaboration or revision of laws and policies governing the respective roles of the executive authority, other democratic institutions and the judiciary."

This article is relevant to the right to equality concerning its connection to the rule of law and democratic institutions.

Chapter III of the CoE Draft Convention on AI is about applying AI-systems in providing goods, facilities, and services, starting with Article 8, which regulates the obligations relating to public and private actors. This article brings the application of an AI-system to be in accordance with domestic and international law because they bring obligation to respect human rights. The provisions listed thereunder are mainly related to social and economic rights, as it mentions the AI-systems concerning *"health, family care, housing, energy consumption, transport, food supply, education, employment, finance, environmental protection, digital information, media and communication,"* also reminding the areas of artificial intelligence use listed in the second chapter of this

study. The first subparagraph of the article states that AI-systems should abide by the existing human rights law, which is an obligation that should already be realized. Therefore, while being an essential provision, it carries only a reminder nature. The second paragraph of Article 8 mandates states to provide effective guidance on *"how to prevent and mitigate any adverse impacts of the application of an artificial intelligence system on the enjoyment of human rights and fundamental freedoms."*

Defending Human Dignity in Artificial Intelligence

Human dignity is among the provisions relating to human rights in general that could be relevant for the right to equality and non-discrimination. Human dignity is addressed in the OECD AI Recommendation under the title "human-centred values and fairness" solely by its name, while in the UNESCO AI Recommendation, a detailed definition of human dignity is followed by relevant provisions mentioning that humans should not be harmed or subordinated in any form such as mental, physical, social, economical in any stage of AI-lifecycle.[1450] Respect for human dignity is also found in the EU Ethics Guidelines for Trustworthy AI, mentioning that the development of AI-systems should *"respect, serve and protect"* humans to fulfill their basic needs, physical and mental integrity, and personal and cultural identity.[1451] In addition, the EU Ethics Guidelines for Trustworthy AI further notes that human dignity of the humans should be respected equally.[1452] The Guidelines state that such understanding moves further than non-discrimination by the discrimination grounds.[1453] While that is accurate, how equal respect for human dignity in artificial intelligence would be realized is unclear in the Guidelines, despite the principles and methods explained thereunder.

Article 9 of the CoE Draft Convention on AI is dedicated to the *"preservation of individual freedom, human dignity, and autonomy,"* underlining the importance of "*the ability to reach informed decisions free from undue influence, manipulation or detrimental effects"* that could negatively impact human rights. Article 9 could be illustrated when considering the right to equality and non-discrimination with an example mentioned prior in the study. Suppose a machine is trained on biased data within a healthcare system. The decision-making process of the doctor or the patient could be

[1450] OECD, *supra* note 991 at Principle 1.2.a; UNESCO, *supra* note 1004 at 6 paras. 13-14.
[1451] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 10.
[1452] *Id.* at 11.
[1453] *Id.*

impacted. Alternatively, if a judge is deciding on a case with the help of an AI-system, their judgment can be affected by it. This article obliges states with a positive *obligation to protect*.

Algorithmic processes also have the potential to be manipulative and influence people's emotions and may also hinder equality in society and lead the way to discrimination as such.[1454] Regarding the EU's approach, Article 5(1)(a) of the EU AI Act Proposal mentions that mental manipulation without cognizance is a prohibited act to protect human autonomy and dignity. Article 5(1)(b) of the EU AI Act Proposal remarks that an AI-system exploiting the vulnerabilities of a group to distort their behavior mentally should not be placed into the market. While not a direct provision on discrimination, human dignity related provisions are crucial for ensuring equality.

Protecting Societal Well-Being Whilst Adopting Artificial Intelligence

The EU Ethics Guidelines for Trustworthy AI also emphasize societal well-being extending to environment in the context of artificial intelligence in light of the fairness principle throughout the AI-lifecycle.[1455] Societal well-being also encompasses the effects of artificial intelligence on democratic institutions.[1456]

The EU AI Act Proposal refers to societal well-being in the Explanatory Memorandum, mentioning that artificial intelligence should be a tool for improving the well-being of humans.[1457] Accordingly, societal well-being would require a human-centric approach to artificial intelligence, ensuring trust in the technology. [1458] Societal and environmental well-being is addressed in the May 2023 proposed amendments to the EU AI Act Proposal, which suggests the inclusion of well-being as a general principle for all AI-systems, viewing that they should be designed and operated to benefit all humans and their long-term impacts on society, democracy, and the individuals should be evaluated in parallel to EU HLEG's assessment list about artificial intelligence.[1459] Similarly, the CoE Draft Convention on AI notes societal well-being in the context of

[1454] Council of Europe Committee of Ministers, Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes, Decl(13/02/2019)1 paras. 6, 8, 9 (2019), https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b (last visited Aug 4, 2022).
[1455] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 19.
[1456] *Id.*
[1457] European Commission, *supra* note 12 at Explanatory Memorandum 1.
[1458] *Id.*
[1459] BRANDO BENIFEI AND IOAN-DRAGOȘ TUDORACHE, *supra* note 1131 at 143; High-Level Expert Group on AI (AI HLEG), *supra* note 1000 at 19.

innovation, considering the potential of artificial intelligence to foster societal well-being.[1460]

OECD AI Recommendation is another document addressing societal well-being and human-centered values.[1461] Accordingly, artificial intelligence stakeholders should take action in safeguarding beneficial outcomes for humans and the planet by ensuring social, gender and economic equalities, protecting human competency, supporting inclusion, caring the environment.[1462] Human-centered values include respecting human rights, equality and non-discrimination, diversity, fairness, privacy, the rule of law, democratic values, and social justice.[1463] Similarly, Asilomar AI Principles include several provisions relating to societal well-being, constructing them with human values such as dignity, human rights, and freedoms and aiming at benefiting the greatest number of persons and humanity.[1464]

Artificial intelligence could fail to benefit humans if human and societal well-being is not accomplished.[1465] Concerns for societal well-being could prevail even in legally handled and safe systems.[1466] Human rights could be hindered if artificial intelligence technologies are improved without considering society members' mental and physical health.[1467] Therefore, societal well-being should be prioritized in AI-systems.

Ensuring Diversity and Inclusivity in Artificial Intelligence

Diversity and inclusion, which have significance for reaching substantive equality, are mentioned in some of the documents.[1468] From the standpoint of this study, substantive equality should be the aim and practices that will ensure that it is required in all stages of artificial intelligence, including design, development, and implementation, hence, promoting diversity and inclusion in artificial intelligence documents is meaningful in this direction.[1469]

[1460] Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, *supra* note 13 at Preamble para. 4.
[1461] OECD, *supra* note 991 at Principles 1.1-1.2.
[1462] *Id.* at Principle 1.1.
[1463] *Id.* at Principle 1.2.
[1464] Future of Life Institute, *supra* note 991.
[1465] IEEE, *supra* note 991 at 21.
[1466] *Id.*
[1467] *Id.*
[1468] Ienca and Vayena, *supra* note 972 at 51.
[1469] *Id.*

Aiming for the public benefit in artificial intelligence research could support substantive equality if the interests of marginalized and vulnerable individuals and groups are considered in the relevant processes, as pointed out by the CoE "Recommendation on the Human Rights Impacts of Algorithmic Systems" as an obligation of the states.[1470]

Inclusive approach to ensuring fairness in artificial intelligence could extend to availability and accessibility of artificial intelligence to different groups and individuals.[1471] For example, the UNESCO AI Recommendation notes that should be done by taking "specific needs" of different groups such as people with disabilities, marginalized and vulnerable persons, girls and women, into consideration.[1472] UNESCO AI Recommendation further mentions equity among individuals despite differences in grounds such as race, ethnicity, gender, in the context of access and participation the AI-lifecycle.[1473] That indicates the other outlook of equality for artificial intelligence, that is sharing the benefits of artificial intelligence fairly in different parts of the world among different states.[1474]

In UNESCO's approach to diversity and inclusivity in artificial intelligence, a very substantive equality approach is seen, as it is said to be *"done by promoting active participation"* of persons belonging to and from different backgrounds, extending to considering different thoughts, expressions, beliefs and lifestyles during the whole AI-lifecycle.[1475]

A substantive equality approach for ensuring inclusivity throughout the AI-lifecycle can be made by extending inclusivity in the education of artificial intelligence, which is suggested by the CoE Resolution on "Preventing Discrimination Caused by the Use of Artificial Intelligence" with a specific mention to girls, women, and minorities, endorsing research on the bias as well as promotion of digital literacy for accessing to all society.[1476] Similar provisions on diversity, inclusivity, and equity for ensuring the right to equality

[1470] Council of Europe Committee of Ministers, *supra* note 991 at B.6 under the "Guidelines on addressing the human rights impacts of algorithmic systems"".
[1471] UNESCO, *supra* note 1004 at 7–8 para. 28.
[1472] *Id.* at 28.
[1473] *Id.*
[1474] *Id.* at paras. 28, 30.
[1475] *Id.* at Principles 19-20.
[1476] Council of Europe Parliamentary Assembly, *supra* note 848 at paras. 6, 12.

and non-discrimination in artificial intelligence throughout the AI-lifecycle are set in the Toronto AI Declaration by mentioning the participation of different groups.[1477]

Inclusivity, diversity, and fairness are addressed in the EU Ethics Guidelines for Trustworthy AI by mentioning avoiding bias, accessibility to AI-systems, and stakeholder participation during the whole stages of the AI-lifecycle.[1478] The EU Ethics Guidelines for Trustworthy AI also mention an "equal opportunity" to access education, good services, and technology.[1479]

Article 7(1)(f) of the EU AI Act Proposal includes a specific measure to update the list of high-risk AI-systems due to their impact on vulnerable groups due to age, disabilities, and physical or mental to use that vulnerability. Social scoring is in general prohibited by public authorities, which can lead to discriminatory outcomes under Article 5(2)(c) of the EU AI Act Proposal. Article 5(1)(d) of the EU AI Act Proposal limits the real-time biometric identification is limited to a strict application area with considerations to context, an essential measure for preventing potential discriminatory behavior such as profiling by public authorities. These measures aim to consider diversity in equality.

Article 10 of the CoE Draft Convention on AI is dedicated to "access to public debate and inclusive democratic processes", introducing a positive obligation to fulfill that that can enhance the right to equality, and help to prevent from non-discrimination by incorporating "*all interested parties, groups and individuals enjoy equal and fair access to public debate and inclusive democratic processes"*

Article 18 of the CoE Draft Convention on AI is titled "public consultation and additional measures," bringing a socially conscious and action-directed provision that could help to protect and realize the right to equality and non-discrimination. It could allow the discriminated individuals to come together, have a voice, and raise awareness for artificial intelligence-related issues. It may help the artificial intelligence developers to be aware of these. Promoting digital literacy and skills for everyone will surely educate individuals on artificial intelligence and thus, help achieving substantive equality.

[1477] Amnesty International and Access Now, *supra* note 1006 at paras. 18-21.
[1478] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 18–19.
[1479] *Id.* at 12.

Aiming at Fairness in Regulating Artificial Intelligence

Fairness, preventing and mitigating biases that may result in discrimination are commonly mentioned in the context of justice in the non-binding documents.[1480] Yet, achieving justice and fairness differs among these documents as mentioned in the first chapter.[1481] As discovered in the first chapter of this study, fairness may have different meanings, and it will continue to do so when the field leaves out legal regulation. Therefore, for the current scene, non-binding impact assessments include adopting a fairness definition, why that definition is preferred, whether this definition is included in setting up the AI-system, consultation with potentially impacted people, a quantitative measurement for that definition, and mechanisms for ensuring fairness.[1482] For example, a system design that will give suggestions based on women's shopping habits in the online market may continue to suggest things that most women think and see so far that they would prefer. Alternatively, if policing in places regarded as "criminal activity areas" gives more “effective” results in catching criminals, it may not enable the polices to switch to a mechanism that will observe that place, for example, make the policing system stop matching with the location. One may follow the measures when there is a legal consequence, such as sanctions, in return if s/he is not complying with the measures. Even if one complies with ethical guidelines, assessment lists, and ethical principles set by their institutions and audits, one should know they may face legal consequences.

