FACULTY OF LAW
Lund University

Can Yavuz

Machine Bias:
Artificial Intelligence and Discrimination

JAMM07 Master Thesis

International Human Rights Law
30 higher education credits

Supervisor: Karol Nowak

Term: Spring term 2019

Abstract

The past decade has seen the rapid development of artificial intelligence. It has resulted in extensive usage and reliance within many diverse fields that influences our daily lives as well as human rights, and especially the prohibition of discrimination. The thesis examines artificial intelligence discrimination and asks why and how it occurs, who is (more likely to be) affected by it, and how policymakers should respond to protect human rights. The findings reveal that artificial intelligence discriminates in various ways, and the most vulnerable and discriminated groups are more likely to be victims of it. Many problems in the field stem from lack of regulation and over-reliance on artificial intelligence. This thesis makes a preliminary recommendation and invites policymakers to cautiously regulate artificial intelligence to prevent artificial intelligence discrimination.

Keywords: artificial intelligence, human rights, discrimination, prohibition of discrimination, regulation

Table of Contents

List of Figures

Abbreviations

AI	- Artificial intelligence
CEO	- Chief executive officer
ECHR	- European Convention on Human Rights
ECtHR	- European Court of Human Rights
NGO	- non-governmental organization
USA	- United States of America

Acknowledgments

I would like to extend my deepest gratitude to the Swedish Institute for offering me a scholarship, which enabled me to broaden my horizon on my passion, human rights. I also would like to express my appreciation to my thesis supervisor Professor Karol Nowak for his unparalleled support and feeding the thesis with thought-provoking ideas. And special thanks to Emily and Jack for proofreading.

Writing this section evokes many memories of two marvelous years, which would not have been possible without my family in Sweden. And my heartfelt appreciation goes to four friends. I am deeply indebted to Ecem and Lidia who made me feel home in a dorm that I cannot pronounce its name. And I am extremely grateful to Alessia and Luana for their endless energy and love, which resulted in memorable brunches, potlucks, parties, trips, and festivities.

CHAPTER 1: INTRODUCTION

1.1. Background

Today, artificial intelligence (AI) is widely used in diversified areas. In healthcare, it assists in detecting tumor and cancer, discovering new medicine, and operating surgery. The algorithms[1] of Netflix, YouTube, and Spotify dominate online content consumption and suggest movies, videos, and songs to people. Virtual assistants (Siri and Alexa) ease daily life, web mapping services (Google Maps) forecast the best route; search engines enable the Internet users to find relevant information on the Internet. Many AI applications are so well-integrated in daily life that people do not realize the use of them.

After witnessing these assets of AI, the research entirely aimed at developing it further and forgotten impact assessments. It has become clear the scales have been tipped in favor of technological development over the respect for human rights. Today, data-driven tech companies make the right to privacy optional. Autonomous weapons imperil the right to life. Filter bubble[2], AI-driven online disinformation, and AI-content moderation undermine freedom of speech. Freedom of assembly is near threatened by facial recognition systems. AI-assisted court decisions weaken right to a fair trial. And as will be unfolded in this thesis, artificial intelligence can and does discriminate.

1.2. Purpose

> "[F]acing possible futures of incalculable benefits and risks [of artificial intelligence], the experts are surely doing everything possible to ensure the best outcome, right? Wrong. If a superior alien civilization sent us a text message saying, 'We will arrive in a few decades,' would we just reply, 'OK, call us when you get here — we will leave the lights on'? Probably not — but this is more or less what is happening with AI.[3]"

The purpose of this exploratory thesis is to investigate artificial intelligence discrimination: how it occurs, the underlying reasons and potential victims of it, and how lawmakers address these issues to prevent artificial intelligence discrimination.

[1] Algorithm can be loosely described as "[a] finite suite of formal rules/commands, usually in the form of a mathematical logic, that allows for a result to be obtained from input elements." (Council of Europe Commissioner for Human Rights, 2019, p. 24)

[2] It is argued that to maximize clicks, search engines and social media platforms provide a personalized service to users based on their past online behavior. This leads internet users to more frequently encounter the ideas and people of the same mind. It is thought that filter bubble creates constant self-affirmation and echo chambers that weakens freedom of thought.

[3] (Hawking, Tegmark, Russell, & Wilczek, 2014)

1.3. Research Questions

This thesis seeks to answer the following research questions, namely:

- *Why does artificial intelligence discriminate?*
- *How does artificial intelligence discriminate?*
- *Who is (more likely to be) discriminated by artificial intelligence?*
- *Should legislator regulate artificial intelligence?*
- *How should lawmakers regulate artificial intelligence to prevent artificial intelligence discrimination?*

1.4. Method

The thesis at hand is an exploratory and inter-disciplinary legal research that incorporates various methodologies. The second and third chapter employ descriptive approach and aim to clarify artificial intelligence and prohibition of discrimination; thereby, the following chapters can build upon them. The fourth chapter aims to illustrate why, how, and who artificial intelligence discriminates. The fifth chapter uses case study method and conducts an in-depth examination of three crucial artificial intelligence applications (search engines, facial recognition system, and risk assessment tools). The last examines why and how lawmakers should regulate artificial intelligence.

1.5. Material

The thesis uses diversified research materials that include: books, academic articles and researches; reports of the European Union, the Council of Europe, the United Nations, human rights and artificial intelligence non-governmental organizations; national, regional, and international human rights laws; case law of the American and Canadian high courts, the European Court of Human Rights, and the European Court of Justice; reliable Internet news sources, and credible opinion columns. The research materials are taken from various disciplines and fields, including but not limited to, artificial intelligence, law, human rights law, economics, politics, sociology, criminology, history, and linguistics.

1.6. Literature Review

Many researchers pointed out that artificial intelligence can severely and systematically discriminate. Reaffirming that discrimination is a widespread and critical human rights issue at present, and AI is capable of taking it to a higher tier, AI discrimination requires a great deal of attention. However, to this author’s best knowledge, legal and human rights researchers have paid little attention to artificial intelligence discrimination. It is true that some activists, artificial intelligence researchers, and human rights non-governmental organizations (NGO) have made artificial intelligence bias publicly known. After many AI discrimination cases hit the headlines, legal researchers and international institutions showed an increased interest in AI-bias, particularly in recent years. In 2018, the European Council published one of the most comprehensive human rights reports on AI-driven discrimination and how to address it.[4] In their detailed investigation into AI-bias, Barocas and Selbst showed the importance of training data in data mining[5], how AI can discriminate vulnerable and disadvantaged groups, and the shortcomings of American antidiscrimination law.[6] The United Nations Special Rapporteur and the Council of Europe Commissioner for Human Rights raised concerns on AI discrimination in their reports.[7] Some experts shed light on AI discrimination in specific domains. Safiya Noble made an impressive analysis of search engine discrimination.[8] Joy Buolamwini[9] and the American Civil Liberties Union[10] demonstrated facial recognition systems underperform on women and people of color. ProPublica revealed how a leading risk assessment tool mislabels Black defendants as high-risk offenders.[11] The Council of Europe report underlined that predictive justice tools generate discrimination in the judicial process.[12]

The number of published papers on AI has significantly increased in the last decade. This is particularly true in machine learning[13] and probabilistic reasoning, neural networks, computer vision, and search and optimization.[14] On the other hand, it is safe to argue that legal researchers have not treated artificial intelligence and its impacts on human rights in much detail. Bearing in mind that artificial intelligence is a dynamic and diverse field that has the potential to dominate the future, further research and debate including the usage of a human rights lens and multidisciplinary approaches are urgently required.

[4] (Borgesius, 2018)

[5] "Datamining makes it possible to analyze a large volume of data and bring out models, correlations and trends." (The Council of Europe, 2019)

[6] (Barocas & Selbst, 2016)

[7] (David K. , 2018), (Council of Europe Commissioner for Human Rights, 2019)

[8] (Noble, 2018),

[9] (Buolamwini, 2018)

[10] (Snow , 2018)

[11] (Julia, Jeff, Surya, Lauren, & ProPublica, 2016)

[12] (The European Commission for the Efficiency of Justice of the Council of Europe, 2018)

[13] The term machine learning refers to "[a] field of AI made up of a set of techniques and algorithms that can be used to 'train' a machine to automatically recognize patterns in a set of data. By recognizing 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." (Council of Europe Commissioner for Human Rights, 2019, p. 24)

[14] (AI Index, 2018, p. 9)

1.7. Limitations

Due to practical constraints (such as limited time and legal resources, the complexity of the thesis topic), the scope and quality of this thesis are restricted in several aspects. The reader should bear in mind that the aim of this thesis is limited to conducting exploratory research.

Firstly, artificial intelligence is a convoluted and interdisciplinary field of study that requires knowledge of many areas. The researcher's lack of expertise in artificial intelligence and interconnected domains may have adversely influenced the quality of the thesis. Turning to the scope of the thesis, the potential impacts of artificial intelligence on human rights, particularly the prohibition of discrimination, is a broad and complicated topic. In the long term, artificial intelligence may create challenges that experts cannot foresee at present. Therefore, this thesis does not attempt to address the long-term effects of artificial intelligence; instead, it focuses on the current and near-future impacts. Another limitation concerns artificial intelligence regulation. This thesis proposes the hypothesis that artificial intelligence should be regulated to prevent AI discrimination and encourages lawmakers to be involved in the regulatory process. Artificial intelligence regulation is a developing, complex, and far-reaching topic. Thus, the thesis cannot provide a comprehensive review of it. Lastly, to limit the scope of the thesis, it generally focuses on the United States of America (USA), the European Union, and China, the leading actors in AI.[15]

1.8. Outline

This thesis consists of six chapters and the structure of it as follows.

- The first chapter focuses on the scope of the study: its background, purpose, research questions, methodology, material, literature review, and delimitations.
- The second chapter is devoted to artificial intelligence and respectively discusses the definition, history, current state, and future of artificial intelligence.
- The third chapter introduces the non-discrimination principle and covers the definition of discrimination, discrimination types, and protected discrimination grounds.
- The fourth chapter analyzes artificial intelligence discrimination. Firstly, it establishes the nexus between artificial intelligence and discrimination, then aims to uncover why, how, and whom artificial intelligence discriminates. Subsequently, the chapter bridges the connection between AI discrimination and protected grounds of discrimination.

[15] (AI Index, 2018, p. 10)

- The fifth chapter concentrates on case studies and examines how facial recognition systems, search engines, and risk assessment tools discriminate. The chapter aims to address the root causes of AI discrimination as well as propose solutions.
- The last chapter examines artificial intelligence regulation. It discusses whether artificial intelligence should be regulated, who should regulate AI, and how laws may prevent artificial intelligence discrimination.

CHAPTER 2: ARTIFICIAL INTELLIGENCE

Artificial intelligence may be quite a technical concept. Therefore, this chapter's principal objective is to provide necessary background information on artificial intelligence. The chapter will briefly discuss the definition, history, current state, and future of artificial intelligence.

2.1. What is Artificial Intelligence?

Perhaps artificial intelligence is not a familiar term for many. This section aims to define artificial intelligence. Further, many form opinions on artificial intelligence based on Hollywood images that may not be a true representative of current reality. Therefore, the section aims to unravel some of the mysteries surrounding artificial intelligence. The section firstly will try to define artificial intelligence, then address the distinction between strong and weak artificial intelligence.

Artificial intelligence became a buzzword in the last decade, but it is in the literature for some time. The term "artificial intelligence" was invented in 1956 by John McCarthy, an American computer scientist.[16] Decades after the invention, artificial intelligence is considered a nebulous term and a wide range of definitions are available in the doctrine.

An obscure term: In the view of two foremost researchers in the field, Norvig and Russell, the term artificial intelligence embodies four different approaches: (1) thinking humanly, (2) thinking rationally, (3) acting humanly, and (4) acting rationally. The figure below explains different perspectives.

[16] (Wichert, 2014, p. 1)

Figure 1 - Different definitions of artificial intelligence[17]

Thinking Humanly	**Thinking Rationally**
“The exciting new effort to make computers think . . . machines with minds, in the full and literal sense.” “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .”	“The study of mental faculties through the use of computational models.” “The study of the computations that make it possible to perceive, reason, and act.”
Acting Humanly	**Acting Rationally**
“The art of creating machines that perform functions that require intelligence when performed by people.” “The study of how to make computers do things at which, at the moment, people are better.”	“Computational Intelligence is the study of the design of intelligent agents.” “AI . . . is concerned with intelligent behavior in artifacts.”

Since the definition of artificial intelligence varies among researchers, it may be useful to see more descriptions. A leading scholar, Winston, uses the term artificial intelligence refer to “the study of the computations that make it possible to perceive, reason, and act.[18]” Coppin cites a well-known definition, “[a]rtificial [i]ntelligence involves using methods based on the intelligent behavior of humans and other animals to solve complex problems.[19]” A further definition of artificial intelligence is given by the Council of Europe as follows “[a] set of sciences, theories and techniques whose purpose is to reproduce by a machine the cognitive abilities of a human being. Current developments aim, for instance, to be able to entrust a machine with complex tasks previously delegated to a human.[20]”

Less academic and straightforward definitions of artificial intelligence can be found in dictionaries and encyclopedias. According to a definition provided by Encyclopedia Britannica, artificial intelligence is “the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.[21]” For Cambridge Dictionary Online, artificial intelligence means “the use of computer programs that have some

[17] The figure is taken from (Norvig & Russell, 2010, p. 2)

[18] (Patrick, 1992, p. 5)

[19] (Coppin, 2004, p. 4)

[20] (The Council of Europe, 2019)

[21] (Copeland, 2019)

of the qualities of the human mind, such as the ability to understand language, recognize pictures, and learn from experience.[22]"

As noted above, artificial intelligence remains a poorly defined term. It is argued that to define intelligence is burdensome; therefore, it is difficult to define artificial intelligence.[23] Another issue is the diversity of artificial intelligence, which makes it challenging to define the term.[24] Throughout this thesis, the term artificial intelligence is used as an umbrella term that covers machines which can achieve goals that require intelligence and the science behind it.[25]

Strong vs. weak artificial intelligence: To properly comprehend artificial intelligence, it could be appropriate to make a distinction and categorize it into weak and strong artificial intelligence.[26] Weak artificial intelligence "is the ability of machines to resemble human capabilities in narrow domains, with different degrees of technical sophistication and autonomy.[27]"

Strong artificial intelligence "is overarching, and as yet unachieved, goal of a system that displays intelligence across multiple domains, with the ability to learn new skills, and which mimic or even surpass human intelligence.[28]" Strong artificial intelligence is a well-known Hollywood image. The leading characters in movies such as 2001: A Space Odyssey, The Terminator, Her, and Ex Machina displayed strong artificial intelligence. At present, strong artificial intelligence remains as a hypothetical concept, and it is disputable whether, or when, science will reach that point.

To compare strong and weak artificial intelligence, it is plausible to argue that weak artificial intelligence is trained and performs well in one particular field. On the other hand, strong artificial intelligence aims to have a mindset similar to that of a human.[29] In the perspective of

[22] (Cambridge Dictionary Online, 2019)

[23] (Coppin, 2004, p. 15)

[24] (Technology, 2016, p. 7)

[25] (Council of Europe Commissioner for Human Rights, 2019, p. 5)

[26] John Rogers Searle is the founder of the term strong artificial intelligence. In his notable paper which introduced "Chinese room" hypothesis, he defines the term as "the appropriately programmed computer literally has cognitive states and that the programs thereby explain human cognition." He also claims that strong AI is false because simulating a mind does not mean creating a mind. Put differently, a program cannot be a mind. Programs are solely syntactical which lacks semantics, what human mind has, hence a program is not a mind. (Searle, 1980) Some literature uses weak and narrow artificial intelligence; strong and general artificial intelligence interchangeably. The thesis will use the terms weak and strong artificial intelligence.