Fairness is recognized as a principle in the EU Ethics Guidelines for Trustworthy AI, touching on its substantive and procedural dimensions.[1483] To realize the first, benefits and costs should be distributed equally and just, free from discrimination, stigma, and bias.[1484] To realize the latter, it mentions effective redress mechanisms for decisions made by AI-systems or their human operators.[1485] In this direction, the Assessment List for the EU Ethics Guidelines for Trustworthy AI also includes questions concerning fairness such as the preferred definition of fairness, stakeholder consultation and analysis of fairness metrics.[1486] UNESCO AI Recommendation has a heading “Fairness and Non-discrimination” which address diversity, bias and discrimination and digital

[1480] Ienca and Vayena, *supra* note 972 at 51.
[1481] *Id.*
[1482] High-Level Expert Group on AI (AI HLEG), *supra* note 1000 at 16–17.
[1483] HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), *supra* note 12 at 12.
[1484] *Id.*
[1485] *Id.* at 13.
[1486] High-Level Expert Group on AI (AI HLEG), *supra* note 1000 at 17.

knowledge.[1487] OECD on the other hand mention fairness next to "human values" by solely calling its name without further detail.[1488] Finally, none of the forthcoming legally binding documents on artificial intelligence from the EU or the CoE mention fairness as a principle in their provisions.

Formulating Equality and Non-Discrimination as Principles for Artificial Intelligence

When nearly every document, including non-binding ones, references the discriminatory aspects of artificial intelligence, and many documents mention equality, this may not always imply equality and non-discrimination in the legal sense as defined under human rights law. This is also one of the reasons why the different meanings and understandings of equality in law have been examined. This calls for addressing discrimination through artificial intelligence from the perspective of human rights law.[1489] Considering the long-term effects of the AI-system is again relevant for ensuring equality and fairness, especially on a social level, during the design and deployment steps of the AI-lifecycle.[1490]

When the non-binding documents are inquired for principles on equality and non-discrimination, Ienca and Vayena observe that it is a more frequent concern in academic and governmental documents compared to those from the private sector.[1491]

Non-discrimination is itself listed as a principle that is expected to be achieved.[1492] Nevertheless, while it is a principle, it is to be accomplished with the existence of all the other criteria. It is like a bouquet with its standing but with various other criteria. Non-discrimination could be realized with and through inclusion and diversity.[1493] It is also required at the whole lifecycle of an AI-system.[1494] Avoiding introducing and amplifying biases should be aimed when input selection and during the design.[1495] There should be testing and monitoring tools for existing or potential biases.[1496] Discrimination grounds

[1487] UNESCO, *supra* note 1004 at 7–8 paras. 28-30.
[1488] OECD, *supra* note 991 at Principle 1.2.
[1489] See McGregor, Murray, and Ng, *supra* note 1; Smuha, *supra* note 19.
[1490] FJELD ET AL., *supra* note 990 at 58.
[1491] Ienca and Vayena, *supra* note 972 at 51.
[1492] See HIGH-LEVEL EXPERT GROUP ON AI (AI HLEG), THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) FOR SELF ASSESSMENT. (2020), https://data.europa.eu/doi/10.2759/791819 (last visited Aug 6, 2022).
[1493] *Id.* at 16.
[1494] High-Level Expert Group on AI (AI HLEG), *supra* note 1000.
[1495] *Id.* at 16.
[1496] *Id.*

may be tested ahead of time for specific target groups. There, of course, is a need for a mechanism for informing biases.[1497]

The UNESCO AI Recommendation mentions that *"AI actors should make all reasonable efforts"* to minimize and avoid reinforcing or perpetuating discrimination or biased outcomes throughout the AI-lifecycle as well as providing an effective remedy.[1498] While addressing "effective remedy" is plausible, the said approach of "all reasonable efforts" and the choice of the words "minimization" and "avoiding" is concerning for their capacity to contribute to genuine equality.[1499]

The Toronto AI Declaration addresses the prevention of discrimination with numerous provisions, addressing both states and private actors in their responsibility to "proactively prevent discrimination," and when prevention is not possible, instant addressing of the harms with effective remedies is required to conform with human rights law, reminding that existing discriminatory states and inequalities may be exacerbated and propagated.[1500]

Toronto AI Declaration refers to the human rights obligations of the states mentioning that they should not engage in or support AI-systems in a public context that are violating the right to equality and non-discrimination.[1501] It also reminds the states their positive obligations to protect individuals from discrimination by private actors. [1502]

There are various provisions found concerning equality in the CoE Draft Convention on AI. The equality and non-discrimination are mentioned therein at two different times.. Article 3 of the CoE Draft Convention on AI mentions the subject of this thesis, that is, the principle of equality and non-discrimination. First, in Article 3, where it is introduced as a blanket and general equality provision, and secondly, in Article 12 under Chapter IV as a principle. Article 3 is a standalone provision and aims to be an umbrella for the whole convention. The provision lists out some discrimination grounds and underlines that it is not exhaustive; these grounds include *"gender, sexual orientation, race, color, language, age, religion, political or any other opinion, national or social origin, association with a national minority, property, birth, state of health, disability".* The provision addresses

[1497] *Id.*
[1498] UNESCO, *supra* note 1004 at 8 para. 29.
[1499] *Id.* at 29.
[1500] Amnesty International and Access Now, *supra* note 1006 at paras. 14-17.
[1501] *Id.* at para. 22.
[1502] *Id.* at para. 24.

intersectional discrimination, which may often occur in AI-systems, by stating that "based on a combination of one or more of these grounds."

Article 12 under the Chapter IV of the CoE Draft Convention on AI presents the "principle of equality and anti-discrimination", instructing the equality as a separate principle:

> "Each Party shall, within its jurisdiction and in accordance with its domestic law, ensure that the design, development and application of artificial intelligence systems respect the principle of equality including gender equality and rights related to discriminated groups and individuals in vulnerable situations."

This article is the most relevant about the right to equality and non-discrimination. Extending the range to *"design, development, and application processes"* is good because, as explored in the study, the right to equality and non-discrimination can be hindered and violated at any stage. Article 12 mandates parties *to "ensure that (...) AI systems respect the principle of equality".* The first issue with the wording of the article is while "ensuring" refers to an *obligation to fulfill*, "respecting" refers to an *obligation to protect*. Secondly, how and with which measures the AI-systems can respect the principle of equality is not clear. If such measures are not explained, this article will be less impactful, and their fate may end up similar to the user. What are the measures for respecting the right to equality and non-discrimination, and what can a state do to say that this obligation is conducted should be elaborated. Thirdly, there needs to be more clarity regarding how to make a system respect the principle of equality, reminding the artificial intelligence personhood discussions and leaving a lacuna for the attribution of responsibility when the system does not respect the right to equality. There can be cases where the state takes measures, but the AI-system may still not respect the right to equality. In such a case, the state may not be at fault, yet there can be no other responsible to hold accountable. Fourthly, the extent of respect for the equality principle needs to be elucidated. How and when is the principle of equality (not) respected by an AI-system?

Another problem in the CoE Draft Convention on AI is that the meaning of the principle of equality is not defined thereunder. Differences in the meaning given to equality are more of a hypothetical question in today's world, yet still, a prospective party to this convention may have recognized the respect for equality bar low in their domestic law. Another problem here is that "within its jurisdiction." Indeed, it would conflict with

international law and even human rights law to oblige a state more than its jurisdiction. However, artificial intelligence can be designed and developed somewhere and then applied outside that jurisdiction. Since this article is not only limited to public authorities and applies to states to take the relevant measures for the private sector too, where its ability to oversee is limited. If a state that the AI-system is applied is also a party to this convention, then the answer could be answered on some level since the application of the artificial intelligence is in states parties to the convention, and they are obliged to take measures for the application of artificial intelligence to respect equality. Nevertheless, the potential outcomes are ambiguous if a state is not a party to this convention concerning the jurisdiction because the artificial intelligence can be applied outside of a state's jurisdiction.

CONCLUSION

Artificial intelligence has transformed human lives. This transformation is reflected in the legal field as human rights impacts. One of the human rights impacts of artificial intelligence relates to the right to equality and non-discrimination. In this direction, this study examines artificial Intelligence in terms of the right to equality and non-discrimination. The study builds on the view that equality and discrimination-related matters in the context of artificial intelligence should be dealt with as a human rights law problem, which is insufficient in the current landscape of artificial intelligence. Thus, providing a legal perspective to artificial intelligence requires adapting it to human rights law. In doing so, international human rights law could be applied as is, yet it could also be improved. Accomplishing that requires scrutinizing the legally binding regulative preparations of artificial intelligence and the non-binding documents to get inspiration from the AI-ethics, diagnosing their weaknesses, and then reinforcing them as legal matters by looking at the regulatory frameworks on artificial intelligence.

The first chapter of the study presents the definitions relevant to artificial intelligence and, building on them, makes essential explanations on the right to equality and non-discrimination in the first step. The study builds on the approach of handling artificial intelligence with human rights law and supports that discrimination through artificial intelligence yields a right to equality and non-discrimination problem. The emphasis on handling discrimination through artificial intelligence as a legal problem stems from the deficiencies in addressing the issue as a legal matter. The preference for adopting the term artificial intelligence is, as also implied by the EU and the CoE, the technically changeable definitions of artificial intelligence should not limit the law and that it affects human life in ways that may include different methods, emphasizing inhumanity as it will not make decisions that affect human life independently of humans. Whether artificial intelligence makes direct decisions or is used as part of a system, it is called artificial intelligence, referring to its non-human feature. Indeed, artificial intelligence as machine learning is different from general artificial intelligence and calling each automated system artificial intelligence may not be correct. Nevertheless, making rigid preferences about which sub-branch or exactly which form of artificial intelligence is not essential when dealing with questions about equality and discrimination. The study in the first chapter

mentions the overviews of other relevant terms, such as machine learning and deep learning, to bring a reader from the legal background to provide a basic understanding of the upcoming discussions. Explanations of the types of biases are due for the same reason: to show how the bias that causes discrimination may take forms in artificial intelligence.

Apart from this, since human rights concerns are perceived as more of an ethical problem in the literature, the notion of fairness comes to the fore when mentioning equality and non-discrimination in artificial intelligence. In this direction, the study probes fairness discussions to find the equivalents of these concepts in law and how equality and discrimination are formulated in the legal framework to reconcile the notions. In conclusion, the issue of equality and fairness in intelligence is linked to the right to equality and the prohibition of discrimination. However, it should be noted that the notion of fairness, which is frequently preferred in the artificial intelligence field, does not express its entire legal applicability. However, it could strengthen the law's view on discrimination through artificial intelligence. Equality, both as a principle and as a right, is one of the most fundamental elements of human rights law. While reminding the essentials of the right to equality and the non- discrimination, the necessity of using these concepts in artificial intelligence to match their legal meanings is underlined because, ultimately, when there is a judicial problem, their meanings in law will manifest.

The second chapter presents how artificial intelligence might lead to discrimination in two steps. In the first stage, artificial intelligence was examined according to its different areas of use, and then by illustrating discrimination under the grounds for discrimination recognized in international human rights law. The study observed that since the use of artificial intelligence in various fields is already present, it may lead to discrimination in numerous ways emphasizing the urgent need for improving legal understanding in the artificial intelligence field. In this direction, these different areas of artificial intelligence use were a selection encompassing healthcare, judicial systems, criminal justice, employment, education, online platforms, generative artificial intelligence, and online marketing. Since discrimination might come about in diverse ways in such areas of use, examining with examples has demonstrated the gravity of the subject matter. The types of biases noted in the first chapter have been exemplified, and a basis has been made so the reader can understand how discrimination through artificial

intelligence might occur. Afterward, discrimination cases through artificial intelligence were examined in light of the discrimination grounds accepted in human rights law.

The other step in the second chapter deals with how the discrimination examination in international human rights law is carried out to examine how an examination could be made if a discrimination case goes to the ECtHR. While doing this, an adaptative analysis is made by taking on the discrimination test of the ECtHR. Since the ECHR is a living instrument, it has the prospect of adapting to new forms of discrimination in the context of artificial intelligence. In this respect, the study determined new set of positive and procedural obligations may arise for artificial intelligence. Ensuring non-discrimination in artificial intelligence presents a body of such obligations on states extending to effective procedural mechanisms. This also connects to regulations and principles specific to artificial intelligence discussed in the third chapter. The second chapter of the study concludes that the discrimination framework of international human rights law is ready and available for application to artificial intelligence. That being said, particular challenges exist, such as the subsidiary nature of the right in the majority of the international human rights law documents and not being able to claim a social rights related discrimination in the context of the ECtHR, the momentum and gravity of the growth of artificial intelligence, the variety in the ways of potential violation of the right to equality and non-discrimination. These challenges and the ongoing regulatory preparations point out the need for an examination of the right to equality and the prohibition of discrimination in artificial intelligence specific regulatory documents.

The third chapter primarily provides an overview of the regulatory works and preparations for artificial intelligence in the context of the right to equality and non-discrimination. The need for regulating artificial intelligence may stem from different reasons. It should be noted that artificial intelligence will have various legal consequences besides human rights issues; hence, they must be examined separately. In the context of the present study, it is seen that the regulation of artificial intelligence in terms of the right to equality and non-discrimination is a topic continuously discussed. Nevertheless, there is a prevailing trend to address problems about the right to equality and non-discrimination in artificial intelligence, mainly through non-binding regulations, which are both worthwhile and risky. Non-binding regulations are valuable because they could improve the existing legal frameworks, offer theoretical insights, incentivize the artificial

intelligence field to take action, and help them take responsibility. Currently, no international legally binding regulation is devoted directly to artificial intelligence in the legal realm.

On the other hand, the EU and the CoE, being influential globally in their regulatory impacts, have made substantial progress in their legally binding regulatory preparations on artificial intelligence. The regulatory works are examined based on the theoretical framework in the non-binding regulations and in light of the principles for artificial intelligence encompassing transparency and explainability, accountability and responsibility, human control and oversight, security and safety, and equality and fairness. The examination involves evaluating how the forthcoming legal documents address these principles in the extent of their relation to the right to equality and non-discrimination, as well as whether they could deliver satisfactory legal protection. Additionally, the third chapter continues to build on the existing human rights law, especially the CoE Draft Framework Convention on AI is steering in this regard as it aims to adapt the existing human rights law framework to artificial intelligence.

However, although the CoE Draft Framework Convention on AI includes essential provisions, more is needed to protect the right to equality besides other human rights. For a legal framework to function, there needs to be more clarity in the provisions. On the other hand, as explained in the third chapter, the EU AI Act Proposal is righteously criticized for not providing sufficient human rights protection. While there are potentially effective measures in the EU AI Act Proposal, concerns such as benefiting from the economics of artificial intelligence are indeed reflected thereunder. It is difficult to say that the primary purpose of the EU AI Act Proposal is to protect human rights in artificial intelligence. In this case, the fact that both the EU and the CoE refer to each other in their initial studies, along with the CoE's acceptance of the EU becoming a party to the Draft Framework Convention on AI, the members of the EU being also members of the CoE may imply that it may be necessary to turn to the CoE for the human rights dimensions regarding artificial intelligence. As the CoE Draft Framework Convention on AI further accepts members outside the CoE as signatories to the Convention, expecting that it could have more impact worldwide.

Furthermore, the principles such as transparency, accountability, and safety and security may come up as human rights obligations in realizing the right to equality and

non-discrimination in artificial intelligence. Including these principles in the CoE Draft Framework Convention on AI implies that. Likewise, fulfilling the obligations to be brought in terms of the EU AI Act Proposal, although not in the form of human rights rules, will still be beneficial towards protecting human rights. The principles for artificial intelligence are required to ensure the right to equality and non-discrimination, which is more feasible with an specific framework for artificial intelligence. However, more than the current and forthcoming regulatory works is needed to protect the right to equality and non-discrimination in artificial intelligence. Equality and fairness find a place in the regulatory documents as separate artificial intelligence principles. However, more than the mere existence of these principles is needed to protect the right to equality and non-discrimination in artificial intelligence. It is possible to ensure the right to equality and non-discrimination when all other principles are ensured and observing them. This study supports the stance on establishing these principles for states as positive and procedural obligations for artificial intelligence and supporting them with human rights due diligence methods for private actors in terms of providing and protecting the right to equality and non-discrimination.

Lastly, in order to limit the study's scope, certain aspects have been excluded. For example, the relation of discrimination through artificial intelligence to personal data protection has also been excluded except for several mentions as it would require an entire focus to be given to that aspect, and more studies are already conducted on the matter. Similarly, national perspectives, besides the examples given in the study, are again excluded to keep the focus of the study as a matter of international human rights law. Numerous examples, perspectives on principles, and documents in the literature had to be excluded to keep the study's volume manageable. The potential benefits of artificial intelligence on the right to equality and non-discrimination are likewise excluded. Research with an even more limited scope could be conducted in future research, for example, regarding a principle, an approach, an organization, or a document discussed in the study. More research in such ways should be conducted in the field of artificial intelligence, as it is subject to new developments even on the day the last sentence of this thesis is written.

BIBLIOGRAPHY

1. Books

Alpaydın E, *Introduction to Machine Learning* (4th edn, MIT Press 2020)

Alpaydın E, *Machine Learning: The New AI* (Revised and updated edition, The MIT Press 2021)

Aristotle, *Aristotle: Nicomachean Ethics* (Roger Crisp ed, Roger Crisp tr, Cambridge University Press 2000)

Chesterman S, *We, the Robots?: Regulating Artificial Intelligence and the Limits of the Law* (Cambridge University Press 2021)

De Schutter O, *International Human Rights Law: Cases, Materials, Commentary* (Cambridge University Press 2011)

Deakin S and Markou C (eds), *Is Law Computable?: Critical Perspectives on Law and Artificial Intelligence* (1st edn, Hart Publishing 2020)

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

European Parliament Directorate General for Parliamentary Research Services, Castelluccia C and Métayer DL, *Understanding Algorithmic Decision-Making: Opportunities and Challenges* (Publications Office of the European Union 2019)

European Parliament Directorate General for Parliamentary Research Services, *The Impact of the General Data Protection Regulation on Artificial Intelligence* (Publications Office 2020)

European Union Agency for Fundamental Rights, *Preventing Unlawful Profiling Today and in the Future: A Guide* (Publications Office of the European Union 2018)

European Union Agency for Fundamental Rights, *Getting the Future Right: Artificial Intelligence and Fundamental Rights* (Publications Office of the European Union 2020)

European Union Agency for Fundamental Rights, *Bias in Algorithms: Artificial Intelligence and Discrimination.* (Publications Office of the European Union 2022)

European Union Agency for Fundamental Rights., European Court of Human Rights., and Council of Europe (Strasbourg)., *Handbook on European Non-*

Discrimination Law: 2018 Edition (Publications Office of the European Union 2018)

Fredman S, *Discrimination Law* (2nd edn, Oxford University Press, Incorporated 2012)

Han J, Pei J and Tong H, *Data Mining: Concepts and Techniques* (Morgan Kaufmann 2022)

Harris DJ and others, *Law of the European Convention on Human Rights* (Oxford University Press 2009)

IEEE, *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems* (First Edition, IEEE 2019)

Janis MW, Kay RS and Bradley AW, *European Human Rights Law: Text and Materials* (Oxford University Press 2008)

Norvig P and Russell S, *Artificial Intelligence: A Modern Approach, Global Edition* (4th edition, Pearson 2021)

O'Neil C, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1st edition, Crown 2016)

Orwat C, *Risks of Discrimination through the Use of Algorithms* (Federal Anti-Discrimination Agency 2020)

Pasquale F, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015)

Say C, *50 Soruda Yapay Zeka* (7th edn, Bilim ve Gelecek Kitaplığı 2018)

Smith N, *Basic Equality and Discrimination: Reconciling Theory and Law* (Ashgate Pub 2011)

Zuiderveen Borgesius FJ, *Discrimination, Artificial Intelligence, and Algorithmic Decision-Making* (Directorate General of Democracy, Council of Europe 2018)

2. Book Chapters

Ala-Pietilä P and Smuha NA, 'A Framework for Global Cooperation on Artificial Intelligence and Its Governance' in Bertrand Braunschweig and Malik Ghallab (eds), *Reflections on Artificial Intelligence for Humanity* (Springer International Publishing 2021)

Boddington P, 'Normative Modes: Codes and Standards' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (Oxford University Press 2020)

Bryson JJ and Theodorou A, 'How Society Can Maintain Human-Centric Artificial Intelligence' in Marja Toivonen and Eveliina Saari (eds), *Human-Centered Digitalization and Services* (Springer 2019)

Calders T and Žliobaitė I, 'Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures' in Bart Custers and others (eds), *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases* (Springer 2013)

Clayton R, Tomlinson H and George C, 'Freedom From Discrimination in Relation to Convention Rights', *The Law of Human Rights*, vol 1 (1st edn, Oxford University Press 2000)

Clifford J, 'Equality' in Dinah Shelton (ed), *The Oxford Handbook of International Human Rights Law* (Oxford University Press 2013)

Collins H and Khaitan T, 'Indirect Discrimination Law: Controversies and Critical Questions' in Hugh Collins and Tarunabh Khaitan (eds), *Foundations of Indirect Discrimination Law* (1st edn, Hart Publishing 2018)

Cruz-Roa AA and others, 'A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection' in Camille Salinesi, Moira C Norrie and Óscar Pastor (eds), *Advanced Information Systems Engineering*, vol 7908 (Springer Berlin Heidelberg 2013)

de Kleijn R, 'Artificial Intelligence Versus Biological Intelligence: A Historical Overview' in Bart Custers and Eduard Fosch-Villaronga (eds), *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice* (TMC Asser Press 2022)

Dervanović D, 'I, Inhuman Lawyer: Developing Artificial Intelligence in the Legal Profession' in Marcelo Corrales, Mark Fenwick and Nikolaus Forgó (eds), *Robotics, AI and the Future of Law* (Springer 2018)

European Commission for the Efficiency of Justice (CEPEJ), 'Which Uses of AI in European Judicial Systems?', *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (Council of Europe 2018)

Fredman S, 'Direct and Indirect Discrimination: Is There Still a Divide?' in Hugh Collins and Tarunabh Khaitan (eds), *Foundations of Indirect Discrimination Law* (1st edn, Hart Publishing 2018)

Gasser U and Schmitt C, 'The Role of Professional Norms in the Governance of Artificial Intelligence' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (Oxford University Press 2020)

Gosepath S, 'Equality' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Summer 2021, Metaphysics Research Lab, Stanford University 2021)

Havelková B, 'Judicial Scepticism of Discrimination at the ECtHR' in Hugh Collins and Tarunabh Khaitan (eds), *Foundations of Indirect Discrimination Law* (1st edn, Hart Publishing 2018)

Ienca M and Vayena E, 'AI Ethics Guidelines: European and Global Perspectives' in The CAHAI Secretariat, *Towards Regulation of AI Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law* (Council of Europe 2020)

Kazim E and others, 'Automation and Fairness' in Charles Kerrigan, *Artificial Intelligence: Law and Regulation* (Edward Elgar Publishing 2022)

Le Bui M and Noble SU, 'We're Missing a Moral Framework of Justice in Artificial Intelligence: On the Limits, Failings, and Ethics of Fairness' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (Oxford University Press 2020)

Liu C, Tan Z and He M, 'Overview of Artificial Intelligence in Medicine' in Manda Raz, Tam C Nguyen and Erwin Loh (eds), *Artificial Intelligence in Medicine: Applications, Limitations and Future Directions* (Springer Nature Singapore 2022)

Mantelero A, 'Analysis of International Legally Binding Instruments. Final Report' in The CAHAI Secretariat, *Towards Regulation of AI Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law* (Council of Europe 2020)

Moeckli D, 'Equality and Non-Discrimination' in Daniel Moeckli and others (eds), *International Human Rights Law* (2nd edition, Oxford University Press 2014)

Muller C and The CAHAI Secretariat, 'The Impact of AI on Human Rights, Democracy and the Rule of Law', *Towards Regulation of AI Systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law* (Council of Europe 2020)

Rab S, 'Telecoms and Connectivity' in Charles Kerrigan, *Artificial Intelligence: Law and Regulation* (Edward Elgar Publishing 2022)

Réaume D, 'Dignity, Equality, and Comparison' in Deborah Hellman and Sophia Moreau (eds), *Philosophical Foundations of Discrimination Law* (Oxford University Press 2013)

Ronsin X and Lampos V, 'In-Depth Study on the Use of AI in Judicial Systems, Notably AI Applications Processing Judicial Decisions and Data' in European Commission for the Efficiency of Justice (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment* (Council of Europe 2018)