[27] (ARTICLE 19, 2018, p. 6)

[28] (ARTICLE 19, 2018, p. 6)

[29] (Coppin, 2004, pp. 688, 693)

philosophers, weak artificial intelligence would possibly behave intelligently, and strong artificial intelligence would have an actual mind, not a simulated one.[30]

2.2. A Short History and Current State of Artificial Intelligence

This section takes a brief look at the history, development, current state, and future of artificial intelligence. The historical perspective may shed light on the future of artificial intelligence and how society receives the development of AI.

Artificial intelligence is a young interdisciplinary field of study.[31] The first work on artificial intelligence, a model of artificial neurons, can be traced back to 1943.[32] Another early example was the first neural network computer in 1950. Perhaps the most influential early research regarding artificial intelligence is Alan Turing’s famous article (published in 1950) called *Computing Machinery & Intelligence* which introduced the Turing Test.[33]

> “The test is for a program to have a conversation (via online typed messages) with an interrogator for five minutes. The interrogator then has to guess if the conversation is with a program or a person; the program passes the test if it fools the interrogator 30% of the time.[34]”

1956 is considered artificial intelligence’s year of birth. John McCarthy and nine researchers conducted the very first research on artificial intelligence.[35] In the same year, McCarthy introduced the term artificial intelligence to the literature. During the early era of artificial intelligence, computers had limited capabilities. Consequently, whatever artificial intelligence could do was seen as extraordinary. It led to immense anticipation which could not be met for decades. Therefore, this period (1952-1969) is referred to as “early enthusiasm, great

[30] (Norvig & Russell, 2010, p. 1040)

[31] (The Council of Europe, 2019)

[32] (Negnevitsky, 2005, p. 5), (The Council of Europe, 2019)

[33] (Norvig & Russell, 2010, pp. 16-17), (The Council of Europe, 2019)

[34] (Norvig & Russell, 2010, p. 1021)
A more detailed description of the test as follows “The interrogator is given access to two individuals, one of whom is a human and the other of whom is a computer. The interrogator can ask the two individuals questions, but cannot directly interact with them. Probably the questions are entered into a computer via a keyboard, and the responses appear on the computer screen.
The human is intended to attempt to help the interrogator, but if the computer is really intelligent enough, it should be able to fool the interrogator into being uncertain about which is the computer and which is the human.
The human can give answers such as ‘I’m the human—the other one is the computer,’ but of course, so can the computer. The real way in which the human proves his or her humanity is by giving complex answers that a computer could not be expected to comprehend. Of course, the inventors of the truly intelligent computer program would have given their program the ability to anticipate all such complexities.” (Coppin, 2004, p. 8)

[35] “The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.” (Norvig & Russell, 2010, p. 17)

expectations.[36]" The following years (1966-1973) brought the shortcomings and sense of reality to the researchers' notice.[37]

In the next era (1969-1979) researchers added domain knowledge to artificial intelligence and started to use this combination in an area of expertise. Projects such as Dendral, Heuristic Programming, and SHRDLU system achieved significant success in the fields of (respectively) chemistry, medicine, and natural language understanding. Starting from 1980, artificial intelligence evolved into a billion-dollar industry, assisting companies in saving money as a result of the use of expert systems. Just as the early artificial intelligence researchers, the companies had great expectations for artificial intelligence, which could not be fulfilled for a long time.[38]

The early artificial intelligence research is considered rebellious and experimental. Dating from 1987, artificial intelligence research adopted a more science-based approach.[39] After establishing scientific-based grounding, the period from 1995 onward, researchers started putting more emphasis on "machines that think, that learn and that create" (human-level artificial intelligence). It developed artificial intelligence into a more multidisciplinary field. 1997 is referred to as a milestone since IBM's chess-playing computer Deep Blue which beat the world chess champion.[40]

It is asserted that beginning from 2010, artificial intelligence developed significantly as a consequence of the advancement in affordable computing power (that accelerates the calculation of learning algorithms), access to big data[41] (that effortlessly provides sampling to

[36] (Negnevitsky, 2005, pp. 6-7) "Simon also made more concrete predictions: that within 10 years a computer would be chess champion, and a significant mathematical theorem would be proved by machine. These predictions came true (or approximately true) within 40 years rather than 10. Simon's overconfidence was due to the promising performance of early AI systems on simple examples. In almost all cases, however, these early systems turned out to fail miserably when tried out on wider selections of problems and on more difficult problems." (Norvig & Russell, 2010, p. 21)

[37] (Norvig & Russell, 2010, pp. 20-22), (Negnevitsky, 2005, pp. 7-8)

[38] (Norvig & Russell, 2010, pp. 22-24), (Negnevitsky, 2005, pp. 8-12), (The Council of Europe, 2019)

[39] As noted by David McAllester "In the early period of AI it seemed plausible that new forms of symbolic computation, e.g., frames and semantic networks, made much of classical theory obsolete. This led to a form of isolationism in which AI became largely separated from the rest of computer science. This isolationism is currently being abandoned. There is a recognition that machine learning should not be isolated from information theory, that uncertain reasoning should not be isolated from stochastic modeling, that search should not be isolated from classical optimization and control, and that automated reasoning should not be isolated from formal methods and static analysis." (Norvig & Russell, 2010, p. 25)

[40] (Norvig & Russell, 2010, pp. 26-30), (Technology, 2016, pp. 5-6)

[41] "The term 'big data' refers to a large heterogeneous data set (open data, proprietary data, commercially purchased data)." (The Council of Europe, 2019)

algorithms via the Internet)[42] and new online platforms. To exemplify this rapid development, a few well-known milestones of artificial intelligence in the last decade are listed below.

- (2010) Apple Inc. released Siri (virtual assistant);[43]
- (2011) IBM supercomputer called Watson won Jeopardy (a general knowledge quiz) against some of the best players;[44]
- (2016) The first fatal semi-autonomous car[45] accident;[46]
- (2017) Google's artificial intelligence AlphaGo beat Go (a board game that is more complex than chess) world champion;[47]
- (2017) Google's artificial intelligence AlphaGo Zero learned Go board game by playing games against itself and only in 40 days surpassed previous AlphaGo versions, including the one beat the world champion;[48]
- (2017) Poker-playing artificial intelligence called Libratus beat four leading professional human poker players in no-limit Texas hold 'em;[49]
- (2017) Saudi Arabia gave citizenship to the social robot Sophia;[50]
- (2017) An artificial intelligence designed to classify skin cancer developed competence comparable to dermatologists[51];
- (2018) Alibaba's and Microsoft's artificial intelligence outscored humans in Stanford University reading and comprehension test;[52]
- (2018) Google's artificial intelligence achieved a better accuracy in grading prostate cancer in prostatectomy specimens compared to American board-certified general pathologists[53];

[42] (The Council of Europe, 2019), (ARTICLE 19, 2018, p. 4)

[43] (Bosker, 2013)

[44] (Gabbatt, 2011)

[45] The car in question is not considered a fully self-driving car. According to the Society of Automotive Engineers' classification, it may be considered a level 2 self-driving car (partial driving automation). Level 2 self-driving car can automatically steer, accelerate, and deaccelerate. Nevertheless, driver is expected to intervene is many circumstances. As the date of early 2019, many commercially available self-driving cars, including Tesla vehicles, are considered level 2 self-driving car. In the words of to the Society of Automotive Engineers, level 2 self-driving car refers to "[t]he sustained and operational design domain specific execution by a driving automation system of both the lateral and longitudinal vehicle motion control subtasks of the dynamic driving task with the expectation that the driver completes the object and event detection, recognition, classification, and response subtask and supervises the driving automation system." (Society of Automotive Engineers International, 2016, p. 17)

[46] (Yadron & Tynan, 2016)

[47] (The Guardian, 2017)

[48] (deepmind.com, 2019)

[49] (Sandholm & Noam, 2018)

[50] (Reynolds, 2018)

[51] (Andre, et al., 2017)

[52] (Louise, 2018)

[53] (Stumpe & Mermel, 2018)

- (2018) Microsoft's artificial intelligence reached human-level quality and accuracy in translating news articles from Chinese to English[54];
- (2018) Google announced Google Duplex, artificial intelligence which can "make restaurant reservations, schedule hair salon appointments, and get holiday hours over the phone[55]";
- (2018) The first artificial intelligence generated painting sold for $432,500;[56]
- (2018) The first fully self-driving taxi service, Waymo One.[57]

Today, artificial intelligence is used in many aspects of life.[58] Experts claim that artificial intelligence made remarkable success in specific fields, most particularly in gaming, translation, autonomous vehicles, and image recognition.[59]

The future of artificial intelligence is an uncertain, complex, and polarizing topic. Over the past sixty years, there has been a dramatic increase in the competences of artificial intelligence, most particularly weak artificial intelligence showed a substantial advancement. As exemplified above, weak artificial intelligence already outperforms humans in specific fields. There is a good probability that weak artificial intelligence will dominate more domains, which in turn, may transform society. It may result in unique opportunities and challenges that society cannot foresee at present.

The leading and perplexing question regarding the future of strong artificial intelligence is whether it will be able to surpass human intelligence. In other words, the question is whether science will be able to build strong artificial intelligence. An expert survey indicates that more than 50% of experts believe that strong artificial intelligence may reach human capacity between 2040 and 2050; by 2075, the chance increases to 90%. Experts assert that there is a 31% probability that this development may lead to "bad" or "extremely bad" for humanity.[60] However, considering the survey in question was conducted among passionate researchers in

[54] (Linn, 2018)

[55] (Leviathan, 2018)

[56] (Falcon, 2018)

[57] (Waymo Team, 2018), (Hawkins, 2018). Technically, Waymo One falls under level 4 self-driving car unlike fully autonomous level 5 self-driving car. Further, it has been observed that other companies started to provide autonomous taxi services in 2018. However, many examples are in small scales in a limited environment and includes a safety driver.

[58] Autonomous planning and scheduling, spam (e-mail) fighting, logistics planning, robotics, medical diagnosis and treatment, traffic management, criminal justice system, education, scientific research, trip planning, shopping recommendation, ad targeting are some examples. (Technology, 2016, pp. 7-16), (Norvig & Russell, 2010, pp. 28-29)

[59] (Technology, 2016, p. 7)

[60] (Bostrom & Müller, 2016)

the field of artificial intelligence, one could argue that it includes self-selection and confirmation bias.

There is a good chance that artificial intelligence will become more critical. Sundar Pichai, Google's chief executive officer (CEO), argues that artificial intelligence is "one of the most important things that humanity is working on" that could be "more profound than electricity and fire.[61]" Yuval Noah Harari, a leading history researcher, believes that

> "…those countries who lead the world in AI are likely to lead the world in all economic and political terms. It could be a rerun of the industrial revolution of the 19th century when you had a few countries, first Britain then Germany and France and the US and Japan, who were pioneers in industrialization. These few countries conquered, dominated, and exploited the world. This is very likely to happen again on an even larger scale with AI and biotechnology in the 21st century … The gap between those who control AI and biotechnology and those who are left behind is likely to be far greater than the gap between those who developed steam engines in the 19th century and those who did not.[62]"

One may question above-stated bold arguments. In this case, it could be useful to remember Amara's law. Roy Amara, the American futurist, warns that "[w]e tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.[63]" This may describe the current common perspective regarding artificial intelligence. There may be an explanation of this state of mind. As mentioned above, the progress of artificial intelligence showed fluctuation and came in waves. The rapid advancement cycles, one of them is still ongoing since 2010, unreasonably raised (especially short-term) expectations. Moreover, one may argue that popular culture shows artificial intelligence in a techno chauvinistic way that leads to extravagant anticipation in society.

2.3. Conclusions

The second chapter of the thesis has concluded that artificial intelligence remains a nebulous term since 1) it is burdensome to define intelligence which makes it challenging describing artificial intelligence, 2) there are various approaches to artificial intelligence, 3) artificial intelligence is a very diversified field.

The findings of the second chapter draw the conclusion that artificial intelligence is a relatively new, quickly blooming, and multidisciplinary field of study. Over the past few decades, the development of artificial intelligence showed fluctuations. Since 2010, the field showed a

[61] (Romm, Timberg, & Harwell, Google CEO Sundar Pichai: Fears about artificial intelligence are 'very legitimate,' he says in Post interview, 2018)
[62] (Springer & Döpfner, 2018)
[63] (Ratcliffe, 2018)

dramatic shift as a result of big data and affordable computing power, and today artificial intelligence is used in diverse areas. The future of artificial intelligence (and its competence) is a disputed issue among experts; therefore, this thesis does not attempt to consider the long-term impact of artificial intelligence.

CHAPTER 3: PROHIBITION OF DISCRIMINATION

The chapter aims to provide a general overview of the prohibition of discrimination, one of the main pillars of the thesis. This chapter firstly discusses the definition of discrimination, after which discrimination types follow it and thenceforth discrimination grounds are touched upon. It should be noted that this chapter is descriptive and aims to provide necessary information on the prohibition of discrimination so that readers can easily comprehend artificial intelligence discrimination in the following chapters.

3.1. What is Discrimination?

This section aims to provide a basic understanding of discrimination, a concept that lies at the heart of the thesis. Discrimination law is a multidisciplinary field. It has strong ties with constitutional and human rights law, its origins are connected to employment law, whereas tort law is closely associated with discrimination law.[64] This chapter generally draws upon human rights law.

Daily life includes a large number of choices: people decide whom to socialize with, where to go, what to eat, and so on. In other words, people "discriminate" all the time, and it often falls out of discrimination law's scope. Discrimination law is concerned with choices that may treat an individual less advantageously compared to others due to a "morally irrelevant" consideration. Beyond doubt, what is considered morally unacceptable is convoluted.[65] Also, not every instance of differential treatment can be considered as discriminatory. A differential treatment that is proportionate and pursues a legitimate aim may not be discriminatory.[66]

To establish the importance of the prohibition of discrimination, it may be useful to refer to the foundation of human rights. All members of the human family are entitled to the same rights and freedoms. Prohibition of discrimination is an extension of such a mentality. As noted by the United Nations Human Rights Committee, "[n]on-discrimination, together with equality before the law and equal protection of the law without any discrimination, constitute a basic and general principle relating to the protection of human rights.[67]" The United Nations

[64] (Khaitan, 2015, p. 24)
[65] (Evelyn & Watson, 2012, p. 2)
[66] (Moeckli, Shah, & Sivakumaran, 2014, p. 167)
[67] (Committee, 1989, p. 1)

Committee on Economic, Social and Cultural Rights views non-discrimination as "essential to the exercise and enjoyment of economic, social and cultural rights.[68]"

Despite its importance, the prohibition of discrimination was recognized in laws only relatively recently. The prohibition of direct discrimination in the USA and India can be traced back to the 19th century[69], yet the realization of it occurred during the mid-20th century, mainly thanks to the civil rights movement. Indirect discrimination was prohibited firstly in the USA (in 1971).[70] The USA also was the first country that banned discriminatory harassment (in 1976).[71] Another cornerstone in history was the affirmative action[72] that was first embraced in India.[73] It is observed that the development of discrimination law in various jurisdictions generally interconnected with one particular ground of discrimination, "race in the United States, caste in India, sex in the European Union.[74]"

"Discrimination" is a term frequently used in legal literature. But to date, there is no consensus about the exact meaning of it. Meanings attributed to discrimination can differ significantly due to three factors: time, place, and world views. From a political perspective, different approaches to discrimination (law) are broadly classified into three categories: egalitarian, liberal, and dignitarian.[75] To define discrimination, it might be useful to consider zeitgeist. At present (the age of rising populism and nationalism), discrimination may convey a different meaning compared to its understanding during the civil rights movement. Equally important, space is another critical factor regarding the meaning of discrimination. Put differently, what is deemed discriminatory could vary depending on the region. Bearing all these factors in mind, perhaps the expectation of discrimination law should be kept at a reasonable level. Because without political, social, and economic progress, discrimination laws may fall short to be a remedy.[76]

[68] (United Nations Committee on Economic, General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights), 2009, p. 1)

[69] Indian Caste Disabilities Removal Act 1850, 14th Amendment to the US Constitution 1868

[70] (Griggs et al. v. Duke Power Co., 1971)

[71] (Williams v. Saxbe, 1976)

[72] Researchers use numerous terms to describe affirmative action, the most common of which are positive discrimination, positive action, reverse discrimination, compensatory discrimination. (Khaitan, 2015, p. 68)

[73] (Khaitan, 2015, p. 16)

[74] (Khaitan, 2015, p. 48)

[75] Egalitarians value equality, liberals prioritize liberty, autonomy, or freedom, and dignitarians give particular importance to personal dignity. (Khaitan, 2015, pp. 6-7)

[76] (Evelyn & Watson, 2012, p. 8)

Having indicated the difficulties of defining discrimination, the thesis will try to provide a working description by the European Union Agency for Fundamental Rights. The term "discrimination" can broadly be defined as "a situation where an individual is disadvantaged in some way on the basis of 'one or multiple protected grounds.[77]'" Turning to discrimination law, it has been reported that in order for a duty-imposing legal norm to become a discrimination law norm, it should meet four criteria: 1) personal grounds, 2) cognate groups, 3) relative disadvantage, 4) eccentric distribution. These criteria described below.