Shaw P and Kerrigan C, 'Ethics', *Artificial Intelligence: Law and Regulation* (Edward Elgar Publishing 2022)

Shellshear E, Tremeer M and Bean C, 'Machine Learning, Deep Learning and Neural Networks' in Manda Raz, Tam C Nguyen and Erwin Loh (eds), *Artificial Intelligence in Medicine: Applications, Limitations and Future Directions* (Springer Nature 2022)

Tanna M and Dunning W, 'Bias and Discrimination' in Charles Kerrigan (ed), *Artificial Intelligence: Law and Regulation* (Edward Elgar Publishing 2022)

van Otterlo M, 'A Machine Learning View on Profiling' in Mireille Hildebrandt and Katja de Vries, *Privacy, Due Process and the Computational Turn* (1st edn, Routledge 2013)

Wagner B, 'Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?' in Emre Bayamlioglu and others (eds), *Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen* (Amsterdam University Press 2018)

Xenidis R and Senden L, 'EU Non-Discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination' in Ulf Bernitz and others (eds), *General Principles of EU Law and the EU Digital Order* (Kluwer Law International 2020)

Yeung K and Weller A, 'How Is "Transparency" Understood by Legal Scholars and the Machine Learning Community?' in Mediarep.Org and others, *Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen* (Amsterdam University Press 2018)

Yeung K, Howes A and Pogrebna G, 'AI Governance by Human Rights–Centered Design, Deliberation, and Oversight: An End to Ethics Washing' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (Oxford University Press 2020)

3. Articles, Papers and Preprints

'International AI Ethics Panel Must Be Independent' (2019) 572 Nature 415

'State v. Loomis - Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing' (2017) 130 Harvard Law Review 1530

Aivodji U and others, 'Fairwashing: The Risk of Rationalization', *Proceedings of the 36th International Conference on Machine Learning* (PMLR 2019) <https://proceedings.mlr.press/v97/aivodji19a.html>

Aletras N and others, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 PeerJ Computer Science 1

Ashraf C, 'Exploring the Impacts of Artificial Intelligence on Freedom of Religion or Belief Online' (2022) 26 The International Journal of Human Rights 757

Baeza-Yates R, 'Bias on the Web' (2018) 61 Communications of the ACM 54 <https://dl.acm.org/doi/10.1145/3209581> accessed 15 August 2023

Barnard C and Hepple B, 'Substantive Equality' (2000) 59 The Cambridge Law Journal 562

Barocas S and Selbst AD, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671

Bietti E, 'From Ethics Washing to Ethics Bashing: A View on Tech Ethics from within Moral Philosophy', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2020) <https://dl.acm.org/doi/10.1145/3351095.3372860>

Binns R, 'Fairness in Machine Learning: Lessons from Political Philosophy', *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (PMLR 2018) <https://proceedings.mlr.press/v81/binns18a.html> accessed 10 May 2023

Binns R, 'On the Apparent Conflict between Individual and Group Fairness', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (ACM 2020) <https://dl.acm.org/doi/10.1145/3351095.3372864> accessed 10 May 2023

Bohr A and Memarzadeh K, 'The Rise of Artificial Intelligence in Healthcare Applications' [2020] Artificial Intelligence in Healthcare 25

Borji A, 'A Categorical Archive of ChatGPT Failures' 1

Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', *Conference on fairness, accountability and transparency* (PMLR 2018)

Calders T, 'Machine-Learning Discrimination: Bias in, Bias Out', *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)* (2019) <https://ieeexplore.ieee.org/document/9014827>

Çalışkan A, Bryson JJ and Narayanan A, 'Semantics Derived Automatically from Language Corpora Contain Human-like Biases - Supplementary Material' (2017) 356 Science 183

Chui M and others, 'Notes from the AI Frontier: Insights from Hundreds of Use Cases' [2018] McKinsey Global Institute 32

Cofone IN, 'Algorithmic Discrimination Is an Information Problem' (2019) 70 HASTINGS LAW JOURNAL 1389

Creel K and Hellman D, 'The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems' (2022) 52 Canadian Journal of Philosophy 26

Datta A, Tschantz MC and Datta A, 'Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination' (2015) 2015 Proceedings on Privacy Enhancing Technologies 92

De Boer RJ, Heems W and Hurts K, 'The Duration of Automation Bias in a Realistic Setting' (2014) 24 The International Journal of Aviation Psychology 287

Dencik L and Sanchez-Monedero J, 'Data Justice' (2022) 11 Internet Policy Review <https://policyreview.info/articles/analysis/data-justice> accessed 23 July 2023

Dwork C and others, 'Fairness Through Awareness' 1

Ebers M, 'Liability for Artificial Intelligence and EU Consumer Law' (2021) 12 Journal of Intellectual Property, Information Technology and Electronic Commerce Law 204

Gaynor M, 'Automation and AI Sound Similar, but May Have Vastly Different Impacts on the Future of Work' (*Brookings*) <https://www.brookings.edu/articles/automation-and-artificial-intelligence-sound-similar-but-may-have-vastly-different-impacts-on-the-future-of-work/>

Greene D, Hoffmann AL and Stark L, 'Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning' [2019] Hawaii International Conference on System Sciences 2019 (HICSS-52) 2122

Gunning D and Aha D, 'DARPA's Explainable Artificial Intelligence (XAI) Program' (2019) 40 AI Magazine 44

Hacker P, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination under EU Law' (2018) 55 Common Market Law Review <https://kluwerlawonline.com/journalarticle/Common+Market+Law+Review/55.4/COLA2018095> accessed 9 November 2021

Hagendorff T, 'The Ethics of AI Ethics -- An Evaluation of Guidelines' (2020) 30 Minds and Machines 99

Hassani BK, 'Societal Bias Reinforcement through Machine Learning: A Credit Scoring Perspective' (2021) 1 AI and Ethics 239 <https://doi.org/10.1007/s43681-020-00026-z> accessed 15 August 2023

Hellman D, 'Measuring Algorithmic Fairness' (2020) 106 Virginia Law Review 811

Hoffmann AL, 'Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse' (2019) 22 Information, Communication & Society 900

Hooker S, 'Moving beyond "Algorithmic Bias Is a Data Problem"' (2021) 2 Patterns (New York, N.Y.) 100241

Jobin A, Ienca M and Vayena E, 'The Global Landscape of AI Ethics Guidelines' (2019) 1 Nature Machine Intelligence 389

Joseph M and others, 'Fair Algorithms for Infinite and Contextual Bandits' 1

Joseph M and others, 'Rawlsian Fairness for Machine Learning' 1

Kadar C, Maculan R and Feuerriegel S, 'Public Decision Support for Low Population Density Areas: An Imbalance-Aware Hyper-Ensemble for Spatio-Temporal Crime Prediction' (2019) 119 Decision Support Systems 107

Kaplan A and Haenlein M, 'Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence' (2019) 62 Business Horizons 15

Katz DM, Bommarito MJ and Blackman J, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States' <https://papers.ssrn.com/abstract=2463244>

Kroll JA and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 74

Liu Y and others, 'Calibrated Fairness in Bandits' 1

Lane L, 'The Horizontal Effect of International Human Rights Law in Practice: A Comparative Analysis of the General Comments and Jurisprudence of Selected United Nations Human Rights Treaty Monitoring Bodies' (2018) 5 European Journal of Comparative Law and Governance 5 <https://brill.com/view/journals/ejcl/5/1/article-p5_5.xml>

Lowry S and Macpherson G, 'A Blot on the Profession' (1988) 296 British Medical Journal (Clinical research ed.) 657

Lum K and Isaac W, 'To Predict and Serve?' (2016) 13 Significance 14

Makridakis S, 'The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms' (2017) 90 Futures 46

Mayson SG, 'Bias In, Bias Out' (2019) 128 The Yale Law Journal 2218

McGregor L, ‘Accountability for Governance Choices in Artificial Intelligence: Afterword to Eyal Benvenisti’s Foreword’ (2019) 29 European Journal of International Law 1079

McGregor L, Murray D and Ng V, ‘International Human Rights Law as a Framework for Algorithmic Accountability’ (2019) 68 International and Comparative Law Quarterly 309

Mehrabi N and others, ‘A Survey on Bias and Fairness in Machine Learning’ (2021) 54 ACM Computing Surveys 115:1

Miller DD and Brown EW, ‘Artificial Intelligence in Medical Practice: The Question to the Answer?’ (2018) 131 The American Journal of Medicine 129

Mittelstadt BD and others, ‘The Ethics of Algorithms: Mapping the Debate’ (2016) 3 Big Data & Society 1

Nemitz P, ‘Constitutional Democracy and Technology in the Age of Artificial Intelligence’ (2018) 376 Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences 20180089

Parasuraman R and Manzey DH, ‘Complacency and Bias in Human Use of Automation: An Attentional Integration’ (2010) 52 Human Factors 381

Pasquale F and Citron DK, ‘The Scored Society: Due Process for Automated Predictions’ (2014) 89 Washington Law Review 1

Rawls J, ‘Justice as Fairness’ (1958) 67 The Philosophical Review 164

Reiling AD (Dory), ‘Courts and Artificial Intelligence’ (2020) 11 International Journal for Court Administration 8

Risse M, ‘Human Rights and Artificial Intelligence: An Urgently Needed Agenda’ (2019) 41 Human Rights Quarterly 1

Saxena N and others, ‘How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness’ <http://arxiv.org/abs/1811.03654> accessed 6 October 2022

Schönberger D, ‘Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications’ (2019) 27 International Journal of Law and Information Technology 171

Sert MF, Yıldırım E and Haşlak İ, 'Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court' (2022) 40 Social Science Computer Review 1416

Silva S and Kenney M, 'Algorithms, Platforms, and Ethnic Bias: An Integrative Essay' (2018) 55 Phylon (1960-) 9

Smuha NA, 'Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea' (2021) 34 Philosophy & Technology 91

Smuha NA, 'How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act' 1

Speicher T and others, 'Potential for Discrimination in Online Targeted Advertising', *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (PMLR 2018) <https://proceedings.mlr.press/v81/speicher18a.html> accessed 23 April 2023

Strauss DA, 'The Illusory Distinction Between Equality of Opportunity and Equality of Result' (1992) 34 William and Mary Law Review 171

Suresh H and Guttag JV, 'A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle', *Equity and Access in Algorithms, Mechanisms, and Optimization* (2021) <http://arxiv.org/abs/1901.10002> accessed 4 October 2022

Sweeney L, 'Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising' (2013) 11 Queue 10

Taylor L, 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally' (2017) 4 Big Data & Society 2053951717736335 <https://doi.org/10.1177/2053951717736335>

Tolan S, 'Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges' [2019] arXiv preprint arXiv:1901.04730

Turing AM, 'Computing Machinery and Intelligence' (1950) LIX Mind 433

van Maanen G, 'AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics' (2022) 1 Digital Society 9

van Nuenen T and others, 'Transparency for Whom? Assessing Discriminatory Artificial Intelligence' (2020) 53 Computer 36

Veale M and Binns R, 'Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data' (2017) 4 Big Data & Society 1

Verma S and Rubin J, 'Fairness Definitions Explained', *Proceedings of the International Workshop on Software Fairness* (Association for Computing Machinery 2018) <https://doi.org/10.1145/3194770.3194776> accessed 3 August 2022

Wachter S, 'Affinity Profiling and Discrimination By Association in Online Behavioral Advertising' (2020) 35 Berkeley Technology Law Journal 368

Wachter S, Mittelstadt B and Russell C, 'Why Fairness Cannot Be Automated: Bridging the Gap between EU Non-Discrimination Law and AI' (2021) 41 Computer Law & Security Review 105567

Westen P, 'The Empty Idea of Equality' (1982) 95 Harvard Law Review 537

Woodruff A and others, 'A Qualitative Exploration of Perceptions of Algorithmic Fairness', *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (ACM 2018) <https://dl.acm.org/doi/10.1145/3173574.3174230> accessed 19 June 2022

Wu X and Zhang X, 'Automated Inference on Criminality Using Face Images' <https://arxiv.org/pdf/1611.04135v2.pdf> accessed 21 February 2023

Wu X and Zhang X, 'Responses to Critiques on Machine Learning of Criminality Perceptions (Addendum of ArXiv:1611.04135)' 1

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

Xiang A and Raji ID, 'On the Legal Compatibility of Fairness Definitions' [2019] 1

Yeung K, 'Recommendation of the Council on Artificial Intelligence (OECD)' (2020) 59 International Legal Materials 27

Zalnieriute M, Moses LB and Williams G, 'The Rule of Law and Automation of Government Decision-Making' (2019) 82 The Modern Law Review 425

Završnik A, 'Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings' (2021) 18 European Journal of Criminology 623

Završnik A, 'Criminal Justice, Artificial Intelligence Systems, and Human Rights' (2020) 20 ERA Forum 567

Zhang T and others, 'Fairness in Semi-Supervised Learning: Unlabeled Data Help to Reduce Discrimination' (2022) 34 IEEE Transactions on Knowledge and Data Engineering 1763

Žliobaitė I, 'Measuring Discrimination in Algorithmic Decision Making' (2017) 31 Data Mining and Knowledge Discovery 1060

Zucker I and Prendergast BJ, 'Sex Differences in Pharmacokinetics Predict Adverse Drug Reactions in Women' (2020) 11 Biology of Sex Differences 32

Zuiderveen Borgesius FJ, 'Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence' (2020) 24 The International Journal of Human Rights 1572

4. Online Resources

'AI Act: A Step Closer to the First Rules on Artificial Intelligence | News | European Parliament' (11 May 2023) <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> accessed 11 May 2023

American Academy of Dermatology Association, 'Melanoma: Who Gets and Causes' <https://www.aad.org/public/diseases/skin-cancer/types/common/melanoma/causes> accessed 3 December 2022

Amos Toh, 'Dutch Ruling a Victory for Rights of the Poor' (*Human Rights Watch*, 6 February 2020) <https://www.hrw.org/news/2020/02/06/dutch-ruling-victory-rights-poor>

Apple Support, 'Get the Most Accurate Measurements Using Your Apple Watch' (*Apple Support*, 20 September 2022) <https://support.apple.com/en-gb/HT207941> accessed 2 January 2023

'Artificial Intelligence – Ethical and Legal Requirements: Feedback and Statistics: Proposal for a Regulation' (*European Commission*, 6 August 2021) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback_en?p_id=24212003> accessed 20 May 2023

Asimov I, 'Runaround' (1942) <https://web.williams.edu/Mathematics/sjmiller/public_html/105Sp10/handouts/Runaround.html> accessed 12 April 2023

Australian Government Department of Industry, Science and Resources, 'Australia's AI Ethics Principles | Australia's Artificial Intelligence Ethics Framework | Department of Industry, Science and Resources' (*https://www.industry.gov.au/node/91877*, 5 October 2022) <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles> accessed 13 April 2023

Balasubramanian R, Libarikian A and McElhaney D, 'Insurance 2030—The Impact of AI on the Future of Insurance | McKinsey' (*McKinsey & Company*, 12 March 2021) <https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance> accessed 8 June 2023

Bellan R, 'Microsoft Lays off an Ethical AI Team as It Doubles down on OpenAI | TechCrunch' (14 March 2023) <https://techcrunch.com/2023/03/13/microsoft-lays-off-an-ethical-ai-team-as-it-doubles-down-on-openai/> accessed 15 March 2023

'ChatGPT' (*ChatGPT*) <https://chat.openai.com/chat> accessed 14 February 2023

'DeepMind's Health Team Joins Google Health' <https://www.deepmind.com/blog/deepminds-health-team-joins-google-health> accessed 20 December 2022

Diakopoulos N and others, 'Principles for Accountable Algorithms and a Social Impact Statement for Algorithms :: FAT ML' <https://www.fatml.org/resources/principles-for-accountable-algorithms>

European Commission, 'Non-Discrimination - What to Do If Your Rights Have Been Breached' (*European Commission - European Commission*) <https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/know-your-rights/equality/non-discrimination_en> accessed 11 September 2022

European Parliament, 'Proposal for a Regulation on a European Approach for Artificial Intelligence | Legislative Train Schedule' (*European Parliament*, 20 April 2023) <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence> accessed 26 May 2023

‘Get Gender-Specific Translations - Google Translate Help’ <https://support.google.com/translate/answer/9179237?hl=en&ref_topic=7011755&sjid=14977587573544606156-EU> accessed 7 April 2023

Google Developers, ‘Framing: Key ML Terminology | Machine Learning’ (*Machine Learning Foundational Courses Crash Course*, 18 July 2022) <https://developers.google.com/machine-learning/crash-course/framing/ml-terminology> accessed 10 August 2022

IBM, ‘IBM Global AI Adoption Index 2022’ (2022) <https://www.ibm.com/downloads/cas/GVAGA3JP>

IBM, ‘AI Fairness 360 - Resources’ (*IBM Research Trusted AI*) <https://aif360.mybluemix.net/aif360.mybluemix.net/resources> accessed 13 August 2022

International Society of Automation, ‘What Is Automation? - ISA’ (*isa.org*, 2022) <https://www.isa.org/about-isa/what-is-automation> accessed 25 July 2022

Jordan M, ‘Artificial Intelligence — The Revolution Hasn’t Happened Yet’ (*Medium*, 30 April 2018) <https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7> accessed 12 August 2022

Koebler J and Cox J, ‘The Impossible Job: Inside Facebook’s Struggle to Moderate Two Billion People’ (*Vice*, 23 August 2018) <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works> accessed 23 April 2023

Kroll J, ‘Accountable Algorithms (A Provocation)’ (*Media@LSE*, 10 February 2016) <https://blogs.lse.ac.uk/medialse/2016/02/10/accountable-algorithms-a-provocation/> accessed 24 May 2023

Lee O, ‘Camera Misses the Mark on Racial Sensitivity’ (*Gizmodo*, 15 May 2009) <https://gizmodo.com/camera-misses-the-mark-on-racial-sensitivity-5256650> accessed 8 April 2023

Lee P, ‘Learning from Tay’s Introduction’ (*The Official Microsoft Blog*, 25 March 2016) <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/> accessed 16 February 2023

Lipton AZC, ‘From AI to ML to AI: On Swirling Nomenclature & Slurried Thought’ (*Approximately Correct*, 5 June 2018)

<https://www.approximatelycorrect.com/2018/06/05/ai-ml-ai-swirling-nomenclature-slurried-thought/> accessed 12 August 2022

Mauro G and Schellmann H, '"There Is No Standard": Investigation Finds AI Algorithms Objectify Women's Bodies | Artificial Intelligence (AI) | The Guardian' (8 February 2023) <https://www.theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies> accessed 13 February 2023

Microsoft Bing, 'Introducing the New Bing' <https://www.bing.com/new> accessed 16 February 2023

'National Fair Housing Alliance et al. v. Facebook | Civil Rights Litigation Clearinghouse' <https://clearinghouse.net/case/17155/> accessed 21 March 2023

New Zealand Government, 'Algorithm Charter For Aotearoa New Zealand' (2020) <https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf> accessed 13 April 2023

OECD.AI (2021), 'Powered by EC/OECD (2021), Database of National AI Policies' <https://oecd.ai/en/dashboards/overview> accessed 6 June 2023

Pasternack A, 'GPT-Powered Deepfakes Are a "Powder Keg"' (*Fast Company*, 22 February 2023) <https://www.fastcompany.com/90853542/deepfakes-getting-smarter-thanks-to-gpt> accessed 9 June 2023

Peck D, 'They're Watching You at Work' (*The Atlantic*, 21 November 2013) <https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/> accessed 8 March 2023

'Responsible AI Principles from Microsoft' (*Microsoft*) <https://www.microsoft.com/en-us/ai/responsible-ai> accessed 10 June 2023

Sankaran V, 'Apple Sued Over "Ineffectiveness" of Apple Watch Blood Oxygen Reader for People of Colour' (*The Independent*, 29 December 2022) <https://www.independent.co.uk/tech/apple-watch-sued-oxygen-reader-bias-b2252807.html> accessed 2 January 2023

'Settlement Agreement | National Fair Housing Alliance Et Al. V. Facebook | Civil Rights Litigation Clearinghouse' (18 March 2019) <https://clearinghouse.net/doc/102171/> accessed 21 March 2023

Sharp G, 'Nikon Camera Says Asians: People Are Always Blinking - Sociological Images' (29 May 2009)

<https://thesocietypages.org/socimages/2009/05/29/nikon-camera-says-asians-are-always-blinking/> accessed 8 April 2023

T.C. Sağlık Bakanlığı, 'NeyimVar' (*NeyimVar*) <https://neyimvar.gov.tr/giris> accessed 31 July 2022

T.C. Sağlık Bakanlığı, 'NeyimVar Privacy, Use and Copyrights' (*NeyimVar*) <https://neyimvar.gov.tr/tanitim> accessed 29 November 2022

The PAIR Team Google, 'Know Your Data' <https://knowyourdata.withgoogle.com> accessed 26 July 2022

The PAIR Team Google, 'Know Your Data Documentation' <https://knowyourdata.withgoogle.com/docs/> accessed 26 July 2022

Treaty Office of Council of Europe, 'Chart of Signatures and Ratifications of Treaty 177' (*Treaty Office - Conventions*) <https://www.coe.int/en/web/conventions/full-list> accessed 14 August 2022

U.S. Department of Justice U.S. Attorney's Office Southern District of New York, 'United States Attorney Resolves Groundbreaking Suit Against Meta Platforms, Inc., Formerly Known As Facebook, To Address Discriminatory Advertising For Housing' (21 June 2022) <https://www.justice.gov/usao-sdny/pr/united-states-attorney-resolves-groundbreaking-suit-against-meta-platforms-inc-formerly> accessed 21 March 2023

Varshney KR, 'Introducing AI Fairness 360, A Step Towards Trusted AI - IBM Research' (*IBM Research Blog*, 19 September 2018) <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/> accessed 13 August 2022

Varshney KR, 'Introducing AI Fairness 360, A Step Towards Trusted AI - IBM Research' (*IBM Research Blog*, 19 September 2018) <https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/> accessed 25 July 2022

Vincent J, 'Google "Fixed" Its Racist Algorithm by Removing Gorillas from Its Image-Labeling Tech' (*The Verge*, 12 January 2018) <https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai> accessed 6 April 2023

Wagner K and Dolmetsch C, 'Meta Settles Claims That Ads Violated U.S. Fair Housing Laws' (*Time*, 21 June 2022) <https://time.com/6189595/meta-settles-ads-fair-housing-laws/> accessed 21 March 2023

WHO, 'WHO Calls for Safe and Ethical AI for Health' <https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health> accessed 8 June 2023

Woollacott E, 'Apple Sued Over "Racial Bias" of Apple Watch' (*Forbes*, 29 December 2022) <https://www.forbes.com/sites/emmawoollacott/2022/12/29/apple-sued-over-racial-bias-of-apple-watch/> accessed 2 January 2023

5. Legal Cases

Abdu v Bulgaria [2014] European Court of Human Rights App. No. 26827/08

Abdulaziz, Cabales and Balkandali v The United Kingdom [1985] European Court of Human Rights App. Nos. 9214/80; 9473/81; 9474/81, A94

Aktaş v Turkey [2003] European Court of Human Rights App. No. 24351/94, Rep Judgm Decis 2003-V Extr

Anguelova v Bulgaria [2002] European Court of Human Rights App. No. 38361/97, Rep Judgm Decis 2002-IV

Assessment of creditworthiness, authority, direct multiple discrimination, gender, language, age, place of residence, financial reasons, conditional fine [2018] National Non-Discrimination and Equality Tribunal of Finland 216/2017

Biao v Denmark [2016] European Court of Human Rights App. No. 38590/10, Rep Judgm Decis 2016

Big Brother Watch and Others v the United Kingdom [2021] European Court of Human Rights [GC] App. Nos. 58170/13, 62322/14 and 24960/15

BS v Spain [2012] European Court of Human Rights App. No. 47159/08

Buckley v The United Kingdom [1996] European Court of Human Rights App. No. 20348/92, Rep 1996-IV

Burden v The United Kingdom [2008] European Court of Human Rights App. No. 13378/05, Rep Judgm Decis 2008

Carson and Others v the United Kingdom [2010] European Court of Human Rights App. No. 42184/05, 2010–II Eur Court Hum Rights Rep Judgm Decis 407

Case 'Relating to Certain Aspects of the Laws on the Use of Languages in Education in Belgium' v Belgium "Belgian Linguistics Case" [1968] European Court of Human Rights App. Nos. 1474/62 1677/62 1691/62 1769/63 1994/63 2126/64, 1 EHRR Eur Court Hum Rights Rep Judgm Decis 252