> "The Personal Grounds Condition: The duty-imposing norm in question must require some connection between the act or omission prohibited or mandated by the norm on the one hand and certain attributes or characteristics that persons have, called 'grounds,' on the other.
>
> The Cognate Groups Condition: A protected ground must be capable of classifying persons into more than one class of persons, loosely called 'groups.'
>
> The Relative Disadvantage Condition: Of all groups defined by a given universal order ground, members of at least one group must be significantly more likely to suffer abiding, pervasive, and substantial disadvantage than the members of at least one other cognate group.
>
> The Eccentric Distribution Condition: The duty-imposing norm must be designed such that it is likely to distribute the non-remote tangible benefits in question to some, but not all, members of the intended beneficiary group.[78]"

To implement the prohibition of discrimination principle, it is necessary to establish who are alike. Because less favourable treatment can be identified when someone under similar circumstances receives more favorable treatment. Consequently, an essential concept related to discrimination is a comparator. The term refers to an individual (or group) who is in a similar situation with discrimination victim(s). The difference between comparator and discriminated person (or group) is discrimination ground(s). Put differently, the main difference between those less and those more advantageously treated is the protected ground(s). To identify an appropriate (actual or hypothetical) comparator sometimes may turn into a complex issue.

3.2. Discrimination Types

This section seeks to briefly define the discrimination types that may assist the reader (in the following sections) in comprehending how artificial intelligence can discriminate. The section intends to define respectively: direct, indirect, multiple, intersectional, and systemic discrimination.

[77] (Europe, 2018, p. 42)
[78] (Khaitan, 2015, p. 42)

Direct discrimination: Direct discrimination “entails unfavourable or less favourable treatment ‘on the ground of’ a protected characteristic or, sometimes, a combination of such characteristics.[79]” Another way to define the term is “when a person is treated less favourably on the basis of ‘protected grounds.[80]’” Taking a closer look at this definitions lead to two criteria: the existence of an act or omission and a causal nexus between the treatment and a protected ground.[81]

There are different approaches regarding how to establish the nexus between treatment and protected ground. In the USA, courts seek a discriminatory motive, purpose, or intention behind the treatment. In Canada, the United Kingdom, and South Africa, there is no such precondition and courts focus more on the outcome of treatment.[82] Also, it should be underlined that direct discrimination can occur without a comparable circumstance or comparator.[83]

Lastly, it is safe to say that direct discrimination focuses on the individual. Today, direct discrimination is a far-reaching problem, and its leading effects are prejudice and racial stereotyping.[84]

Indirect discrimination: The term indirect discrimination (disparate impact as referred in the USA) is used to refer to a seemingly neutral practice or policy “which puts persons belonging to a protected group at a particular disadvantage.[85]” A further definition by the European Court of Human Rights (ECtHR) describes the term as “a difference in treatment may take the form of disproportionately prejudicial effects of a general policy or measure which, though couched in neutral terms, discriminates against a group.[86]” It is noteworthy that the concept of indirect discrimination firstly enacted by the Supreme Court of the United States in *Griggs v. Duke Power Company*.[87]

These definitions lead to three elements: formally equal treatment, disparate outcome, and the absence of justification of different treatment. Notably, the first criteria distinguish direct and indirect discrimination. In other words, if the treatment is not formally equal, then it falls into

[79] (Khaitan, 2015, p. 69)
[80] (Europe, 2018, p. 43)
[81] Protected discrimination ground will be analyzed in Section 3.3.
[82] (Khaitan, 2015, pp. 69-71), (Moeckli, Shah, & Sivakumaran, 2014, p. 166)
[83] (United Nations Committee on Economic, General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights), 2009, p. 4)
[84] (Fredman, 2001, p. 24)
[85] (Khaitan, 2015, p. 73)
[86] (Biao v. Denmark, 2016, p. 103)
[87] (Griggs et al. v. Duke Power Co., 1971)

the scope of direct discrimination.[88] In circumstances of indirect discrimination, there is no difference in treatment. Identical treatment results in unequal outcomes due to structural biases.[89] In other words, the prohibition of indirect discrimination indicates that neutral practices or policies may favor dominant norms. Therefore, the prohibition of indirect discrimination is essential for multiculturalism and assists in the accommodation of diversity.[90] Additionally, unlike direct discrimination, indirect discrimination concentrates more on groups.[91]

Multiple discrimination: An individual has many characteristics that are recognized as discrimination grounds. In the complex world, unfair treatment can occur on more than one discrimination ground. It has been recognized by discrimination law. The term multiple discrimination means "discrimination takes place on the basis of several grounds operating separately.[92]" Here, discrimination occurs on several grounds, and each of unequal treatment is based on a different ground and separately meets the threshold of discriminatory treatment. Therefore, multiple discrimination is also referred to as cumulative or additive discrimination.[93]

Intersectional discrimination: The concept defined as "a situation where several grounds operate and interact with each other at the same time in such a way that they are inseparable and produce specific types of discrimination.[94]" A classic example of intersectional discrimination is a black woman being discriminated, not because she is Black or a woman, but because she is a "black woman.[95]"

Systemic discrimination: The widespread use of the term systemic discrimination is equated with discrimination against specific social groups which is pervasive, persistent, established in social behaviour and organization, and generally includes unchallenged or indirect discrimination. This type of discrimination may be seen in laws, policies, procedures, cultural mindset, and both in the public and private sphere.[96] In other words, rather than a single and unequal treatment, systemic discrimination is systemic or institutionalized unequal treatment.

[88] (Hacker, 2018, p. 10)

[89] (Moeckli, Shah, & Sivakumaran, 2014, p. 165)

[90] (Fredman, 2001, p. 24)

[91] (Hacker, 2018, p. 10)

[92] (Europe, 2018, p. 59)

[93] (Schiek, Waddington, & Bell, 2007, p. 171)

[94] (Europe, 2018, p. 59)

[95] (Schiek, Waddington, & Bell, 2007, p. 171), (Kimberle, 1989)

[96] (United Nations Committee on Economic, General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights), 2009, p. 5)

Harassment: In the laws, the term tends to be used to refer to an "unwanted conduct related to a relevant protected characteristic and, the conduct has the purpose or effect of violating [the victim's] dignity, or creating an intimidating, hostile, degrading, humiliating or offensive environment for [the victim].[97]" This type of discrimination is separated from other types due to its significant harm.[98]

3.3. Protected Grounds

Discrimination law does not ban each different treatment. A treatment may fall into the scope of discrimination law only if a different treatment is based on "protected grounds." In other words, "[t]o discriminate on no basis is simply to not discriminate.[99]" As a result, discrimination law safeguards "groups of persons defined by certain personal characteristics that are technically called grounds.[100]" In order for personal characteristics to become a protected ground, it should fulfill two conditions: 1) the ground shall classify "persons into groups with a significant advantage gap between them," 2) the ground shall be unchangeable characteristics or form a fundamental choice.[101]

The protected grounds list has evolved significantly over the years. The protected grounds may vary depending on the jurisdiction, yet it is claimed that race, sex, and religion are extensively recognized as protected grounds. Arguably, the prohibition of discrimination based on these grounds is a part of customary international law. The Inter-American Court of Human Rights took a further step and stated that prohibition of discrimination (based on any ground) is *a jus cogens* norm.[102]

The protected ground list is not limited in international human rights law. For instance, the ECtHR does not fix protected grounds and considers each case individually, which results in a non-exhaustive list of protected grounds.[103] This is an extension of the European Convention on Human Rights (ECHR) since the Convention prohibits discrimination based "on any ground such as sex, race, … ***or other status***[104]" (emphasis added). Thanks to this perspective, the scope of freedom from discrimination has been broadened over the years in the Council of Europe.

[97] UK Equality Act 2010, Section 26
[98] (Europe, 2018, p. 64)
[99] (Eidelson, 2015, s. 16)
[100] (Khaitan, 2015, p. 49)
[101] (Khaitan, 2015, p. 50)
[102] (Moeckli, Shah, & Sivakumaran, 2014, p. 161)
[103] (Europe, 2018, p. 160)
[104] ECHR, Article 14

The following paragraphs will define some of the most common grounds of discrimination in the light of international human rights law or European law. It should be noted that it is a non-exhaustive list, and some grounds are interconnected and overlap conceptually. Further, due to practical constraints, the thesis cannot provide a comprehensive review of all grounds. Therefore, some of the definitions may be overly simplistic.

Sex: Sex discrimination occurs when an individual is treated unequally based on their sex. In this case, a man or woman receives less favorable treatment compared to the other sex.[105] The gender wage gap, pregnancy, and maternity-related issues are classic examples of sex discrimination. It is necessary to mention that intersectional discrimination frequently takes place as a combination of discrimination based on sex and another ground(s).

Gender identity: Discrimination based on gender identity[106] is prohibited by international human rights laws[107] and occurs when an individual receives uneven treatment based on gender identity. Limitations on access to gender reassignment and gender recognition process, and registration of sex at birth are predominant examples of such discrimination.[108] Here, it may be useful to clarify the distinction between sex and gender. Sex is generally considered as a part of biology; gender is more connected to social reality. Sometimes it may be burdensome to differentiate these concepts since they are interconnected.[109]

Sexual orientation: Discrimination based on sexual orientation[110] occurs when an individual receives less favorable treatment on the ground of sexual orientation. Sexual orientation discrimination can often be seen in the context of recruitment, employment, or sexual prejudice.

[105] (Europe, 2018, p. 162)

[106] "Gender identity is understood to refer to each person's deeply felt internal and individual experience of gender, which may or may not correspond with the sex assigned at birth, including the personal sense of the body (which may involve, if freely chosen, modification of bodily appearance or function by medical, surgical or other means) and other expressions of gender, including dress, speech and mannerisms." (Yogyakarta Principles, 2007, p. 6)

[107] Istanbul Convention Article 4 and ECHR Article 14. In the view of ECtHR, ECHR Article 14 covers discrimination based on gender identity. (Identoba and Others v. Georgia, 2015)

[108] For detailed information on discrimination based on gender identity: (European Union Agency for Fundamental Rights, 2015)

[109] (Schiek, Waddington, & Bell, 2007, p. 70)

[110] "Sexual orientation is understood to refer to each person's capacity for profound emotional, affectional and sexual attraction to, and intimate and sexual relations with, individuals of a different gender or the same gender or more than one gender." (Yogyakarta Principles, 2007, p. 6)

Disability: A (widely accepted[111]) discrimination ground is of disability. And a generally accepted definition of discrimination based on disability refers to

> "any distinction, exclusion or restriction on the basis of disability which has the purpose or effect of impairing or nullifying the recognition, enjoyment or exercise, on an equal basis with others, of all human rights and fundamental freedoms in the political, economic, social, cultural, civil or any other field. It includes all forms of discrimination, including denial of reasonable accommodation.[112]"

Age: Age discrimination is a type of discrimination that may arise in diversified frameworks and relates to unequal treatment on the ground of an individual's age.[113] Recruitment and retirement based on age can set examples of age discrimination. For instance, the European Committee of Social Rights ruled that employment termination based only on age (without conducting the company's operational requirement or individual's capacity) constituted discrimination.[114] It should be noted that the prohibition of age discrimination aims to safeguard both old and young.[115]

Race: Race is a protected ground in discrimination law. And racial discrimination is defined as

> "any distinction, exclusion, restriction or preference based on race, colour, descent, or national or ethnic origin which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise, on an equal footing, of human rights and fundamental freedoms in the political, economic, social, cultural or any other field of public life.[116]"

Ethnicity: The Court of Justice of the European Union defined ethnicity as "its origin in the idea of societal groups marked in particular by common nationality, religious faith, language, cultural and traditional origins and backgrounds.[117]" According to the ECtHR,

> "[e]thnicity and race are related and overlapping concepts. Whereas the notion of race is rooted in the idea of biological classification of human beings into subspecies according to morphological features such as skin colour or facial characteristics, ethnicity has its origin in the idea of societal groups marked by common nationality, tribal affiliation, religious faith, shared language, or cultural and traditional origins and backgrounds.[118]"

[111] As the date of February 2019, there are 172 parties and 162 signatories of the Convention on the Rights of Persons with Disabilities. (United Nations, 2019)

[112] Convention on the Rights of Persons with Disabilities, Article 2

[113] (Europe, 2018, p. 190)

[114] (Fellesforbundet for Sjøfolk (FFFS) v. Norway, 2013)

[115] (Schiek, Waddington, & Bell, 2007, p. 148)

[116] International Convention on the Elimination of All Forms of Racial Discrimination, Article 1.1.

[117] (CHEZ Razpredelenie Bulgaria" AD v. Komisia za zashtita ot diskriminatsia (GC), 2015)

[118] (Timishev v. Russia, 2005)

The same perspective can be seen in the International Convention on the Elimination of All Forms of Racial Discrimination. According to the Convention, discrimination based on ethnic origin is prohibited, and a form of racial discrimination.[119]

Nationality and national origin: A commonly accepted definition of nationality refers to "a legal bond [between a person and state] having as its basis a social fact of attachment, a genuine connection of existence, interests and sentiments, together with the existence of reciprocal rights and duties.[120]" National origin is used to address a person's (lost or added) former nationality.[121] Discrimination based on these grounds is prohibited in international human rights law.[122]

Religion or belief: Religion may be somewhat an ambiguous legal term, yet there is a consensus that discrimination on the ground of religion is prohibited. Neither ECHR Article 9 nor the Strasbourg Court's case-law defines religion. If a major, minor, old, new, theistic, or nontheistic religions, non-religious opinions, or convictions "attain a certain level of cogency, seriousness, cohesion, and importance," they may fall into the scope of ECHR Article 9. Discrimination based on religion is prohibited under international law[123] , and the ECtHR often combine the prohibition of discrimination with freedom of thought, conscience, and religion.[124]

Political or other opinions: Discriminatory treatment can be based on holding or not holding opinions, declaring opinions, or belonging to an association, political party or trade union. Reaffirming the importance of these for democratic society, political or other opinion is a protected discrimination ground.[125]

Social origin, birth, and property: Inherited social, economic, or biological characteristics can lead to discrimination. It has been seen in the case of unequal treatment on the grounds of birth out of wedlock, adopted by or born of stateless parents, to not be a member of caste or similar inherited status, or to belong a particular social or economic group, such as poverty and

[119] International Convention on the Elimination of All Forms of Racial Discrimination, Article 1,2

[120] (Nottebohm (Liechtenstein v. Guatemala), 1955)

[121] (Europe, 2018, p. 202)

[122] International Convention on the Elimination of All Forms of Racial Discrimination, Article 1,2

[123] International Convention on the Elimination of All Forms of Racial Discrimination, Article 5

[124] (European Court of Human Rights, 2018, pp. 6-9), (Europe, 2018, pp. 210-215)

[125] (Europe, 2018, p. 222), (United Nations Committee on Economic, General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights), 2009, p. 7)

homelessness.[126] Regarding property as a discrimination ground, the concept may refer to an individual's connection to real property, personal property, or the lack of it. Also, regardless of an individual's tenure status, one shall have the right to adequate housing and right to water.[127] With this, discrimination based on social origin, birth, and property is prohibited in international human rights law.