Centrum för rättvisa v Sweden [2021] European Court of Human Rights [GC] App. No. 35252/08

Chassagnou and Others v France [1999] European Court of Human Rights App. Nos. 25088/94, 28331/95 and 28443/95, Rep Judgm Decis 1999-III

Danilenkov and Others v Russia [2009] European Court of Human Rights App. No. 67336/01, Rep Judgm Decis 2009 Extr

Delfi AS v Estonia [2015] European Court of Human Rights [GC] App. No. 64569/09

DH and Others v The Czech Republic [2007] European Court of Human Rights App. No. 57325/00, 2007–IV Eur Court Hum Rights Rep Judgm Decis 241

Di Trizio v Switzerland [2016] European Court of Human Rights App. No. 7186/09

EB v France [2008] European Court of Human Rights App. No. 43546/02, Eur Court Hum Rights Rep Judgm Decis

ECLI:NL:RBDHA:2020:1878, Rechtbank Den Haag, C-09-550982-HA ZA 18-388 (English) [2020] Rb Den Haag ECLI:NL:RBDHA:2020:1878

Engel and Others v The Netherlands [1976] European Court of Human Rights App. Nos. 5100/71; 5101/71; 5102/71; 5354/72; 5370/72, 1 Eur Court Hum Rights Rep Judgm Decis 647

Fábián v Hungary [2017] European Court of Human Rights App. No. 78117/13, Rep Judgm Decis 2017 Extr

Fabris v France [2013] European Court of Human Rights App. No. 16574/08, Reports of Judgments and Decisions 2013 Eur Court Hum Rights Rep Judgm Decis

FH Zwaan-de Vries v Netherlands [1987] CCPR Communication No. 182/1984, UN Doc CCPRC29D1821984

Hoogendijk v The Netherlands [2005] European Court of Human Rights App. No. 58641/00, Eur Court Hum Rights Rep Judgm Decis

Hugh Jordan v The United Kingdom [2001] European Court of Human Rights App. No. 24746/94, Eur Court Hum Rights Rep Judgm Decis

Khamtokhu and Aksenchik v Russia [2017] European Court of Human Rights App. nos. 60367/08 and 961/11, Rep Judgm Decis 2017

Kjeldsen, Busk Madsen and Pedersen v Denmark [1976] European Court of Human Rights App. Nos. 5095/71; 5920/72; 5926/72), 1 Eur Court Hum Rights Rep Judgm Decis 711

Konstantin Markin v Russia [2012] European Court of Human Rights App. no. 30078/06, Rep Judgm Decis 2012 Extr

Lithgow and Others v The United Kingdom [1986] European Court of Human Rights App. nos. 9006/80; 9262/81; 9263/81; 9265/81; 9266/81; 9313/81; 9405/81, A102

Magyar Kétfarkú Kutya Párt v Hungary [2020] European Court of Human Rights [GC] App. No. 201/17

Marckx v Belgium [1979] European Court of Human Rights App. No. 6833/74, A31 Eur Court Hum Rights Rep Judgm Decis

Molla Sali v Greece [2018] European Court of Human Rights App. No. 20452/14

Nachova v Russia [2005] European Court of Human Rights App. Nos. 543577/98 and 43579/98, Rep Judgm Decis 2005-VII

National Fair Housing Alliance; Fair Housing Justice Center, Inc; Housing Opportunities Project For Excellence, Inc; Fair Housing Council Of Greater San Antonio v Facebook, Inc [2019] United States District Court Southern District Of New York 1:18-cv-02689

Opuz v Turkey [2009] European Court of Human Rights App. No. 33401/02, Rep Judgm Decis 2009

Oršuš and Others v Croatıa [2010] European Court of Human Rights App. No.15766/03, Rep Judgm Decis 2010

Petrovic v Austria [1998] European Court of Human Rights App. No. 20458/92, Rep 1998-II

Pine Valley Developments Ltd and Others v Ireland [1991] European Court of Human Rights App. No. 12742/87, A222 Eur Court Hum Rights Rep Judgm Decis

Pla and Puncernau v Andorra [2004] European Court of Human Rights App. No. 69498/01, Rep Judgm Decis 2004-VIII

Rasmussen v Denmark [1984] European Court of Human Rights App. No. 8777/79, 17 ECHR

Salman v Turkey [2000] European Court of Human Rights App. No. 21986/93, Rep Judgm Decis 2000-VII

SAS v France [2014] European Court of Human Rights App. No. 43835/11, Rep Judgm Decis 2014

Sigurður Einarsson and Others v Iceland [2019] European Court of Human Rights App. No. 39757/15

State v Loomis - 2016 WI 68, 371 Wis 2d 235, 881 NW2d 749

Stec and Others v The United Kingdom [2006] European Court of Human Rights App. Nos. 65731/01 and 65900/01, Rep Judgm Decis 2006-VI

Tapayeva and Others v Russia [2021] European Court of Human Rights App. No. 24757/18

Timishev v Russia [2005] European Court of Human Rights App. Nos. 55762/00 and 55974/00, Rep Judgm Decis 2005-XII

Tyrer v the United Kingdom [1978] European Court of Human Rights App. No. 5856/72, A26

Ünal Tekeli v Turkey [2004] European Court of Human Rights App. No. 29865/96, Rep Judgm Decis 2004-X Extr

Weller v Hungary [2009] European Court of Human Rights App. No. 44399/05

Willis v The United Kingdom [2002] European Court of Human Rights App. No. 36042/97, Rep Judgm Decis 2002-IV

Zarb Adami v Malta [2006] European Court of Human Rights App. No. 17209/02, Eur Court Hum Rights Rep Judgm Decis

6. Legal Documents

6.1 Treaties, Conventions and Regulations

Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11, 14 and 15, supplemented by Protocols Nos. 1, 4, 6, 7, 12, 13 and 16, E.T.S. No. 5 (Nov. 4, 1950) (entered into force Sept. 3, 1953).

Consolidated version of the Treaty on the Functioning of the European Union (TFEU) 2012 47

Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) 1985 (ETS No 108)

European Social Charter (ETS No. 35) 1961 (ETS No 35) (entered into force Feb. 36, 1975).

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 3 (ICESCR)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) 2016 1

The EU Charter of Fundamental Rights of the European Union 2000 391

6.2 The Forthcoming Legally Binding Documents

European Commission, Annexes to the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts 2021 [COM(2021) 206 final]

European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts 2021 [COM(2021) 206 final]

Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law 2023 [CAI(2023)01]

6.3 Other

Council of Europe Committee of Ministers, 'Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes' <https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b> accessed 4 August 2022

Council of Europe Committee of Ministers, 'Declaration by the Committee of Ministers on the Risks of Computer-Assisted or Artificial-Intelligence-Enabled Decision Making in the Field of the Social Safety Net' <https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a1cb98> accessed 4 August 2022

Council of Europe Committee of Ministers, 'Recommendation CM/Rec(2019)1 of the Committee of Ministers to Member States on Preventing and Combating Sexism' <https://search.coe.int/cm/pages/result_details.aspx?ObjectId=0900001680a217ca> accessed 4 August 2022

Council of Europe Directorate General of Human Rights and Rule of Law, 'Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108) - Guidelines on Artificial Intelligence and Data Protection' <https://rm.coe.int/human-rights-and-business-recommendation-cm-rec-2016-3-of-the-committe/16806f2032> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Recommendation 2185 (2020) of the Parliamentary Assembly on Artificial Intelligence in Health Care: Medical, Legal and Ethical Challenges Ahead' <https://pace.coe.int/pdf/04b0ea76003aa33e8c0008819521f70b2694252d33bafdf3f375d859d9e4058b/rec.%202185.pdf> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Resolution 2342 (2020) of the Parliamentary Assembly on Justice by Algorithm – The Role of Artificial Intelligence in Policing and Criminal Justice Systems' <https://pace.coe.int/pdf/b6c2cd92ba3c001b7cc14df75b50b1c3718bbfcc5f9c65438620eaf61d4cf648/res.%202342.pdf> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Recommendation 2102 (2017) of the Parliamentary Assembly on Technological Convergence, Artificial Intelligence and Human Rights' <https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Recommendation 2181 (2020) of the Parliamentary Assembly on the Need for Democratic Governance of Artificial Intelligence' <https://pace.coe.int/pdf/114e7524b1e69823f01fc2cb4080f5401933355f1c631e797e23f45fcf041f95/rec.%202181.pdf> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Recommendation 2182 (2020) of the Parliamentary Assembly on Justice by Algorithm – The Role of Artificial Intelligence in Policing and Criminal Justice Systems'

<https://pace.coe.int/pdf/bf427be8505b6833866638cc75e52970b11091824f077b5caaf5f581f21f61c2/rec.%202182.pdf> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Recommendation 2183 (2020) of the Parliamentary Assembly on Preventing Discrimination Caused by the Use of Artificial Intelligence' <https://pace.coe.int/pdf/1cf0f22ba28c172524b4e7261f78f2d45c508160b6c058ec2b2d2fadfca9c3bc/rec.%202183.pdf> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Recommendation 2186 (2020) of the Parliamentary Assembly on Artificial Intelligence and Labour Markets: Friend or Foe?' <https://pace.coe.int/pdf/9a966efa8fb3a95f25194ea42052817fdc1b50640493ed75defe1805d0212a19/rec.%202186.pdf> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Resolution 2341 (2020) of the Parliamentary Assembly on the Need for Democratic Governance of Artificial Intelligence' <https://pace.coe.int/en/files/28803/html> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Resolution 2343 (2020) of the Parliamentary Assembly on Preventing Discrimination Caused by the Use of Artificial Intelligence' <https://pace.coe.int/pdf/263ef53d02a37aaf864a1f7cf9a6427a0c3bb47fbedc7a706ec8bf4811d15244/res.%202343.pdf> accessed 4 August 2022

Council of Europe Parliamentary Assembly, 'Resolution 2345 (2020) of the Parliamentary Assembly on Artificial Intelligence and Labour Markets: Friend or Foe?' <https://pace.coe.int/pdf/2a6ef5d63825348a8311e7591d1990bffeae7075476b2aafce0e361424936534/res.%202345.pdf> accessed 4 August 2022

Council of Europe, 'Explanatory Report to the Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms' (Council of Europe 2000) <https://rm.coe.int/09000016800cce48>

Council of Europe, 'Recommendation CM/Rec(2016)3 of the Committee of Ministers to Member States on Human Rights and Business' <https://rm.coe.int/human-rights-and-business-recommendation-cm-rec-2016-3-of-the-committe/16806f2032> accessed 4 August 2022

Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators 2022

Ethics and governance of artificial intelligence for health: WHO guidance 2021

European Commission for the Efficiency of Justice (CEPEJ), 'Guidelines on Electronic Court Filing (e-Filing) and Digitalisation of Courts' (Council of Europe 2021) CEPEJ(2021)15 <https://rm.coe.int/cepej-2021-15-en-e-filing-guidelines-digitalisation-courts/1680a4cf87> accessed 20 May 2023

European Commission, 'Commission Staff Working Document Impact Assessment Accompanying the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts' (European Commission 2021) SWD(2021) 84 final <https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence>

European Commission, 'Communication from the Commission Artificial Intelligence for Europe' (European Commission 2018) COM(2018) 237 final <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN> accessed 9 August 2022

European Commission, 'Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions - Building Trust in Human-Centric Artificial Intelligence' 11 <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>

European Commission, 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust' (European Commission 2020) COM(2020) 65 final <https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf> accessed 6 August 2022

European Committee on Legal Co-operation (CDCJ) Guidelines of the Committee of Ministers of the Council of Europe on online dispute resolution mechanisms in civil and administrative court proceedings 2021

European Economic and Social Committee and Catelijne Muller, 'Opinion of the European Economic and Social Committee on AI/Regulation Proposal for a

Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/regulation-artificial-intelligence>

European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment 2018

High-Level Expert Group on AI (AI HLEG), 'Ethics Guidelines for Trustworthy AI' (European Commission 2019) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419>

OECD Recommendation of the Council on Artificial Intelligence 2019

Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems 2020

The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment 2020

UNESCO Recommendation on the Ethics of Artificial Intelligence 2021 21

UNESCO, 'Beijing Consensus on Artificial Intelligence and Education - UNESCO Digital Library' (UNESCO 2019) <https://unesdoc.unesco.org/ark:/48223/pf0000368303> accessed 23 April 2023