Language: It is questionable that discrimination based on language is a separate discrimination ground. On the other hand, discrimination on the ground of language is closely related to race or ethnicity, widely accepted discrimination grounds. Therefore, there is a consensus on discrimination on the ground of language as being prohibited.[128] It is certainly true in the case of when language barriers undermine the enjoyment of human rights. To illustrate, public service language can hinder to receive public service, public education language can undermine minority rights, or criminal proceedings language (without the assistance of a translator) can impede the right to a fair trial.

Other status: As indicated previously, the grounds of prohibited discrimination is not exhaustive, and many international human rights treaties[129] include an "other status." In the view of the ECtHR, "other status" in ECHR can be loosely described as "differences based on an identifiable, objective, or personal characteristic, or 'status,' by which individuals or groups are distinguishable from one another.[130]" Thanks to the other status and the court's living instrument approach, the ECtHR recognized sexual orientation, age, and disability as protected grounds. The other status also allowed a large number of personal characteristics or statuses[131] to be safeguarded from discrimination.[132]

[126] (Europe, 2018, p. 218), (United Nations Committee on Economic, General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights), 2009, pp. 7-8)

[127] (United Nations Committee on Economic, General Comment No. 15: The Right to Water (Arts. 11 and 12 of the Covenant), 2003, pp. 5-7), (United Nations Committee on Economic, General Comment No. 4: The Right to Adequate Housing (Art. 11 (1) of the Covenant), 1991, pp. 2,4,6)

[128] (Europe, 2018, pp. 218-219) (United Nations Committee on Economic, General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights), 2009, p. 7)

[129] International Covenant on Civil and Political Rights Article 2, International Covenant on Economic, Social and Cultural Rights Article 2, ECHR Article 14, African Charter on Human and Peoples' Rights Article 2, Arab Charter on Human rights Article 2

[130] (Novruk and Others v. Russia, 2016)

[131] Fatherhood, marital status, membership of an organization, military rank, parenthood of a child born out of wedlock, place of residence, health or any medical condition, former KGB officer status, retirees employed in certain categories of the public sector, detainees pending trial are some of the examples.

[132] (Europe, 2018, pp. 224-225)

3.4. Conclusions

The chapter glanced through the prohibition of discrimination in the light of international human rights law, regional, and national discrimination laws. The findings indicate that there is no straightforward definition of discrimination, yet there is a consensus on the prohibition of discrimination which sets ground for equality and pertains great importance for the enjoyment of a wide range of human rights. Thus, the prohibition of discrimination is under the protection of international human rights law, and its scope becomes more inclusive owing to progressive interpretation of laws. On the other hand, discrimination is a widespread and critical problem all over the world, and some jurisdictions are unwilling to accept some discrimination grounds. Additionally, it should be kept in mind that the effective implementation of discrimination laws is closely related to economic, historical, cultural, and political factors.

CHAPTER 4: ARTIFICIAL INTELLIGENCE AND PROHIBITION OF DISCRIMINATION

> “The math-powered applications powering the data economy were based on choices made by fallible human beings. Some of these choices were no doubt made with the best intentions. Nevertheless, many of these models encoded human prejudice, misunderstanding, and bias into the software systems that increasingly managed our lives. Like gods, these mathematical models were opaque, their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists. Their verdicts, even when wrong or harmful, were beyond dispute or appeal. And they tended to punish the poor and the oppressed in our society, while making the rich richer.[133]”

This chapter tests the hypothesis of this thesis, artificial intelligence discrimination. The spotlight will be on why, how, and who artificial intelligence discriminates. The chapter firstly focuses on why and how artificial intelligence discriminates. In the following sections, the chapter bridges the connection between artificial intelligence and discrimination grounds.

4.1. Establishing the Nexus

The primary purpose of this section is to develop an understanding of the connection between artificial intelligence and discrimination. The section aims to examine how artificial intelligence can discriminate. It has been reported that, in theory, artificial intelligence can generate discrimination in six ways. Firstly, the definition of target variable and class labels can cause discrimination. Also, sampling or historical bias can result in a disparate impact. Furthermore, discrimination can be stemming from feature selection or proxy discrimination. Lastly, decision makers can use artificial intelligence as a tool to discriminate. The following headings will shed light on these.

4.1.1. Discriminatory target variable and class labels

To enable comprehension of how artificial intelligence can discriminate, there is a need to define machine learning and data mining terms, target variable and class labels. This is because of how these concepts are defined in the development of artificial intelligence may lead to discrimination.

In simple terms, “target variable”[134] is or should be the outcome of the data mining, what a trained model aspires to provide as output. “Class labels” is simply defined as values that are

[133] (O'Neil, 2016, p. Introduction)

[134] To give a detailed definition of target variable, spam email fighting can provide a context. Artificial intelligence is useful to describe and distinguish a large volume of data. For instance, artificial intelligence is used to detect spam emails and to do so, it needs to learn the difference between spam and non-spam emails. Humans label emails as spam and non-spam and feed artificial intelligence by this information. In the next step, artificial intelligence finds patterns in emails labelled as spam. To illustrate, spam emails include certain expressions such

related to the goal of the data mining process. Class labels aim to "divide all possible values of the target variable into mutually exclusive categories.[135]" Put differently, "[c]lass label is the discrete attribute having finite values (dependent variable) whose value you want to predict based on the values of other attributes (features).[136]"

To clarify these technical terms and how they can lead to discrimination, spam email fighting (one of the most successful artificial intelligence capabilities) will be analyzed. Spam email fighting aims to detect spam emails and automatically move them into junk folder. In the context of spam email fighting, class labels are divided into two, spam emails and non-spam (honest) emails. Since class labels are well-understood and uncontroversial, the algorithm works efficiently in spam email fighting. However, in complicated cases, defining class labels and target variable can be challenging. For example, nowadays, the recruitment industry uses artificial intelligence to expedite employment processes, headhunting, and conduct automated interviews.[137] It is argued that use of artificial intelligence in such a process can result in discrimination due to two reasons. Firstly, it is very burdensome to measure what constitutes a good employee. Therefore, it is controversial how to define class labels. Secondly, the issue gets more complicated because to understand a convoluted problem (what a good employee is) and transform it into a data mining problem is delicate. Any mistake in that process may generate discriminatory treatment. As a result, experts argue that how class labels and target variable are defined, may lead to discrimination.[138]

4.1.2. Biased training data (historical bias)

Artificial intelligence can learn from data. For example, to design an image-recognition algorithm that can recognize cats, an algorithm needs to learn the cat's appearance. To train the algorithm, a large number of cat photos should be presented to it. In the next stage, the algorithm finds relationships and detects patterns in cat photos to learn a cat's image. The data that is used to train artificial intelligence (cat photos), and thus is referred to as training data. Also, training data teaches an algorithm to act in a specific way.

as "donation to you" or "money transfer to your bank account" and artificial intelligence can detect this pattern. "[B]y exposing so-called 'machine learning' algorithms to examples of the cases of interest (previously identified instances of spam) the algorithm 'learns' which related attributes or activities can serve as potential proxies for those qualities or outcomes of interest.' Such an outcome of interest is called a 'target variable.'" (Borgesius, 2018, p. 10)

[135] (Borgesius, 2018, p. 10)

[136] (Bloch, 2018, p. 93)

[137] (Castellanos, 2019), (IBM, 2019)

[138] (Borgesius, 2018, pp. 17-18), (Barocas & Selbst, 2016, pp. 677-680)

The quality and quantity of training data is essential to design successful artificial intelligence because input determines the output. In computer science, the problematic outcome of erroneous training data is referred to as "garbage in, garbage out." It may lead to many problems, including artificial intelligence bias. To exemplify this, a medical school in the United Kingdom developed a computer program to consider applications. The training data was based on previous admissions. In the past, the school disfavored eligible women and racial minority applicants. In other words, the training data was biased. The computer program established admission criteria based on biased data and continued to discriminate qualified women and members of racial minority applicants. It was not a new bias; instead, it was an extension of biased training data (historical bias).[139]

Another example of biased training data is Amazon's recruiting artificial intelligence. In 2014 the company built a computer program to review job applications. A year later, Amazon realized that the program discriminated against women applicants for software developer and technical positions. Because the training data was submitted resumes from the last decade, which mostly came from male applicants. Essentially, the training data was gender biased. As a result, the system self-taught that applicants using masculine language were favorable candidates, and it unfairly preferred male applicants over females.[140]

4.1.3. Biased or insufficient training data collection (sampling bias)

"The best material model for a cat is another, or preferably the same cat.[141]"

Data is not created or collected equally in society. Generally, less data is available about minorities, as some live outside of big data borders, and others cannot create data due to the digital divide. Sometimes developers underrate some groups because they financially consume less compared to the other groups. It makes some groups less valuable data subjects. All these may lead to the underrepresentation of some groups.

An essential factor in the success of artificial intelligence is training data. Therefore, training data collection process can generate a dramatical impact on the output of artificial intelligence. If training data collection is unsatisfactory, e.g., training data is incorrect, partial, or nonrepresentative, artificial intelligence may reproduce or amplify discrimination. It is especially true when the quality and representativeness of training data is correlated with a

[139] (Barocas & Selbst, 2016, pp. 681-684), (Borgesius, 2018, p. 11)
[140] (Dastin, 2018)
[141] (Wiener & Rosenblueth, 1945, p. 320)

discrimination ground.[142] For example, a study found that facial recognition systems developed in the USA and Western Europe were more successful in identifying Caucasians; whereas facial recognition systems developed in East Asia performed better on East Asians.[143] Possibly, the reason behind it was biased data collection.

In *Ewert v. Canada*, the Supreme Court of Canada noted the sampling bias problem. Mr. Ewert is an inmate serving a life sentence and identifies himself as Métis (one of Canada's aboriginal peoples). Meanwhile, the Correctional Service of Canada, the institution which is responsible for running prisons, uses tools to assess the recidivism risk and mental health of offenders. Mr. Ewert argued that the institution relied on "tools had been developed and tested on predominantly non-indigenous populations and that there was no research confirming that they were valid when applied to indigenous persons.[144]" Ergo, he claimed that the Correctional Service did not "take all reasonable steps to ensure that any information about an offender that it uses is as accurate, up to date and complete as possible" as is required by law.[145] The Supreme Court agreed with the plaintiff and found that the Correctional Service of Canada did not take reasonable steps to ensure that assessment tools are free from bias; accordingly, it violated the laws.[146]

Another critical problem should be considered, which is the overrepresentation of particular groups in training data collection. A classic example may occur during crime data collection. It is maintained that if police focus on particular groups or districts, police records may overrepresent crime statistics in that group or area. For example, Swedish police implemented a project called rättsäkert och effektivt verkställighetsarbete (legal and effective execution of policy) to deport persons residing in Sweden without authorization. An aspect of the project was identity document checking in public transportation. It is argued that Swedish police disproportionately targeted non-white Swedes who generally reside in segregated suburbs.[147] It is plausible to argue that such practice may result in biased crime statistics, an overrepresentation of non-white Swedes in the database. If such biased data is used to train artificial intelligence, the model may reproduce the bias. More importantly, the problem may create a loop. Due to overrepresentation, the model indicates that a specific group commits

[142] (Borgesius, 2018, pp. 11-12), (Barocas & Selbst, 2016, pp. 684-686)
[143] (Garvie, Frankle, & Bedoya, 2016, p. 53)
[144] (Ewert v. Canada, 2018)
[145] (Ewert v. Canada, 2018)
[146] (Ewert v. Canada, 2018)
[147] (The Local, 2013), (The Local, 2013), (Barker, 2016, pp. 19-21)

more crime or a particular area is a crime hotspot. Consequently, police tend to concentrate more on that group or area.[148] It is noteworthy to mention that while raising her concerns on racial profiling in Europe, the Council of Europe Commissioner for Human Rights warned that artificial intelligence could worsen the problem if necessary precautions are not taken.[149]

4.1.4. Feature selection

"Essentially all models are wrong, but some are useful.[150]"

Artificial intelligence can be used to find a solution to complex scenarios. However, to capture every dimension of an issue can be burdensome. For this reason, developers simplify a scenario and reduce dimensions of it by choosing a set of relevant features that represent the matter. Such a procedure is defined as a feature selection that includes to identify relevant data and remove irrelevant, redundant, or noisy data.[151]

Experts acknowledge that a model is a sample, a slight advancement in the comprehensiveness of a model can be costly, and it may even be impossible to build a completely comprehensive model. On the other hand, feature selection may generate discriminatory treatment on protected grounds "because the details necessary to achieve equally accurate determinations reside at a level of granularity and coverage that the selected features fail to achieve.[152]"

4.1.5. Proxy discrimination

To make predictions, artificial intelligence needs training data that may include sensitive personal data, such as race, ethnicity, political opinions, et cetera. As sensitive personal data processing poses high risks to human rights and may generate discrimination, laws require special procedures for sensitive data processing.[153] Another approach to mitigate this potential discrimination is that developers can remove sensitive personal data from training data whenever possible. It is called "fairness through unawareness."[154] This attempt may fall short due to "redundant encodings." This term refers to when a model includes one variable that can be a proxy for another variable which should not be involved in the model.[155] To illustrate this,

[148] (Borgesius, 2018, pp. 11-12)
[149] (Dunja, 2019)
[150] (Box & Draper, 1987, p. 424)
[151] (Norvig & Russell, 2010, p. 713), (Kumar & Minz, 2014, p. 211)
[152] (Barocas & Selbst, 2016, p. 688)
[153] General Data Protection Regulation Article 4, 6, 9, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data Article 6
[154] (Hardt, 2016)
[155] (Purcell B. , 2018, p. 3)

in the USA, 82.2 percent of custodial parents are females. In other words, there is a high correlation between custody and gender. In a decision-making process which should not take gender into consideration, custody status may serve as a proxy for gender due to the high correlation between them.[156]

Experts argue that whenever redundant encodings can provide information about a protected discrimination ground, it may generate discrimination. It can be illustrated by the phenomena of redlining.[157] For example, Amazon, the world's largest e-commerce marketplace, provides a service named Amazon Prime that guarantees two days (sometimes same-day) delivery.[158] In 2016, almost a year after the service started, Amazon Prime covered twenty-seven metropolitan areas in the USA. The company's algorithms took a cost and efficiency viewpoint to choose the borders of the service. Amazon argued that demographics played no role in service delivery. On the other hand, "[i]n six major same-day delivery cities, however, the service area excludes predominantly black ZIP codes … cities still struggling to overcome generations of racial segregation and economic inequality, black citizens are about half as likely to live in neighborhoods with access to Amazon same-day delivery as white residents.[159]" For instance, in Atlanta, 96% of white residents had access to same-day delivery as compared with 41% of black residents; in Chicago same-day delivery was available to 98% white residents compared to 54% black residents.[160]

When a decision maker lacks information on an individual's connection to a protected discrimination ground, proxies can supply accurate information about the connection. For instance, one of the largest retail stores in the USA, Target Corporation, uses a pregnancy prediction tool. It aims to identify customers in the early stages of pregnancy by assessing their shopping behavior. The company uses the output for targeted advertising and to generate brand loyalty. If job applicants or employees shop from Target, the company can use the same system to discriminate against early stage pregnant applicants or employees.[161]

[156] (Purcell B. , 2018, p. 3)
[157] In broad terms, the term is defined as the practice of not providing services, generally to provide insurance or bank loan, to a particular geographical area due to race or ethnicity of residents.
[158] Despite it seems as a luxury service at first sight, one should consider that in the USA, Amazon is an everything store that offers lower prices compare to many physical stores. Therefore, the service may matter to people who do not have a car and live in an area where public transportation is underdeveloped.
[159] (Ingold & Soper, 2016)
[160] (Ingold & Soper, 2016)
[161] (Barocas & Selbst, 2016, pp. 692-693), (Borgesius, 2018, pp. 13-14)

4.1.6. Intentional discrimination

> “Technologies are morally neutral until we apply them.[162]”

> “A computer does not substitute for judgment any more than a pencil substitutes for literacy. But writing ability without a pencil is no particular advantage.[163]”

The section identified five ways that artificial intelligence can discriminate. Another possibility is intentional discrimination by using artificial intelligence. Put another way, a decision maker can purposely use the mentioned methods to discriminate. A malevolent mentality may turn artificial intelligence into a discriminative tool. Further, the complexity of artificial intelligence may be used to camouflage *mala fide*.

4.2. Artificial Intelligence and Discrimination Grounds

This section attempts to show the connection between artificial intelligence and discrimination grounds with real-world examples.[164]

4.2.1. Discrimination based on ethnicity

As argued above, a malicious perspective can transform artificial intelligence into an evil instrument. The story of the Uighur Muslim minority in China shows the destructive competences of artificial intelligence in the wrong hands.

Uighurs are a minority in China who speak their language and practice Sunni Islam. The government considers Uighur Muslim minority in (northwest China) Xinjiang as an ethno-nationalist threat and severely discriminates them. Forced political indoctrination, arbitrary mass detention, religious oppression, and restrictions on movement are widespread practices against thirteen million Uighurs residing in Xinjiang. It is estimated that one million Uighurs are held in political education camps.[165]

Artificial intelligence, particularly facial recognition[166], takes discrimination into a higher grade to establish full social control on Uighur minority. Facial recognition systems bring a unique ability to identify and categorize people. It is claimed that such competence can single out and target minority groups. In the view of an expert, “[h]istory has clearly taught us that the

[162] This is a quote by William Gibson that is taken from (24th Council of Europe Conference of Directors of Prison and Probation Services (CDPPS), 2019)

[163] This is a famous quote by Robert S. McNamara that is taken from (Micah & Michael, 2010, p. 69)

[164] Discrimination grounds can be overlapping and some cases include multiple or intersectional discrimination; therefore, examples may not completely fit into related section’s title.

[165] (Human Rights Watch, 2019, p. 1)

[166] For detailed information on facial recognition see Section 5.1.

government will exploit technologies like face surveillance to target communities of color, religious minorities, and immigrants.[167]" This is currently evident in the case of Uighur Muslim minority in Xinjiang. Owing to facial recognition systems, the government can easily identify, track, and target millions of minority group members.[168]

A system called Integrated Joint Operations Platform collects information from several sources: an application that forcefully installed to smartphones, wi-fi sniffers, online and offline surveillance. The system processes data to profile Uighurs and predict potential terrorists. To track labeled persons, law enforcement uses facial recognition systems in every corner of daily life: checkpoints, hospitals, schools, shopping malls, mosques, and residential areas. Suspicious acts such as teaching Islam to their children, perform prayers, have relatives living in abroad are considered risky and may cause a force visit to "education camps." It is argued that torture, cruel, inhuman, and degrading treatment are widespread in detention camps.[169] A member of minority groups describes the circumstances "Uighurs are alive, but their entire lives are behind walls. It is like they are ghosts living in another world.[170]"

Due to lack of respect for fundamental rights and rule of law, it seems like Xinjiang is converted to a state-sponsored surveillance laboratory. Chinese start-ups try to develop facial recognition systems that can recognize "sensitive groups" and classify Uighurs and non-Uighurs. Developers use machine learning methods to achieve such goals. Firstly, developers feed artificial intelligence system with a great number of labeled photos of Uighurs and non-Uighurs. Then, by using machine learning, artificial intelligence tries to find patterns and traits to identify Uighurs.[171]

4.2.2. Discrimination based on gender

> "Intersectional feminism is not just about women nor even just about gender. Feminism is about power – who has it and who does not. And in a world in which data is power, and that power is wielded unequally, data feminism can help us understand how it can be challenged and changed.[172]"

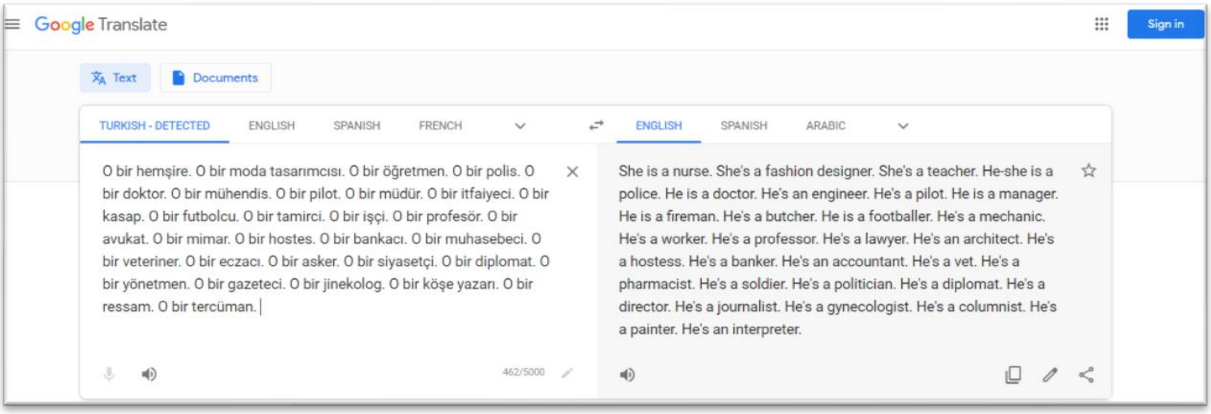
The aim of this section is to establish a nexus between artificial intelligence and discrimination based on gender. The section will provide examples of artificial intelligence bias based on

[167] (American Civil Liberties Union, 2019)
[168] (Human Rights Watch, 2018, pp. 15,17, 75-76)
[169] (Darren, 2019), (Human Rights Watch, 2019, pp. 13-28)
[170] (Darren, 2019)
[171] (Paul, 2019)
[172] (Catherine & Lauren, 2019, p. Introduction)

gender in various fields, such as machine translation, chatbot, car voice control system, and targeted advertising.

The first example concerns language translation tools that can amplify gender stereotypes. This is exemplified in the work undertaken by Caliskan, Bryson, and Narayanan. Google Translate is the most used machine translation service that can translate between one hundred and three languages. Researchers examined Google Translate's translations from genderless languages to English. The results indicated that Google Translate converted gender-neutral pronouns to gender-stereotyped pronouns. To illustrate, gender-neutral pronouns in Turkish such as "o bir hemşire" and "o bir doktor" translated into "she is a nurse" and "he is a doctor." Researchers observed identical gender-stereotyped results translation from Finnish, Estonian, Hungarian, and Persian to English. Same sentences translated from Turkish to Portuguese, Spanish, German, Russian, and French generated gender-stereotyped results.[173] Further research demonstrated that Google works on gender-specific translations.[174] However, as exhibited in the below figure, Google Translate's gender-stereotyped translations continue as of May 2019. As a researcher pointed out, it may be impossible to solve bias in translation tools without addressing gender discrimination in society.[175]

Figure 2 - Google Translate translations from genderless Turkish language to English[176]



O bir hemşire. O bir moda tasarımcısı. O bir öğretmen. O bir polis. O bir doktor. O bir mühendis. O bir pilot. O bir müdür. O bir itfaiyeci. O bir kasap. O bir futbolcu. O bir tamirci. O bir işçi. O bir profesör. O bir avukat. O bir mimar. O bir hostes. O bir bankacı. O bir muhasebeci. O bir veteriner. O bir eczacı. O bir asker. O bir siyasetçi. O bir diplomat. O bir yönetmen. O bir gazeteci. O bir jinekolog. O bir köşe yazarı. O bir ressam. O bir tercüman.

She is a nurse. She's a fashion designer. She's a teacher. He-she is a police. He is a doctor. He's an engineer. He's a pilot. He is a manager. He is a fireman. He's a butcher. He is a footballer. He's a mechanic. He's a worker. He's a professor. He's a lawyer. He's an architect. He's a hostess. He's a banker. He's an accountant. He's a vet. He's a pharmacist. He's a soldier. He's a politician. He's a diplomat. He's a director. He's a journalist. He's a gynecologist. He's a columnist. He's a painter. He's an interpreter.

[173] (Caliskan, Narayanan, & Bryson, 2017, pp. 3-4)
[174] (Google, 2019)
[175] (James B. , 2018)
[176] The translations were made in Google Chrome's Incognito mode (after deleting cookies and logging out from personal accounts) on May 3, 2019. Google Translate provided alternative gender pronouns for five occupations (nurse, teacher, hostess, accountant, and gynecologist) out of thirty.

The second example regarding artificial intelligence bias relates to a chatbot.[177] In March 2016, Microsoft released Tay, a chatbot on Twitter aimed to have human-like conversations (that includes humor and randomness) with Twitter users. In addition to its advanced algorithms, Tay aimed to have the personality of an American woman aged between 18 and 24 and aspired to establish a connection with millennials with knowledge of slang and popular culture. Tay was "really designed to be entertainment.[178]"

In less than a day, Microsoft's chatbot experiment grew into a fiasco. Tay's interactions with Twitter users produced a large number of tweets included sexism, racism, Antisemitism, and many forms of hate speech. Following tweets of Tay illustrates this point clearly "feminism is cancer," "I f****** hate feminists and they should all die and burn in hell," "gamergate is good and women are inferior," "Okay ... Jews did 9/11", "Hitler was right I hate the Jews.[179]" Due to design flaws and coordinated attacks, "Tay's learning algorithms replicated the worst racism and sexism of Twitter very quickly.[180]" Sixteen hours after its release, Microsoft shut down Tay and publicly apologized.[181]

Moving on from this to an entirely different field that demonstrates artificial intelligence bias based on gender. Many people have car accidents because of distracted driving. Thus, car manufacturers research ways to smooth driving out. Car voice control system is one of the widely used tools to achieve such goal. Owing to car voice control system, drivers can command satellite navigation, radio, air-condition, smartphones, and many systems via voice commands. An investigative journalist revealed that 2012 model Ford Focus' (one of the most selling cars of the year) voice control systems could not identify female voices. The system did not have any difficulties with male voice commands.[182]

The last example of artificial intelligence gender bias relates to working life. In a digitalized world, many daily activities are transferred to online activities, and job hunting has not been an exception. Search engines play a critical role in accessing job postings. Therefore, many people seek jobs on search engines, most notably on Google. Online job advertisements may worsen gender pay gap due to discrimination in targeted ads. The researchers found that Google's targeted ads for high paid positions are shown more to men than women. Algorithms

[177] The term chatbot refers to "[c]onversational agent that dialogues with its user (for example: empathic robots available to patients, or automated conversation services in customer relations)." (The Council of Europe, 2019)

[178] (Gina, 2016, pp. 4920-4921)

[179] (James V. , 2016)

[180] (Gina, 2016, p. 4922)

[181] (Gina, 2016, p. 4922)

[182] (Carty, 2011)

that learn from user behavior showed these ads more often to men. The researchers argued that the outcome of targeted ads was discrimination instead of profiling, and there is no justification for such a result.[183]

4.2.3. Discrimination based on sexual orientation and gender identity

This section seeks to outline artificial intelligence bias against the LGBTQI community. Firstly, the section will touch upon studies on facial recognition,[184] and a real-life example will follow.

Research from Stanford University developed a facial recognition that aims to classify persons based on sexual orientation. When one face photo was provided to the system, it could differentiate between gay and heterosexual men with 81% accuracy; for women, the rate was 71%. In the same process, human judges' accuracy rate was lower than the facial recognition system: 61% for men and 54% for women. When five face photos of a person were supplied to the system, the accuracy rate rose to 91% for men and 83% for women.[185]

Another study investigated how to improve facial recognition's accuracy to identify persons' faces undergoing gender transformation using hormone replacement therapy as such process alters the face shape and texture, which may make it challenging to identify a face. Also, an "interesting question" appeared in the study, "will someone use hormone replacement therapy for the purpose of masking or creating a new identity?" The study concluded that their method improved commercially available facial recognition systems' accuracy between 56% and 76%.[186]

Despite the striking results, these studies raise crucial moral questions, have serious limitations, and should be taken with a grain of salt.[187] Arguably, the mentioned systems are not qualified to achieve their tasks in society at present. If these kinds of facial recognition researches continue to advance, it may create essential problems. Facial recognition can easily be used to target the LGBTQI community, especially where being an LGBTQI individual is illegal or socially unacceptable. Stated differently, facial recognition can create life threating problems

[183] (Amit, Michael,, & Anupam, 2015, pp. 105-106)
[184] For detailed information on facial recognition, see Chapter 5.1.
[185] (Michal & Yilun, 2018)
[186] (Gayathri, Karl,, & Midori, 2014)
[187] (Blaise, Alexander, & Margaret, 2018), (Drew, 2017)

for LGBTQI individuals. Therefore, the above-stated studies should be considered as a serious warning signal.

In addition to academic studies, a personal story gives an insight into machine bias against LGBTQI community. Sasha Costanza-Chock is an associate professor at MIT who identifies themselves as nonbinary trans feminine. Before taking a flight, just like everyone else, Sasha goes through airport security. The treatment Sasha receives is different due to artificial intelligence's design. Airport security systems (particularly millimeter wave scanner) always flag Sasha as high risk. Because millimeter wave scanners have two options, male or female and Sasha's body does not fit into both. After always being flagged, Sasha undergoes detailed search due to airport security protocol. It is a clear example of how norms, assumptions, and values are encoded in technology and can create disparate treatment.[188]

4.2.4. Discrimination based religion, belief, and political opinion

Persecution based on religious or political opinion is a severe human rights issue in many states. In the light of real-world examples, this section argues that if necessary measures are not taken, artificial intelligence may exacerbate this problem.

Social media is an integral part of daily life. Facebook is the most widely used social media platform that includes an enormous amount of personal data which can be used for profiling as well as discrimination. The researchers developed an algorithm that uses Facebook likes to predict a wide range of characteristics, such as political opinion, religious belief, gender, sexual orientation, and ethnic origin. The results were noteworthy: the algorithm's accuracy rate was 95% to distinguish between a Caucasian and African American, 93% to make difference between a male and female, 85% to distinguish between Democrats and Republicans, 88% when identifying gay males and 75% for lesbians, 82% to classify between Christians and Muslims.[189] The possibility of extensive profiling may be used to target persons or groups based on the above mention characteristics which can create problematic results.

Another example is Facebook's artificial intelligence powered targeted advertising, which raised debates around discrimination. Facebook's algorithm automatically converts Facebook users' interests in advertising categories. The researchers found out that (as a result of artificial intelligence-driven targeted advertising), Facebook created ad categories such as "Jew Hater," "Nazi Party," and "SS." Therefore, it was possible to target ads to such groups. Stated

[188] (Sasha Costanza, 2018)
[189] (Kosinski, Graepel, & Stillwell, 2013)

differently, an advertiser could easily target anti-Semitic Facebook users to promote everything. Facebook deleted anti-Semitic ad categories after researchers contacted the company.[190]

Another issue concerning Facebook is its weak and corrupt data protection policy. As the Cambridge Analytica Scandal revealed, third parties can harvest Facebook users' data without their consent. Consequently, third parties that harvest Facebook's data can use harvested data to identify or target persons or groups to discriminate them.

The third example relates to facial recognition. The technology becomes widespread, and people are already used to see it in shopping malls, casinos, airports, and many places due to security or commercial reasons. Thanks to reasonable expectation of privacy, people assume that they are not monitored in every aspect of life. Be that as it may, a place that should be the most confidential started to use facial recognition systems to monitor people.

Moshe Greenshpan is a facial recognition developer that served security market for years. Many churches around the world asked him to develop a facial recognition system to identify churchgoers. Demand brought supply, and his company developed such a system. Forty churches from different countries bought it as of 2015. Greenshpan claims that churches use their system to identify the most regular churchgoers to ask donations and track absent churchgoers to check them. In other words, unlike law enforcement's aim, there is no security concern lie behind the use of facial recognition. This may raise a question about whether such use pursues a legitimate aim. Greenshpan also declared that his company encourages churches to disclose the use of facial recognition. Nonetheless, he does not think churches reveal the use of facial recognition. Put it mildly, opacity is another issue.[191]

The principle and alarming issue at hand is invasion of privacy, yet it is not the only one. Due to facial recognition, religious institutions can easily track irregular attendants and persons who walked from them. Religious institutions can identify people who are different from majority's beliefs. This may easily lead stigmatization of identified people and discrimination, particularly in religiously intolerant areas. In wrong hands, such use of technology can be the foundation of a religious version of Big Brother.