United Nations Committee on Economic, Social and Cultural Rights, 'General Comment No. 25 (2020) on Science and Economic, Social and Cultural Rights (Article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights)' 19 <https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=E%2FC.12%2FGC%2F25&Lang=en>

United Nations Committee on the Elimination of Racial Discrimination, 'General Recommendation No.36 . Preventing and Combating Racial Profiling by Law Enforcement Officials' 13 <https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CERD%2FC%2FGC%2F36&Lang=en>

United Nations Committee on the Rights of the Child, 'General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment' 20

<https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CRC%2FC%2FGC%2F25&Lang=en>

United Nations General Assembly, 'UN Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework' (2011) UN Doc A/HRC/17/31

United Nations Human Rights Committee, 'General Comment No. 18 - Non Discrimination' <https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=INT%2FCCPR%2FGEC%2F6622&Lang=en>

United Nations Human Rights Committee, 'General Comment No. 28 - Article 3 (The Equality of Rights between Men and Women)' (United Nations Human Rights Committee 2000) U.N. Doc. CCPR/C/21/Rev.1/Add.10 <https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CCPR%2FC%2F21%2FRev.1%2FAdd.10&Lang=en>

United Nations Human Rights Committee, 'General Comment No. 32 - Article 14: Right to Equality before Courts and Tribunals and to a Fair Trial' <https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/Download.aspx?symbolno=CCPR%2FC%2FGC%2F32&Lang=en>

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

7. Reports

'Improving Access to Remedy in the Area of Business and Human Rights at the EU Level - Opinion of the European Union Agency for Fundamental Rights' (European Union Agency for Fundamental Rights 2017) FRA Opinion – 1/2017 <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2017-opinion-01-2017-business-human-rights_en.pdf>

'Work for a Brighter Future – Global Commission on the Future of Work' (International Labour Organization 2019) <https://www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_662410.pdf>

Access Now and Andersen L, 'Human Rights in the Age of Artificial Intelligence' (Access Now 2018)

<https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>

Ad Hoc Committee on Artificial Intelligence (CAHAI), 'Possible Elements of a Legal Framework on Artificial Intelligence, Based on the Council of Europe's Standards on Human Rights, Democracy and the Rule of Law' (Council of Europe 2021) CAHAI(2021)09rev <https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d> accessed 4 August 2022

Ad Hoc Committee on Artificial Intellıgence (CAHAI), 'Ad Hoc Committee on Artificial Intellıgence (CAHAI) Feasibility Study' (Council of Europe 2020) CAHAI(2020)23 <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da> accessed 4 August 2022

AlgorithmWatch, 'Automating Society: Taking Stock of Automated Decision-Making in the EU' (AW AlgorithmWatch gGmbH & Bertelsmann Stiftung 2019) <https://algorithmwatch.org/de/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf> accessed 25 July 2022

Allen QC R and Masters D, 'Regulating for An Equal AI: A New Role For Equality Bodies Meeting the New Challenges to Equality and Non-Discrimination from Increased Digitisation and the Use of Artificial Intelligence' (EQUINET 2020) <https://equineteurope.org/equinet-report-regulating-for-an-equal-ai-a-new-role-for-equality-bodies/> accessed 1 August 2022

Australian Human Rights Commission, 'Human Rights and Technology Issues Paper' (2018) <https://tech.humanrights.gov.au/sites/default/files/2018-07/Human%20Rights%20and%20Technology%20Issues%20Paper%20FINAL.pdf> accessed 11 June 2023

Bielefeldt H, UN Human Rights Council Special Rapporteur on Freedom of Religion or Belief, and UN Human Rights Council Secretariat, 'Report of the Special Rapporteur on Freedom of Religion or Belief' (UN, 2015) U.N. Doc. A/HRC/31/18 <https://undocs.org/A/HRC/31/18>

Brando Benifei and Ioan-Dragoş Tudorache, 'Draft Compromise Amendments on the Draft Report - Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Artificial Intelligence (Artificial Intelligence

Act) and Amending Certain Union Legislative Acts' (European Parliament Committee on the Internal Market and Consumer Protection and Committee on Civil Liberties, Justice and Home Affairs 2023) KMB/DA/AS <https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf>

Business for Social Responsibility (BSR), 'Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021 Insights and Recommendations' (2022)

Castro D and McLaughlin M, 'Who Is Winning the AI Race: China, the EU, or the United States? — 2021 Update' (Center for Data Innovation 2021) <https://www2.datainnovation.org/2021-china-eu-us-ai.pdf>

Christophe Lacroix, 'Report of the Committee on Equality and Non-Discrimination on Preventing Discrimination Caused by the Use of Artificial Intelligence' (Council of Europe Parliamentary Assembly Committee on Equality and Non-Discrimination 2020) Doc. 15151 <https://pace.coe.int/pdf/afacbf21374dd109246029ed6e5ccd4e9317509f2ef3131cf1db1c6d10b96bb8/doc.%2015151.pdf> accessed 20 May 2023

Chui M and Malhotra S, 'Notes from the AI Frontier: AI Adoption Advances, but Foundational Barriers Remain' (McKinsey & Company 2018) <https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain> accessed 26 July 2022

Dignum V, Muller C and Theodorou A, 'Final Analysis of the EU Whitepaper on AI' (ALLAI 2020) <https://allai.nl/wp-content/uploads/2020/06/ALLAI-Final-Analysis-of-the-EU-Whitepaper-on-AI-consultation.pdf>

Dunja Mijatović, Council of Europe Commissioner for Human Rights, 'Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights' <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64> accessed 27 July 2022

Ernst E, Merola R and Samaan D, 'The Economics of Artificial Intelligence: Implications for the Future of Work' (International Labour Organization 2018) <https://www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_647306.pdf>

EU-US Trade and Technology Council, 'The Impact of Artificial Intelligence on the Future of Workforces in the EU and the US' (European Commission 2022) <https://digital-strategy.ec.europa.eu/en/library/impact-artificial-intelligence-future-workforces-eu-and-us>

European Commission for the Efficiency of Justice (CEPEJ), 'Possible Introduction of a Mechanism for Certifying Artificial Intelligence Tools and Services in the Sphere of Justice and the Judiciary: Feasibility Study' (Council of Europe 2020) CEPEJ(2020)15Rev <https://rm.coe.int/feasability-study-en-cepej-2020-15/1680a0adf4> accessed 20 May 2023

European Union Agency for Fundamental Rights, 'Facial Recognition Technology: Fundamental Rights Considerations in the Context of Law Enforcement' (Publications Office 2020) <https://data.europa.eu/doi/10.2811/524628> accessed 6 August 2022

Fenech M, Strukelj N and Buston O, 'Ethical, Social, and Political Challenges of Artificial Intelligence in Health' (Future Advocacy 2018) <https://wellcome.org/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf>

Fjeld J and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI' (Berkman Klein Center for Internet & Society 2020) Berkman Klein Center Research Publication No. 2020-1 <https://papers.ssrn.com/abstract=3518482>

Future of Privacy Forum, 'Unfairness By Algorithm: Distilling the Harms of Automated Decision-Making' (Future of Privacy Forum 2017) <https://fpf.org/blog/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/>

Gerards J and Xenidis R, 'Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non Discrimination Law' (Publications Office of the European Union 2021) <https://data.europa.eu/doi/10.2838/544956> accessed 4 August 2022

Grother P, Ngan M and Hanaoka K, 'Face Recognition Vendor Test Part 3: Demographic Effects' (National Institute of Standards and Technology 2019) NIST IR 8280

<https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf> accessed 8 April 2023

High-Level Expert Group on AI (AI HLEG), 'A Definition of AI: Main Capabilities and Scientific Disciplines' (European Commission 2019) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651> accessed 6 August 2022

Info-communications Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC), 'Model AI Governance Framework (Second Edition)' (Singapore Digital (SG:D) 2020) <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>

Johan Wolswinkel, 'Comparative Study on Administrative Law and the Use of Artificial Intelligence and Other Algorithmic Systems in Administrative Decision-Making in the Member States of the Council of Europe' (Council of Europe 2022) CDCJ(2022)31 <https://www.coe.int/documents/22298481/0/CDCJ%282022%2931E+-+FINAL+6.pdf/4cb20e4b-3da9-d4d4-2da0-65c11cd16116?t=1670943260563> accessed 19 May 2023

Jones K, 'AI Governance and Human Rights: Resetting the Relationship' (Royal Institute of International Affairs 2023) <https://chathamhouse.soutron.net/Portal/Public/en-GB/RecordView/Index/191961> accessed 21 April 2023

Laura Galindo, 'An Overview of National AI Strategies and Policies' (OECD 2021) DSTI/CDEP(2020)21/FINAL <https://goingdigital.oecd.org/data/notes/No14_ToolkitNote_AIStrategies.pdf>

McKinsey & Company, 'Shaping the Digital Transformation in Europe - Working Paper: Economic Potential' (European Commission 2020) SWD(2021) 84 final <https://digital-strategy.ec.europa.eu/en/library/shaping-digital-transformation-europe-working-paper-economic-potential>

Mueller B, 'How Much Will the Artificial Intelligence Act Cost Europe?' (Center for Data Innovation 2021) <https://www2.datainnovation.org/2021-aia-costs.pdf> accessed 26 February 2023

OECD, 'Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers' (OECD 2021) <, https://www.oecd.org/finance/artificial-intelligence-machine-learningbig-data-in-finance.htm>

Office of the United Nations High Commissioner for Human Rights, 'The Right to Privacy in the Digital Age' (UN Human Rights Office 2021) A/HRC/48/31 <https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session48/Documents/A_HRC_48_31_AdvanceEditedVersion.docx>

Perry WL and others, 'Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations' (RAND Corporation 2013) <https://www.rand.org/pubs/research_reports/RR233.html> accessed 8 June 2023

Raso FA and others, 'Artificial Intelligence & Human Rights: Opportunities & Risks' (Berkman Klein Center for Internet & Society 2018) Berkman Klein Center Research Publication No. 2018-6 <https://papers.ssrn.com/abstract=3259344>

Renda A and others, 'Study to Support an Impact Assessment of Regulatory Requirements for Artificial Intelligence in Europe: Final Report.' (Publications Office 2021) <https://data.europa.eu/doi/10.2759/523404> accessed 21 May 2023

Schennach S, 'Report of the Committee on Social Affairs, Health and Sustainable Development on Artificial Intelligence and Labour Markets: Friend or Foe?' (Council of Europe Parliamentary Assembly Committee on Social Affairs, Health and Sustainable Development 2020) Doc. 15159 <https://pace.coe.int/pdf/7d08c3b0034d4a3cca6bd3e8ce6a369e73850fbce3c2b94d758130438218631d/doc.%2015159.pdf> accessed 4 August 2022

Selin Sayek Böke, 'Report of the Committee on Social Affairs, Health and Sustainable Development on Artificial Intelligence in Health Care: Medical, Legal and Ethical Challenges Ahead' (Council of Europe Parliamentary Assembly Committee on Social Affairs, Health and Sustainable Development 2020) Doc. 15154 <https://pace.coe.int/pdf/a845d11a279c1ca4ce2896a208196db8b11e79b4226d3d4135c23dd8969a23a3/doc.%2015154.pdf> accessed 4 August 2022

Shaheed A, UN Human Rights Council Special Rapporteur on Freedom of Religion or Belief, and UN Human Rights Council Secretariat, 'Report of the Special

Rapporteur on Freedom of Religion or Belief' (UN, 2019) U.N. Doc. A/HRC/40/58 <https://undocs.org/A/HRC/40/58>

Smith G and Rustagi I, 'Mitigating Bias in Artificial Intelligence: An Equity Fluent Leadership Playbook' (Berkeley Haas Center for Equity, Gender and Leadership 2020) <https://haas.berkeley.edu/wp-content/uploads/UCB_Playbook_R10_V2_spreads2.pdf>

Special Rapporteur on freedom of opinion and expression, 'Report on Artificial Intelligence Technologies and Implications for Freedom of Expression and the Information Environment' (United Nations General Assembly 2018) A/73/348 <https://undocs.org/Home/Mobile?FinalSymbol=A%2F73%2F348&Language=E&DeviceType=Desktop&LangRequested=False>