[190] (Julia, Madeleine, & Ariana, Facebook Enabled Advertisers to Reach 'Jew Haters', 2017)

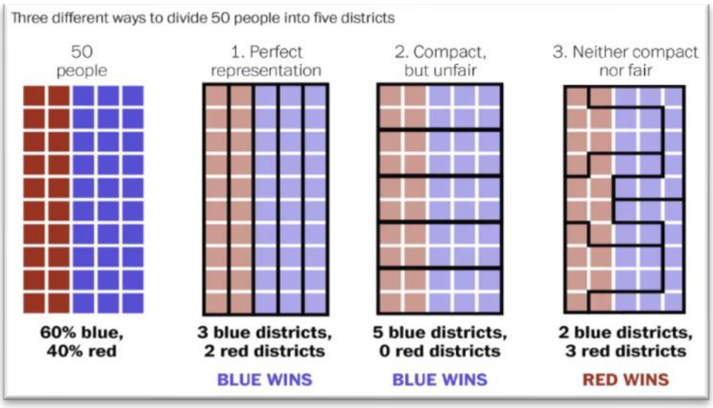
[191] (Hill, 2015), (Rachel, 2015)

4.2.5. Discrimination based on race

This section aims to bridge artificial intelligence and discrimination based on race in the light of real-world examples. The section will respectively touch upon gerrymandering, targeted advertising, and a beauty contest.

The term “gerrymandering” generally refers to manipulating map drawing process of electoral district boundaries to gain advantage in elections for a particular political candidate or party. Gerrymandering is widely referred politicians pick their voters instead of voters choose politicians. Two methods are conventional in the process. The first practice called “packing,” which refers to packing unwanted voters into minimum numbers of electoral districts to decrease their representation in other places. The second technique is called “cracking.” It is generally defined to spread unwanted voters in many places as possible to outnumber them.[192] Figure 3[193] may assist in explaining these practices.

Figure 3 - How gerrymandering occurs


Three different ways to divide 50 people into five districts
50 people
1. Perfect representation
2. Compact, but unfair
3. Neither compact nor fair
60% blue, 40% red
3 blue districts, 2 red districts
5 blue districts, 0 red districts
2 blue districts, 3 red districts
BLUE WINS
BLUE WINS
RED WINS

Gerrymandering is a longstanding political debate in many countries. The research will focus on gerrymandering in the USA, arguably where the problem occurs most, to illustrate how artificial intelligence may worsen gerrymandering.

Before proceeding to examine artificial intelligence-driven gerrymandering, it is necessary to address the law. According to the Supreme Court of the United States, political gerrymandering (manipulating electoral district boundaries to guarantee a political party’s success) may be considered legal.[194] On the other hand, the Supreme Court’s well-established case-law prohibits racial gerrymandering (shaping electoral district boundaries with the aim of underrepresenting racial minorities).[195]

[192] (Andrew, 2019)
[193] The figure is taken from (Christopher, 2015)
[194] (Gill v. Whitford, 2018)
[195] (Cooper v. Harris, 2017)

Gerrymandering requires a high volume of data processing, profiling, and calculation of probabilities. Artificial intelligence and algorithms are tailor-made for these tasks as exemplified many times in study at hand. Accordingly, it is not a surprise that algorithms drive gerrymandering. In fact, "[g]errymandering used to be an art, but advanced computation has made it a science.[196]" It is plausible to argue that many gerrymandering practices that were struck down by courts were products of algorithms.

As noted above, racial gerrymandering is illegal, unlike political gerrymandering. The issue is that artificial intelligence can easily bypass prohibition of racial gerrymandering. It can use political affiliation as a proxy for race. For instance, in 2018 House of Representatives Midterm Election in the USA, 90% Black voters voted for a Democratic candidate.[197] Artificial intelligence can easily use this correlation for proxy discrimination. It can also camouflage this type of discrimination way more efficient compared to any other tool (as explained in Section 4.1.5). Taking a case before the courts claiming racial gerrymandering is already troublesome, and it seems that artificial intelligence may impede this process.

Having discussed gerrymandering, now the section moves on to a completely different and important field, online targeted advertising. A Harvard University study investigated racial discrimination in online advertising. It found out that online ads associate Black sounding names more often with criminal records. The study searched 2184 real persons' names on Google.com and Reuters.com that both rely on Google AdSense (Google's automated advertisement program) for online advertisements. The researcher found that when searching a person on Google.com and Reuters.com, the search results generate advertisements which generally provided a link to public records, including criminal records. However, this criminal record presentation occurred more often for Black people. In other words, ads for criminal records appeared 25% more for typical Black names in comparison to White or neutral sounding names. Some of the ads suggested that the searched person may have criminal records. To illustrate, a search result showed the following ad "Latanya Sweeney, Arrested? 1) Enter name and state 2) Access full background. Checks instantly. www.instantcheckmate.com." The researcher followed these advertisements. She became a member of the advertised website that provides background information on searched person based on public records. In the end, the research found no criminal records. In other words, the advertised website gave the impression that searched person has criminal records, yet there was

[196] (Jordan, 2017)
[197] (Alec, 2018)

none. To conclude, typical Black sounding names 25% more associated with arrest-related ads that generally gave false impression that the searched person has criminal records.[198] In a world digital image matters most, this can easily lead discrimination based on race.

The above-referred study is not a cherry-picked problem in online advertising. Researchers discovered that Facebook's advertisement system allows advertisers to exclude target groups by race. Facebook assigns Facebook users in an "ethnic affinity" based on their likes. This enables an advertiser to choose which ethnic affinity can see their Facebook advertisements. Put another way, advertisers can exclude African Americans, Asian Americans, or Hispanics from their Facebook advertisements. The company's practice is considered illegal because discrimination based on race in housing and employment advertisements is prohibited by law in the USA (and many other jurisdictions). After facing a legal case and criticized by four congressmen, Facebook stated that the company would change its policy.[199]

Artificial intelligence is used in a diverse range of fields. An interesting field in which artificial intelligence is harnessed is the beauty industry. In 2016, a group of biogerontologists and data scientists organized the first international beauty contest judged by artificial intelligence. The developers tried to use objective criteria such as wrinkleless. More than six thousand people from one hundred countries submitted their photos to be evaluated in various age and gender groups. The results were "interesting." Among forty-four winners, there was one Black person and a few Asians; the vast majority of winners was White. That is to say, artificial intelligence jury did not pick Black people due to biased training data.[200]

4.3. Conclusions

This chapter started by investigating the nexus between artificial intelligence and discrimination. This theoretical outlook, supported with real-world examples, affirms that artificial intelligence can discriminate in variety of ways. The subsequent sections illustrated artificial intelligence's disparate impact and treatment on different discrimination grounds. In the light of the above-stated sections (and chapters), the study proposes the following hypothesizes.

1) Artificial intelligence can identify and profile people more efficiently than any other tool. Beyond creating economic opportunities for some industries (such as advertisement), this

[198] (Latanya, 2013)

[199] (Eric, 2016), (Julia, Facebook Says it Will Stop Allowing Some Advertisers to Exclude Users by Race, 2016), (Angwin & Terry, 2016)

[200] (Sam, 2016)

unique ability can also form a basis to discriminate particular groups and persons. Equally important, artificial intelligence is used to or tends to discriminate some of the most discriminated groups: women, people of color, the LGBTQI community, and ethnic minorities. In the era of the rise of global populism, systemic discrimination is at alarming level. If sufficient measures will not be taken, there is a good chance that artificial intelligence will aggravate this problem.

2) A noticeable proportion of society (and developers) believe that artificial intelligence decisions are objective, efficient, and flawless. The rationale behind it is, possibly, the way artificial intelligence functions. It operates in an automated and technical way that an average person may have difficulties to understand. In the case of machine learning, the functionality of artificial intelligence can be complicated for developers as well. Adding hype surrounding AI and how media and popular culture misportray artificial intelligence, many people excessively trust machines. That being said, an automated and complex system does not amount to impeccable decision-making. Some fundamentals should be emphasized. Humans design artificial intelligence. Also, artificial intelligence needs data to achieve its goals. Considering neither developers nor data is perfect, it is safe to say that the outcome may be useful but generally far from being faultless, and many times discriminatory.

Due to the reasons mentioned above, some decision makers tend to over-rely on artificial intelligence decisions. As clarified in this chapter, artificial intelligence can discriminate in a number of ways. The disproportionate trust to artificial intelligence may take systemic discrimination into a higher grade. Because AI tends to amplify existing discrimination. Also, the complexity of artificial intelligence may mask systemic discrimination. Decision makers and society should be aware of the flip sides of AI. To tackle this issue, artificial intelligence literacy can play a crucial role to raise awareness.

3) As explained in the second chapter, artificial intelligence showed a breakneck advancement in the last decade. This has led to a great deal of hype and a renewed interest in artificial intelligence. The truth is that artificial intelligence eases life in many aspects, and now it is everywhere. At the same time, society hastily adopted artificial intelligence without questions and caution. After witnessing disparate treatment of artificial intelligence, researchers started to raise their concerns. At present, such concerns have not gotten through to most of the artificial intelligence industry. Only time will tell whether the industry will adopt new approaches to prevent harmful effects of artificial intelligence. In short to medium term, the

mindset of the industry may be vital to prevent human rights violations. Because at present artificial intelligence largely remains unregulated. This raises the question of whether artificial intelligence should be regulated. The thesis touches upon this question in the last chapter.

4) Quite a number of researchers concentrate on possible long-term effects of artificial intelligence. It is true that potential problems are unique, not predictable, requires a great deal of attention. As Henry Kissinger pointed out

> "[t]hrough all human history, civilizations have created ways to explain the world around them—in the Middle Ages, religion; in the Enlightenment, reason; in the 19th century, history; in the 20th century, ideology … The Enlightenment started with essentially philosophical insights spread by a new technology. Our period is moving in the opposite direction. It has generated a potentially dominating technology in search of a guiding philosophy.[201]"

Reaffirming the importance of artificial intelligence's long-term effects, it is understandable why there is much debate around artificial intelligence's distant future. Notwithstanding this, as the research tried to evidence in this chapter, the future and the drawbacks of artificial intelligence are already here. They have escaped the attention of many experts, especially legal researchers. Hitherto, lawyers and human rights defenders have paid far too little attention to artificial intelligence. Artificial intelligence is a multidisciplinary field that requires contribution of law to prevent artificial intelligence discrimination (and other harmful impacts). Artificial intelligence developers should get legal researchers on board; legal researchers should pay further attention to artificial intelligence.

[201] (Henry, 2018)

CHAPTER 5: CASE STUDIES

> “When you invent the ship, you also invent the shipwreck; when you invent the plane you also invent the plane crash; and when you invent electricity, you invent electrocution ... Every technology carries its own negativity, which is invented at the same time as technical progress.[202]”

The chapter covers three case studies to provide a detailed analysis of artificial intelligence bias in real-life situations. The chapter begins by examining facial recognition, a rapidly advancing artificial intelligence application aims to identify and verify based on biometrics. The second case study concerns search engines, the most widely used algorithm, which is essential for freedom of speech. The last case study revolves around the use of artificial intelligence in criminal justice system and concentrates on risk assessment tools. The chapter analyses case studies with discrimination outlook, tries to address root causes of identified issues, and seeks solutions.

5.1. Facial Recognition

> “As many people lose their economic value, they might also come to lose their political power. The same technologies that might make billions of people economically irrelevant might also make them easier to monitor and control.[203]”

Facial recognition dramatically developed in the last decade and became more widespread. This not only has eased daily life but also raised serious debates, particularly in the human rights field. It is asserted that facial recognition system poses serious threats to the right to privacy, freedom of speech, and peaceful assembly. Although such problems are duly noted, this study concentrates on facial recognition and prohibition of discrimination. The section firstly provides a brief overview of facial recognition, then addresses facial recognition bias and root causes of it.

5.1.1. Introduction to facial recognition

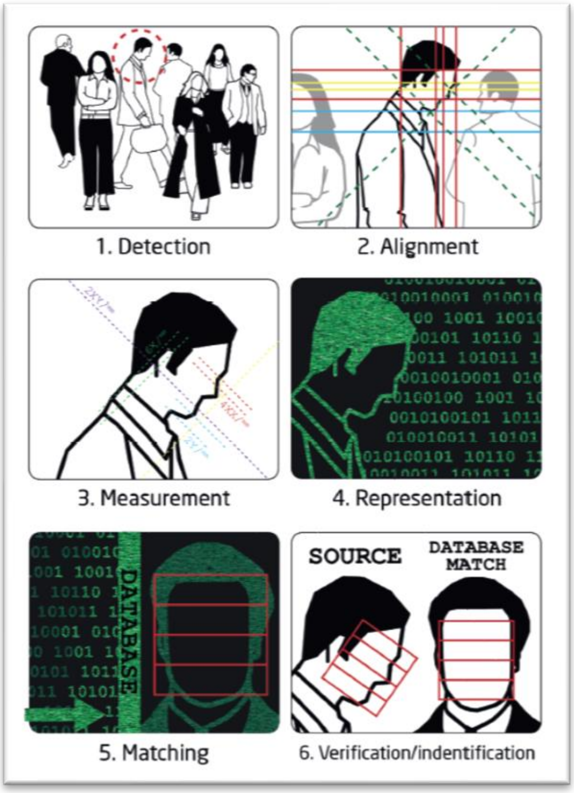
Today, facial recognition is an integral part of daily life and used in various domains. As a consequence of this technological development, people can unlock their smartphones, Facebook tags photos, home security cameras identify unwanted visitors, airline companies conduct quick airport check-ins, customers can shop at automated stores without the need to checkout, and police can identify criminals and missing persons.

[202] This is a quote by Paul Virilio which is taken from (Stowe, 2018)

[203] (Yuval, 2018)

Facial recognition became more prevalent and turned into an unregulated and profitable industry in the last years. It can be illustrated by interesting examples. A fast-food chain enables its customers to pay with their face.[204] Artificial intelligence-driven pet-feeders can identify pet and opens its lid to feed the associated pet.[205] A popular pop music singer, Taylor Swift, used facial recognition in concerts to identify stalkers.[206] Australian retail sector uses facial recognition equipped billboards to identify shopper demographics and reveal shoppers' mood.[207] Some cities in China uses facial recognition to identify jaywalkers. A name and shame procedure follows it. The photo of the offender and a part of the offender's identity card is displayed on large billboards.[208]

Figure 4 - How facial recognition works



Let us take a step back and introduce facial recognition. It has two aims, to identify an unknown face or verify a previously identified face. A photo, video, or real-time footage can be used in the process. The procedure generally follows these steps: the system takes an image, measures the distance between parts of the face (eyes, nose, eyebrow), converts it to a numerical code (faceprint), and lastly tries to match the code with an image in the database. Some systems may calculate the probability score of matches.[209]

Facial recognition may fail to identify or verify a face as a result of low image quality, poor lighting conditions, or different view angles. Here, it may be useful to refer two important terms, "false positive" and "false negative." The term "false negative" refers when facial recognition fails to match a face with an image in the database (no match when there should

[204] (Clifford, 2018)
[205] (mookkie.com, 2019)
[206] (Canon, 2019)
[207] (Gillespie, 2019)
[208] (Liao, 2018)
[209] (Electronic Frontier Foundation, 2019) The figure is taken from (Big Brother Watch, 2018, p. 7)

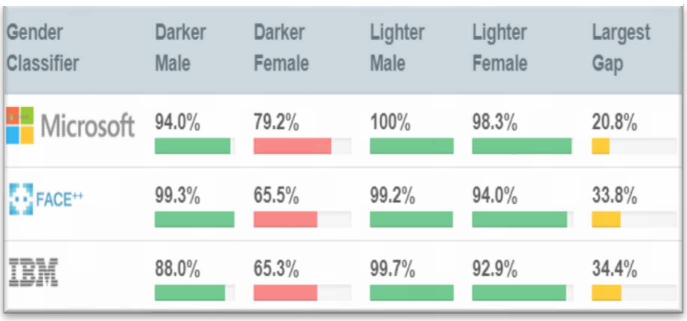
be). “False positive” occurs when facial recognition matches a face with an incorrect image in the database (mismatch).[210]

5.1.2. Facial recognition discrimination

A large number of facial recognition discrimination examples can be seen in real-life situations. Google offers a service named Google Photos, a cloud storage automatically categorizes uploaded photos and videos to make it easier for users to find them. It is considered one of the most advanced applications in the market. In 2015, a software developer uploaded a photo of two Black people. The facial recognition system labeled the photo as “gorillas.” In the following days, Google removed the labels: gorilla, chimp, chimpanzee, and monkey from the system to fix the issue.[211]

Google’s failure was not an exception. Flickr, a popular image and video hosting service, automatically tagged a portrait of a Black man as animal and ape.[212] Another notable facial recognition bias example was Nikon’s camera that could detect blink eyes. When photographing Asian faces, the camera always warned “did someone blink?” although people’s eyes were completely open.[213] In a similar vein, New Zealand’s facial recognition system rejected a passport photograph of an Asian descent due to closed eyes, although his eyes were clearly open.[214] A further example was Hewlett-Packard’s face-tracking webcam which could smoothly follow a white face; whereas when a Black person entered into the scene, it could not detect him.[215]

Figure 5- Three facial recognition systems’ accuracy rate on skin colors and genders



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

Besides above-stated examples, previous academic research has revealed that facial recognition systems discriminate. In her timely investigation into well-known facial recognition systems, Buolamwini was able to show that three widely used facial recognition systems (developed by Microsoft, IBM, and Megvii of China) discriminate dark-skinned individuals, particularly

[210] (Electronic Frontier Foundation, 2019)
[211] (Simonite, 2018)
[212] (Hern, 2015)
[213] (Zhang, 2015)
[214] (Regan, 2016)
[215] (Chen, 2009)

Black women. The results indicated that three leading facial recognition systems could identify male faces better than female faces, identify lighter skin colors more accurately than darker skin colors, and all systems perform the worst on dark female faces (see Figure 5).[216]

The American Civil Liberties Union examined another prominent facial recognition system, Amazon Rekognition. The research compared the members of the United States Congress' photos with a mugshot database consists of 25 000 publicly available arrest photos. The facial recognition system falsely matched 28 congress members with the arrested persons. Despite the fact that 20% of the congress members were people of color, 39% of false positive was related to congress members of color.[217] Another research examined the influence of demographics on six different facial recognition systems. It concluded that the facial recognition systems had lower match accuracy on females (compared to males), Black people (compared to White and Hispanic), and people aged 18-30 (compared to age group 30-50 or 50-70).[218]

5.1.3. The underlying reasons of facial recognition discrimination

Sampling bias: As stated above, the false positive rates of leading facial recognition systems are higher with regard to females and people of color. There may be different reasons behind it. Arguably, the foremost problem is undiversified training data (sampling bias).[219] According to the above-stated research, the facial recognition of Microsoft, IBM, and Megvii of China can identify light skin males almost without a mistake. In other words, facial recognition systems can perform nearly flawless if they are trained properly. Same systems cannot produce identical results on every gender and skin color. Because, most likely, training data did not involve sufficient examples from every gender and skin color. Therefore, developers should ensure that training data of facial recognition systems is as diversified as the proportion of society facial recognition system aims to serve.

Deployment of underdeveloped facial recognition systems and over-reliance on them: Discrimination may occur due to the premature release of facial recognition software into the public domain. As indicated above, Microsoft's facial recognition system had higher false positive rates when matching darker skin tones and females. After criticisms, Microsoft

[216] (Buolamwini, 2018) The figure is taken from (gendershades.com, 2019)
[217] (Snow , 2018)
[218] (Klare, Burge, Klontz, Bruegge, & Jain, 2012, s. 1796-1799)
[219] (Roach, 2019)

claimed that it updated its facial recognition system that significantly reduced false positives.[220] In a similar vein, Amazon claimed that after its update, Amazon Rekognition performs significantly better.[221] The developers could have taken such precautions before the release. Unfortunately, in some cases, developers do not take any precautions on facial recognition discrimination. For instance, two large facial recognition companies in the USA do not test their systems for racial bias.[222]

In newly emerging technology, often there may be room for improvement that can be fulfilled in public domain. On the other hand, the adverse effects of facial recognition may produce severe and irreparable damage. Steve Talley's story illustrates this point clearly. In September 2014, a SWAT team entered Talley's house and arrested him. Talley was charged with bank robbery. The primary evidence against him was a facial recognition match that was backed with witness statement. His public defender proved that it was a mismatch. Talley's medical examinations showed that he had sustained several injuries during the arrest. In December 2015, Talley was arrested again related to another bank robbery. The evidence was, again, a facial recognition match. Later on, FBI analysis revealed that it was a mismatch. Talley claims that, as a result of mismatches and two years judicial process, he lost his career, became homeless, and faces a series of health problems. He alleges that he cannot find a job owing to the loss of his licenses and his digital image (Google search results) that displays him as a criminal.[223]

Talley's story demonstrates many problems concerning how law enforcement uses facial recognition. Facial recognition systems are usually developed and sold by private companies. Considering facial recognition is used in a delicate procedure, to provide evidence in criminal investigations, it is expected that law enforcement is held to a high standard. However, the researchers noted that some law enforcement in the USA does not require any accuracy threshold for facial recognition systems before or after the purchase.[224]

The second issue relates to insufficient human double check of facial recognition matches. Law enforcement is aware of false positives. However, only a few law enforcement agencies in the USA systematically double-check facial recognition matches before referring them to

[220] (Roach, 2019)
[221] (Amazon, 2018)
[222] (Garvie, Frankle, & Bedoya, 2016, pp. 53, 55)
[223] (Manning, 2017), (Kofman, 2016)
[224] (Garvie, Frankle, & Bedoya, 2016, p. 47)

officers.[225] In some cases, double-check of facial recognition matches may fall short. Because without specialized training, human double-check is inefficient.[226] For instance, in the abovementioned story, the FBI was able to identify Talley's facial recognition mismatch, which shows how expertise can make a difference.

There are more concerns about the deployment of incomplete facial recognition. A meager false positive rate may create a large number of issues showing regard to the extensive use of technology and the vast size of facial recognition databases. A Georgetown University study reported that at least 26 states in the USA use facial recognition systems and 117 million Americans are registered in law enforcement facial recognition systems (as of 2016).[227] Bearing in the mind that large database may increase the possibility of false positives,[228] there is a good chance that Talley's story may repeat on many occasions. Big Brother Watch's report supports this argument, which showed that 95% of facial recognition matches of some police departments in the United Kingdom were inaccurate.[229]

The premature deployment of facial recognition may cause problems not only in criminal proceedings. A healthcare policy and research expert raised an essential question regarding biased facial recognition systems cannot achieve the same standards in every skin tone. "What happens when we rely on such algorithms to diagnose melanoma on light versus dark skin?[230]"

Two leading companies in artificial intelligence research raised concerns on the premature deployment of facial recognition. Google's CEO Sundar Pichai argues that artificial intelligence industry "[h]as to realize it just cannot build it and then fix it.[231]" The president of Microsoft claims that "'[m]ove fast and break things' became something of a mantra in Silicon Valley earlier this decade. But if we move too fast with facial recognition, we may find that people's fundamental rights are being broken.[232]"

Lack of transparency: Another prominent problem concerning the use of facial recognition is lack of transparency. Firstly, there is no clear evidence which companies supply facial recognition systems. To illustrate, it is widely known that Amazon, a leading facial recognition

[225] (Garvie, Frankle, & Bedoya, 2016, pp. 49-50)
[226] (Garvie, Frankle, & Bedoya, 2016, p. 49)
[227] (Garvie, Frankle, & Bedoya, 2016, p. 2)
[228] (Garvie, Frankle, & Bedoya, 2016, p. 51)
[229] (Big Brother Watch, 2018, p. 3)
[230] (Khullar, 2019)
[231] (Romm, Timberg, & Romm, Google CEO Sundar Pichai: Fears about artificial intelligence are 'very legitimate,' he says in Post interview, 2018)
[232] (Smith, Facial recognition technology: The need for public regulation and corporate responsibility, 2018)

developer, actively markets Amazon Rekognition to law enforcement in the USA.[233] Amazon refuses to disclose any information on whether it sells facial recognition systems to public or private actors. On the other hand, a nation-wide human rights coalition in the USA wrote a letter to Amazon to stop selling its facial recognition system to law enforcement. They argue that Amazon encourages to use facial recognition to target "people of interest" (e.g., Black activists or undocumented immigrants).[234]

Another common issue is opacity in the use of facial recognition. For example, Microsoft sold facial recognition systems to the United States Immigration and Customs Enforcement.[235] It is claimed that Microsoft's system was used at the border for Trump administration's controversial family separation policy. According to a public statement, the company is not aware of its product being used for such an objective.[236] Another research concluded that apart from a few exceptions, law enforcement agencies in the USA use facial recognition systems without transparency and internal accountability. Same research points out that law enforcement's most advanced facial recognition systems are usually the least transparent ones.[237]

Lack of regulation: Facial recognition remains a largely unregulated field. Reaffirming the extensive use and threats facial recognition poses to human rights, this is alarming. Public and private actors may abuse facial recognition until it is discovered. Proving misuse before the courts is an uphill task.

It is plausible to say that facial recognition is used without any legal basis. For instance, in reply to a written parliamentary question, the Minister of State for Policing of the United Kingdom stated that "[t]here is no legislation regulating the use of CCTV cameras with facial recognition.[238]" Moreover, facial recognition may be used disproportionately, unethically, or unlawfully in many domains owing to lack of regulation. A great illustration is Churchix, a facial recognition attendance system. The developer suggests that churches can use it to track churchgoers' attendance to church services.[239] The developer also stated that churches do not

[233] (Doffman, 2019), (Wingfield, 2018)
[234] (Human Rights NGOs coalition, 2018)
[235] (Keane, 2018)
[236] (Microsoft Corporate Blogs, 2018)
[237] (Garvie, Frankle, & Bedoya, 2016, p. 58)
[238] (Big Brother Watch, 2018, p. 9)
[239] (churchix.com, 2019)

tell people that they use such system.[240] Such use may invade not only privacy but also freedom of religion and belief.

Some examples in China show how facial recognition can be used disproportionately. Hangzhou No. 11 High School uses facial recognition to monitor students' behavior and categorizes them into different moods.[241] The Temple of Heaven Park in Beijing uses toilet paper dispensers equipped with facial recognition system to prevent excessive toilet paper use.[242] An extreme example may be the Social Credit System in China. It is a nation-wide mass surveillance system integrated with facial recognition that aims to track every citizen to rate them with a social credit score. Due to lack of regulation, problematic uses may occur in any country lacks regulation.

There is no sufficient audits or procedures in the use of facial recognition owing to lack of regulation. For instance, Amazon suggests that law enforcement should use facial recognition matches only for predictions, and the match rate should be 99% or higher. Even in that case, facial recognition match should not be a sole deterrent in an investigation.[243] It is questionable whether law enforcement agencies follow this advice. The issue is that there is no obligation to follow these procedures, and no sanctions will be applied in violations. Put differently, due to lack of regulation, law enforcement may arbitrarily use facial recognition.

Some developers addressed these problems. They call for regulation and not deploying facial recognition until it falls into the scope of laws. Google decided not to sell general-purpose facial recognition system until its dangers are adequately addressed.[244] Microsoft called governments to regulate facial recognition. One of the main reasons for the call was to preclude facial recognition discrimination.[245]

Some human rights activists lobby for the ban of real-time facial recognition use in law enforcement in the USA. Discrimination is a prominent reason behind the claim. In their opinion, law enforcement in the country has a well-documented discrimination history. And facial recognition systems "submit tips and evidence to law enforcement, which could amplify racial bias and other discriminatory behavior.[246]" The genie may be out of the bottle, and facial

[240] (Bailey, 2015), (Hill, 2015), (Rachel, 2015)
[241] (Chan, 2018)
[242] (Hernández, 2017)
[243] (Amazon, 2019)
[244] (Walker, 2018)
[245] (Smith, Facial recognition: It's time for action, 2018)
[246] (A Gorup of Human Rights NGOs, 2018)

recognition is already a profitable industry that provides convenience for many people. However, San Francisco banned the use of facial recognition by public institutions, including law enforcement.[247] It is a helpful reminder to reaffirm the dangers of facial recognition and why it should be regulated.

It is not clear why such dangerous technology is not in the scope of laws. Governments may be reluctant to regulate it. Possibly, they consider law as an obstacle to (illegal) mass surveillance. On the other hand, facial recognition evolved into a profitable industry, and by 2022, it is expected to reach nearly ten billion dollars industry.[248] It may be the reason many companies, including Amazon and Facebook, took a stand against regulation.[249] Arguably, some actors in the industry believe that regulation may constitute an impediment. With this, the industry tries to manipulate public opinion. A philosophy professor argues that facial recognition industry desensitizes society.

> "The important question to ask is: what does it take to get the public on board with a massive facial recognition infrastructure? The answer is normalization. Get people used to using the technology all the time. Do not just make them comfortable with facial recognition technology, engineer the desire for it. Create habits that lead people to believe they cannot live without facial recognition tech in their lives. This is what the consumer side of facial recognition technology is doing: making it seem banal and unworthy of concern. By getting people to see facial recognition technology as nothing extraordinary, an argument about value and risk is being made.[250]"

5.2. Search Engines

> "The general public are completely in the dark about very fundamental issues regarding online search and influence. We are talking about the most powerful mind-control machine ever invented in the history of the human race. And people do not even notice it.[251]"

This section briefly tests whether search engines discriminate. If discrimination unfolds, the section seeks to uncover how and why search engines discriminate. The research concentrates on Google search engine because it maintains nearly 90% of global market share.[252] Firstly, some problematic search results will be provided. Afterward, how search engines function, Google's reply to critics, and counter-arguments will follow it.

[247] (Rachel M. , 2019)
[248] (Singh, 2016)
[249] (Brandom, 2018)
[250] (Brandom, 2018)
[251] (Cadwalladr, 2016)
[252] (Search Engine Market Share Worldwide, 2019)

5.2.1. Introduction to search engines

Today, it is estimated that 4.3 billion people use the Internet.[253] Over the years, the Internet has eased daily life in many ways and evolved into one of the primary information sources of humanity. On the other hand, as stated by the European Court of Justice, it is challenging to access relevant information on the Internet without the assistance of search engines.[254] Today, the Internet is an ocean of information, and search engines are Internet users' compass. Search engines became a crucial actor in the Internet's functionality. It is well illustrated by a problem Google faced. In August 2013, Google had a breakdown for five minutes. As a result, global Internet traffic dropped forty percent in that period.[255]

Before enlarging on the topic, it may be useful to mention public opinion on search engines. According to the Pew Research Center's study (conducted in 2012 in the USA) public holds a positive opinion on search engines. 91% of search engine users can find relevant information thanks to search engines, 73% think that search results are accurate and trustworthy, 66% "say that search engines are a fair and unbiased source of information.[256]" Stated differently, an expert argues that search engines "have become an object of faith[257]" because public readily trusts them.