Steering Committee for Media and and Information Society (CDMSI), 'Content moderation - Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation' (Council of Europe 2021) Guidance Note <https://edoc.coe.int/fr/internet/10198-content-moderation-guidance-note.html> accessed 19 May 2023

The Dutch Data Protection Authority (Autoriteit Persoonsgegevens), 'Conclusions Facebook 2017' (Autoriteit Persoonsgegevens 2017) <https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/conclusions_facebook_february_23_2017.pdf> accessed 21 March 2023

The Royal Society, 'Machine Learning: The Power and Promise of Computers That Learn by Example' (The Royal Society 2017) DES4702 <https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

The Working Group on AI and Criminal Law and Sabine Gless, 'European Committee on Crime Problems (CDPC) Feasibility Study on a Future Council of Europe Instrument on Artificial Intelligence and Criminal Law' (Council of Europe 2020) CDPC(2020)3Rev <https://rm.coe.int/cdpc-2020-3-feasibility-study-of-a-future-instrument-on-ai-and-crimina/16809f9b60>

UNI Global Union, 'Top 10 Principles for Ethical Artificial Intelligence' (UNI Global Union 2017)

<http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf> accessed 22 May 2023

Wagner B, 'Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications' (Council of Europe 2018) DGI(2017)12 <https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10> accessed 4 August 2022

Yeung K, 'A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework' (2019) DGI(2019)05 <https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168096bdab>

8. Other Resources

'AI at Google: Our Principles' (*Google*, 7 June 2018) <https://blog.google/technology/ai/ai-principles/> accessed 19 May 2023

'Facebook's five pillars of Responsible AI' (*Meta AI*, 22 June 2021) <https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/> accessed 10 June 2023

'Google AI Principles' (*Google AI*) <https://ai.google/responsibility/principles/> accessed 10 June 2023

'Google Apologises for Photos App's Racist Blunder' *BBC News* (1 July 2015) <https://www.bbc.com/news/technology-33347866> accessed 6 April 2023

Allen A, 'The "Three Black Teenagers" Search Shows It Is Society, Not Google, That Is Racist' *The Guardian* (10 June 2016) <https://www.theguardian.com/commentisfree/2016/jun/10/three-black-teenagers-google-racist-tweet> accessed 6 April 2023

Amnesty International and Access Now, 'The Toronto Declaration - Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems' <https://www.torontodeclaration.org/declaration-text/english/>

Angwin J and others, 'Machine Bias' [2016] *ProPublica* <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Bagli CV, 'Facebook Vowed to End Discriminatory Housing Ads. Suit Says It Didn't.' *The New York Times* (27 March 2018) <https://www.nytimes.com/2018/03/27/nyregion/facebook-housing-ads-discrimination-lawsuit.html> accessed 21 March 2023

Barber I, 'First Thoughts on the Revised Zero Draft of the Council of Europe's AI Treaty | Global Partners Digital' [2023] *Global Partners Digital* <https://www.gp-digital.org/first-thoughts-on-the-revised-zero-draft-of-the-council-of-europes-ai-treaty/> accessed 10 May 2023

Bass D, 'Microsoft's New Chatbot Zo Won't Talk Politics or Racism' *Bloomberg.com* (5 December 2016) <https://www.bloomberg.com/news/articles/2016-12-05/microsoft-s-new-chatbot-zo-won-t-talk-politics-or-racism> accessed 16 February 2023

Bergstrom C and West J, 'Case Study Criminal Machine Learning' [2017] *Calling Bullshit* <https://www.callingbullshit.org/case_studies/case_study_criminal_machine_learning.html?fbclid=>

Booth R, 'Automated UK Welfare System Needs More Human Contact, Ministers Warned' *The Guardian* (22 May 2023) <https://www.theguardian.com/society/2023/may/22/automated-uk-welfare-system-needs-more-human-contact-ministers-warned> accessed 8 June 2023

Bullinaria JA, 'IAI : The Roots, Goals and Sub-Fields of AI' (Introduction to Artificial Intelligence, University of Birmingham, 2005) <https://www.cs.bham.ac.uk/~jxb/IAI/w2.pdf>

Butler P, 'UK Government Faces £150m Bill over Social Welfare Discrimination' *The Guardian* (21 January 2022) <https://www.theguardian.com/society/2022/jan/21/uk-government-bill-welfare-discrimination-universal-credit> accessed 12 June 2023

Capps K, 'Behind HUD's Housing Discrimination Charges Against Facebook' *Bloomberg.com* (28 March 2019) <https://www.bloomberg.com/news/articles/2019-03-28/why-hud-charged-facebook-with-discrimination> accessed 21 March 2023

Charles Duhigg, 'How Companies Learn Your Secrets' *The New York Times* (16 February 2012) <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html> accessed 6 April 2023

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

Dastin J, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women' *Reuters* (10 October 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> accessed 8 March 2023

Dietterich TG, 'Machine Learning' in Edwin D Reilly, Anthony Ralston and David Hemmendinger (eds), *Encyclopedia of Computer Science* (4th edn, John Wiley & Sons, Ltd 2003)

Dietterich TG, 'Machine Learning' in Lynn Nadel (ed), *Encyclopedia of Cognitive Science* (John Wiley & Sons, Ltd 2006) <https://onlinelibrary.wiley.com/doi/pdf/10.1002/0470018860.s00036 DOI: 10.1002/0470018860.s00036>

Eliot L, 'AI Ethics And Legal AI Are Flustered By Deceptive Pretenses Known As AI Ethics-Washing Which Are False Claims Of Adhering To Ethical AI, Including For Autonomous Self-Driving Cars' [2022] *Forbes* <https://www.forbes.com/sites/lanceeliot/2022/06/09/ai-ethics-and-legal-ai-are-flustered-by-deceptive-pretenses-known-as-ai-ethics-washing-which-are-false-claims-of-adhering-to-ethical-ai-including-for-autonomous-self-driving-cars/> accessed 9 June 2023

Etzioni O, 'Opinion | How to Regulate Artificial Intelligence' *The New York Times* (2 September 2017) <https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html> accessed 1 August 2022

European Union Agency for Fundamental Rights, '#BigData: Discrimination in Data-Supported Decision Making' <https://fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making> accessed 6 August 2022

Future of Life Institute, 'The Asilomar AI Principles' (*Future of Life Institute*, 11 August 2017) <https://futureoflife.org/open-letter/ai-principles/>

Google AI, 'Our Principles' (*Google AI*) <https://ai.google/principles/> accessed 10 April 2023

Google Developers, 'Bias (Ethics/Fairness)' <https://developers.google.com/machine-learning/glossary#bias-ethicsfairness> accessed 12 August 2022

Gorner J, 'Chicago Police Use "heat List" to Prevent Violence' *Chicago Tribune* (23 August 2013) <https://www.police1.com/chiefs-sheriffs/articles/chicago-police-use-heat-list-to-prevent-violence-aYQ4dZmhaqIvEB2g/> accessed 21 February 2023

IBM, 'Everyday Ethics for Artificial Intelligence' (IBM 2022) <https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf> accessed 23 May 2023

IBM, 'IBM's Principles for Trust and Transparency' (*IBM Policy*, 30 May 2018) <https://www.ibm.com/policy/trust-principles/>

IBM, 'What Is Natural Language Processing? | IBM' <https://www.ibm.com/topics/natural-language-processing>

Karakaş İ, 'Yargıtay Başkanı Akarca: Yapay Zeka Destekli Yargıtay İçtihat Merkezi, En Geç Haziranda Faaliyete Geçecek' <https://www.aa.com.tr/tr/gundem/yargitay-baskani-akarca-yapay-zeka-destekli-yargitay-ictihat-merkezi-en-gec-haziranda-faaliyete-gececek/2790780> accessed 13 March 2023

Kreger A, 'Council Post: The Future Of AI In Banking' *Forbes* <https://www.forbes.com/sites/forbesbusinesscouncil/2023/03/20/the-future-of-ai-in-banking/> accessed 8 June 2023

Lashbrook A, 'AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind' *The Atlantic* (16 August 2018) <https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/> accessed 3 December 2022

McCarthy J, 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>

McCarthy J, 'What Is Artificial Intelligence?' <http://www-formal.stanford.edu/jmc/whatisai.pdf>

Metzinger T, 'EU guidelines: Ethics washing made in Europe' *Der Tagesspiegel Online* (8 April 2019) <https://www.tagesspiegel.de/politik/ethics-washing-made-in-europe-5937028.html> accessed 16 May 2023

Microsoft Corporation, 'Microsoft Responsible AI Standard v2 General Requirements' (Microsoft 2022) <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5cmFl>

Nicoletti L and Bass D, 'Generative AI Takes Stereotypes and Bias From Bad to Worse' [2023] *Bloomberg* <https://www.bloomberg.com/graphics/2023-generative-ai-bias/> accessed 12 June 2023

Oremus W, 'The Clever Trick That Turns ChatGPT into Its Evil Twin' *Washington Post* (15 February 2023) <https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/> accessed 16 February 2023

Pretz K, 'Stop Calling Everything AI, Machine-Learning Pioneer Says' [2021] *IEEE Spectrum* <https://spectrum.ieee.org/stop-calling-everything-ai-machinelearning-pioneer-says>

Reuters, 'Deepfakes Are Biggest AI Concern, Says Microsoft President' *The Guardian* (25 May 2023) <https://www.theguardian.com/technology/2023/may/25/deepfakes-ai-concern-microsoft-brad-smith> accessed 9 June 2023

Richtel M, 'How Big Data Is Playing Recruiter for Specialized Workers - The New York Times' (27 April 2013) <https://www.nytimes.com/2013/04/28/technology/how-big-data-is-playing-recruiter-for-specialized-workers.html> accessed 8 March 2023

Simonite T, 'When It Comes to Gorillas, Google Photos Remains Blind' [2018] *Wired* <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/> accessed 6 April 2023

Stuart-Ulin CR, 'Microsoft's Politically Correct Chatbot Is Even Worse than Its Racist One' [2018] *Quartz* <https://qz.com/1340990/microsofts-politically-correct-chat-bot-is-even-worse-than-its-racist-one/> accessed 16 February 2023

T.C. Adalet Bakanlığı, 'Yargı Reformu Stratejisi - 2019' <https://sgb.adalet.gov.tr/Resimler/SayfaDokuman/23122019162931YRS_TR.pdf>

The Guardian, 'Microsoft "deeply Sorry" for Racist and Sexist Tweets by AI Chatbot' *The Guardian* (26 March 2016) <https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot> accessed 16 February 2023

Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 'Ulusal Yapay Zekâ Stratejisi 2021-2025' <https://cbddo.gov.tr/uyzs>

University of Virginia, 'What the Heck Is a Deepfake? | Information Security at UVA, U.Va.' (*Information Security at UVA*) <https://security.virginia.edu/deepfakes> accessed 14 June 2023

van Veen C and Cath C, 'Artificial Intelligence: What's Human Rights Got To Do With It?' [2018] *Medium* <https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5> accessed 6 June 2023

Vynck GD, Lerman R and Tiku N, 'Microsoft's AI Chatbot Is Going off the Rails' *Washington Post* (17 February 2023) <https://www.washingtonpost.com/technology/2023/02/16/microsoft-bing-ai-chatbot-sydney/> accessed 18 February 2023

Wakefield J, 'Microsoft Chatbot Is Taught to Swear on Twitter' *BBC News* (24 March 2016) <https://www.bbc.com/news/technology-35890188> accessed 16 February 2023

Weise K, 'Microsoft's Bing Chatbot Offers Some Puzzling and Inaccurate Responses' *The New York Times* (15 February 2023) <https://www.nytimes.com/2023/02/15/technology/microsoft-bing-chatbot-problems.html> accessed 17 February 2023

World Economic Forum Global Future Council on Human Rights 2016-18, 'How to Prevent Discriminatory Outcomes in Machine Learning' <https://www.weforum.org/whitepapers/how-to-prevent-discriminatory-outcomes-in-machine-learning/> accessed 10 February 2022

York C, 'Type "Three Black Teenagers" Into Google And Something Racist Happens. Or Does It?' *HuffPost UK* (8 June 2016)

<https://www.huffingtonpost.co.uk/entry/three-black-teenagers-google-racism_uk_575811f5e4b014b4f2530bb5> accessed 6 April 2023