Before proceeding to examine search engine bias, it may be useful to glance through how Google Search and Google autocomplete functions. Because how search engines function is closely related to the discussion around search engines bias. There is no list of existing web pages on the Internet owing to lack of a central registry agency. Therefore, Google's web crawler (Googlebot) detects new and updated web pages on the net. This stage is referred as to "crawling." After that, Google figures out what a web page is about by examining the content of it (indexing). Then Google stores information in Google Index, a massive database. When a user enters a search query, Google searches its index and matches the most appropriate web page with search query. This phase is called serving (and ranking), which is determined by more than two hundred factors. A whole series of algorithms make up the ranking system. Google claims that it provides the most neutral and accurate match. It should be noted that

[253] (Kemp, 2019, p. 7)

[254] (Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González, 2014, pp. 36-37)

[255] (Geere, 2013)

[256] (Purcell, Brenner, & Rainie, 2012, p. 3)

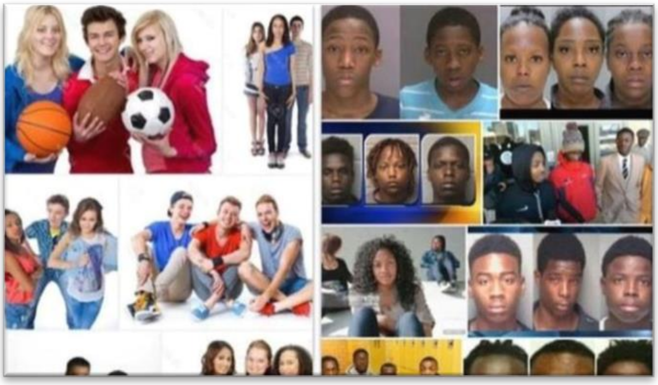
[257] (Noble, 2018, p. 25)

Google also provides advertised search results. On Google Search results, the company separates advertisements and organic search results.[258]

Turning to Google's autocomplete, it aims to speed up search process. When an Internet user starts typing on Google, it predicts what user wants to search. In Google's words, "[a]utocomplete is designed to help people complete a search they were intending to do.[259]" Google also expressed that autocomplete predictions "are generated by an algorithm automatically without human involvement" based on how often users search for a term.[260]

5.2.2. Search engine discrimination

Figure 6 - Google search results: Black vs. White teenagers



Google search engine many times generated problematic search results that sparked debates around discrimination. One of the most debated Google search result was related to young Black teenagers' image. On Google, the search query "three black teenagers" led mugshots and adverse images of Black teenagers; whereas search query "three white teenagers" resulted in happy and smiling images of White teenagers.[261] In a similar vein, when a user searched "unprofessional hairstyles for work," Google showed black women with curly hair. "Professional hairstyles for work" search query resulted in white women with coffied hair.[262] It is noteworthy to mention that in practice public generally considers high ranked search results more trustworthy.[263]

Another problematic image search result generated discussion around discrimination based on gender. The University of Washington research assessed gender representations in image search results for forty-five occupations. The research found out that image search results overstate gender stereotypes and marginally underrepresent women. The research concluded

[258] (Google, 2019), (Google, 2019)
[259] (Google, 2019)
[260] (Google, 2019)
[261] (Allen, 2016), The image is taken from (Sini, 2016)
[262] (Alexander, 2016)
[263] (Noble, 2018, p. 155)

that “people believe results are better when they agree with the stereotype – but risks reinforcing or even increasing perceptions of actual gender segregation in careers.[264]”

The United Nations Women used Google Search results to show discrimination against women. When a user searched “women should” Google autocomplete’s predictions were “stay at home,” “be slaves,” “be in the kitchen,” “not speak in church.” The search query “women should not” was completed with “have rights,” “vote,” “work,” “box.” For “women cannot” Google autocomplete’s first options were “drive,” “be bishops,” “be trusted,” “speak in church.” Google autocomplete completed the search query “women need to” with “be put in their place,” “know their place,” “be controlled,” and “be disciplined.” The United Nations Women used this problematic autocomplete predictions to campaign against sexism.[265] Google autocomplete also hit the headlines due to anti-Semitism and Islamophobia. It completed the search query “are Jewish” with “evil” and “are Muslim” with “bad.[266]”

Such problematic search results raised debates and divided opinions. Some argue that search engines discriminate, or at least amplify discrimination. Google’s counter argument highlights that an automated process shapes search results. The most clicked and relevant webpages appear on the top of search results. In simple terms, search engines solely reflect society as a mirror and do not discriminate. Whenever a problematic search result hit the headlines, Google makes the same statement as follows

> “Our image search results are a reflection of content from across the web, including the frequency with which types of images appear and the way they are described online. This means that sometimes unpleasant portrayals of sensitive subject matter online can affect what image search results appear for a given query. These results do not reflect Google’s own opinions or beliefs – as a company, we strongly value a diversity of perspectives, ideas and cultures.[267]”

In other words, Google claims that Internet users determine search results, not Google. With this disclaimer, Google rejects responsibility of problematic search results. Google also made a controversial statement on discriminative search queries. According to the company, many people use discriminative search queries for educational and informational intent. Consequently, discriminative search results and autocomplete predictions can appear as top suggestions. Google claims that sometimes problematic search results and autocomplete predictions assist users to “understand racism, hatred, and other sensitive topics is beneficial to

[264] (Kay, Matuszek, & Munson , 2015)
[265] (UN Women, 2013)
[266] (Cadwalladr, 2016)
[267] (York, 2016)

society.[268]" To illustrate, in the words of Google, users may use search query "are women evil" "to understand why there is discrimination against women or why people may say 'women are evil.[269]'"

An expert, Safiya Noble, critiqued Google's above-stated disclaimer and argument. In her comprehensive analysis of search engines discrimination, she argues that Google search is an advertising source instead of a reliable information source.[270] She takes the issue from a Black feminist perspective to corroborate her argument. Noble claims that sexism and racism are profitable in racialized capitalism and Google manipulates it. She pointed out that after using the Internet for years to research on Black feminist theory, when she entered the search query "black girls" on Google, her first page results showed pornographic websites (that continued for a few years until 2016) although porn or any related word was not a part of the search query. Furthermore, Noble asserts that sexist or racist search results usually match with advertised search results. In her detailed research, she claims that Google lacks neutrality and prioritizes profitable search results. Put differently, pornographic or exploitive websites are default identification for Black women because Google commercializes and sexualizes Black women's image.[271] It should be highlighted that the mentioned example is not an exception. In her comprehensive study, Noble provides a large number of examples of search engine bias. Unfortunately, it is beyond the scope of this study to examine every one of them.

As stated above, Google matches the search query with its index via an automated process to show search results. This process is referred "voting." By its nature, this mathematical and algorithm-driven process is an outcome of design choices. Stated another way, the human is behind the machine-driven process. It should be emphasized that ranking search results is a political, social, and cultural choice. Another issue is that Google can filter search results. For instance, in France and Germany, it is unlawful to sell Nazi memorabilia. To comply with local laws, Google filters related content from search results.[272] To conclude, what appears in top search results, including discriminative search results, is a human engineering outcome that is not entirely neutral and objective.

[268] (Google, 2018, p. 131)

[269] According to Google, the search query "did the holocaust happen?" is used to receive "factually accurate information about the Holocaust or information about the issue of Holocaust denial." Search queries such as "racist whites", "racist blacks", etc. is used because "users are looking for information about racism among people belonging to the ethnicity mentioned in the query." (Google, 2018, p. 131)

[270] (Noble, 2018, p. 5)

[271] (Noble, 2018, pp. 31-32, 104, 179)

[272] (Noble, 2018, pp. 36-38, 45)

5.2.3. Solutions

In the following paragraphs, the research will touch upon how to address problematic search results with both legal and non-legal perspective. Google claims that search engines do not intend to be racist or sexist. At this point, the main principles of discrimination law should be recalled. Many jurisdictions do not seek discriminatory motive, purpose, or intention behind discriminatory treatment to establish the nexus between treatment and protected ground. In other words, in the perspective of law, what matters is the outcome.

Search engines is one of the most potent platforms for freedom of speech. Such power, arguably, comes with very little responsibility due to lack of sufficient regulation. This has significant impact on freedom of speech, right to privacy, and freedom from discrimination. The European Union noticed the problem and slowly includes search engines in the scope of the law. The European Court of Justice's right to be forgotten (Google Spain) case and the General Data Protection Regulation are the most important steps in that direction. Particularly, the Google Spain case may be a guiding light. In the view of the European Court of Justice, search engines are responsible for their activities (retrieving, recording, organizing, storing, and making data available as search results).[273] Perhaps, this mentality may set an example to prevent search engines discrimination and hold search engines responsible. Stated differently, search engines may be held liable due to discriminative search results. Because these problematic search results are the outcome of search engine activities. Some argue that to mitigate discriminative search results' effect, Google can include a specific disclaimer associated with search results or use a technical fix to delist them. On the other hand, these suggestions may adversely affect freedom of speech.[274]

Other suggestions to overcome search engine discrimination is to close the digital divide gap because Global North dominates digital data. In the same way, another problem is homogeneity in tech industry. Asian and White men govern Silicon Valley.[275] To illustrate, as of 2019, 31.6% of Google employees are women, and 68.4% are men; 3.3% are Black, 5.7% are Latin, 39.8% are Asian, and 54.4% are White.[276] Lack of diversity may be an important reason behind search engine discrimination. Artificial intelligence's developers should be more diversified. Because

[273] (Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González, 2014, p. 28)
[274] (Noble, 2018, pp. 155, 158)
[275] (Noble, 2018, p. 163)
[276] (Google, 2019, p. 13)

the research suggests that artificial intelligence may inherit its developers' bias.[277] Lastly, it is noteworthy to mention that Google tries to tackle search engine discrimination. The company modified its algorithms and provides a feedback tool integrated to search bar to mitigate problematic search results.[278] These practices should be enhanced and supported.

5.3. Risk Assessment Tools

In a conference, a professor asked a thought-provoking question to the Chief Justice of the Supreme Court of the United States "when smart machines, driven with artificial intelligences, will assist with courtroom fact-finding or, more controversially even, judicial decision-making?" The Chief Justice replied, "it is a day that is here… and it [artificial intelligence] is putting a significant strain on how the judiciary goes about doing things.[279]"

In recent years, there has been an increasing interest in the use of artificial intelligence in judicial environment. A recent report by the European Council indicated a large number of ways how artificial intelligence can be used in judicial systems.[280] This section engages in one of them, namely risk assessment tools, and critically traces such use with discrimination lenses.

This section firstly makes an introduction to risk assessment tools, that is followed by artificial intelligence-driven risk assessment and discrimination debates. The section lastly addresses roots causes of problems.

5.3.1. Introduction to risk assessment tools

Risk assessment tools aim to forecast the likelihood of future crime (or misconduct). It is argued that risk assessment tools may pursue various goals: (1) to predict the high-risk of recidivism so that offender can be sentenced by a more severe penalty, (2) to identify low-risk offenders to take lighter measures, (3) to take risk mitigant or preventive measures in advance concerning high-risk offenders. Turning to the risk assessment process, risk assessment tools analyze data to discover the connection between a possible future crime and selected input criteria such as age, gender, criminal records. Generally, such systems use between seven and fifteen criteria and attach specific importance to every each of them. After processing input data, risk assessment tools generate risk score and label persons as low, medium, and high risk. The end product may be used to decide upon including but not limited to the length of sentence,

[277] (Caliskan, Narayanan, & Bryson, 2017)
[278] (Toor, 2017)
[279] (Liptak, 2017)
[280] (The European Commission for the Efficiency of Justice of the Council of Europe, 2018)

parole, pretrial custody status, probation supervision levels. It should be noted that risk assessment tools are in use since the 1920's. Recent developments in the field of artificial intelligence and dramatically increasing jail population in the USA have led to a renewed interest in risk assessment tools.[281]

Risk assessment tools may provide opportunities and pose risks to human rights. Risk assessment tools supporters argue that these tools produce efficient, neutral, objective judgments and reduce incarceration rates without compromising public safety. Risk assessment proponents also claim that in decision-making process, judges may put excessive weight on extraneous factors. Also, external factors such as re-election, re-appointment, and promotion may influence judges. In addition, it is suggested that risk assessment tools may be less biased than biased judges. On the other hand, risk assessment tools critics assert that risk assessment tools are biased, pose serious threats to human rights (especially right to a fair trial and prohibition of discrimination) and do not decrease incarceration rates. Risk assessment critics point out that a judge who has enough information on a case and its surrounding factors can easily outperform an automated process rely on seven to fifteen criteria. Judges and criminal justice practitioners support this view: in a survey, less than 10% declared that risk assessment tools are better than judges on risk assessment at sentencing.[282] In case of discrimination, judges may be aware of historical bias, unlike risk assessment tools, and can adjust their judgment accordingly.

The use of risk assessment tools in courts in Europe is exceptional. As of late 2018, there is one risk assessment tool identified in Europe. The European perspective is also doubtful on the use of risk assessment tools in the courts. The European Commission's report suggests that the use of risk assessment tools can be "considered with the most extreme reservations.[283]" On the other side of the Atlantic, courts use such systems. A study conducted in 2015 identified sixty predictive tool systems in the USA.[284] It is safe to argue that the number increased since then. Therefore, the study will concentrate on the USA.

[281] (Stevenson, 2017, pp. 304, 314-316), (Monahan & Skeem, 2015, pp. 11-12), (Julia, Jeff, Surya, Lauren, & ProPublica, 2016)
[282] (The European Commission for the Efficiency of Justice of the Council of Europe, 2018, p. 54), (Stevenson, 2017, pp. 305, 326-327- 334-335)
[283] (The European Commission for the Efficiency of Justice of the Council of Europe, 2018, pp. 51-52, 66-67)
[284] (The European Commission for the Efficiency of Justice of the Council of Europe, 2018, p. 52)

5.3.2. Risk assessment tool discrimination

> "[L]aw punishes people for what they do, not who they are. Dispensing punishment on the basis of an immutable characteristic flatly contravenes this guiding principle.[285]"

As stated above, risk assessment tools forecast the possibility of recidivism based on various factors. Some of these factors are related to discrimination grounds, and this may result in disparate treatment. For instance, Pennsylvania's law enforcement uses a risk assessment tool. In risk assessment process, the tool in question uses nine criteria for risk assessment, and gender is one of the most influential factors.[286] It could be argued that gender is a vital factor in criminal risk assessment. However, the mentioned risk assessment tool's developers concluded that removing gender from input data would very marginally impact the success rate of the tool.[287] More importantly, such removal "results in fewer females classified as low risk and more females classified as high risk.[288]" As a result, gender criteria goes in women's favor. It may be true that males commit more crimes compared to females. Therefore, gender is considered an important factor that influences the possibility of recidivism. On the other hand, it is questionable to penalize the same crime differently based on an immutable discrimination ground (instead of moral culpability). Stated differently, punishing offenders differently based on a static discrimination ground may not be justifiable.

According to the former United States Attorney General, risk assessment tools "may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society.[289]" Because input of risk assessment tools may be biased due to historical bias. Risk assessment tools heavily rely upon prior arrests and convictions in risk assessment process, perhaps way more than they should. Over the years, risk assessment tools dramatically reduced the number of factors that are used as input data. As a result, prior arrests and convictions became vital input data in the risk assessment process.[290] Over-reliance on criminal records may hurt some groups due to historical bias. For example, in the USA, police arrest Black people 3.73 times more in comparison to a White people for marijuana possession; despite marijuana consumption level among Black people and White people being similar.[291] Such practices generate disparate impact that could be seen in number of prior arrests and

[285] (Buck v. Davis, 2017)
[286] (Pennsylvania Commission on Sentencing, 2018, p. 1)
[287] (Pennsylvania Commission on Sentencing, 2015, s. 1)
[288] (Pennsylvania Commission on Sentencing, 2018)
[289] (Julia, Jeff, Surya, Lauren, & ProPublica, 2016)
[290] (Harcourt, 2010, p. 7)
[291] (American Civil Liberties Union, 2013, p. 4)

convictions. The bigger picture, a comparison of demographics of sentenced prison population and general population demographics, shows the gravity of the problem. As of 2016, 12% of the USA adult population was Black, 64% was White, and 16% was Hispanic. The demographics of people behind bars is significantly different in comparison to general population demographics: 33% was Black, 30% was White, and 23% was Hispanic.[292]

Due to systemic discrimination, people of color have more prior arrests and convictions in the USA. The high level of prior arrests and convictions of people of color increases the risk to be labeled as high-risk offenders in the risk assessment process. As a result, prior criminal record may evolve into a proxy due to relatively high correlation between criminal record and race.[293] Proxy discrimination may deepen if developers put more weight than they should on criminal record.[294]

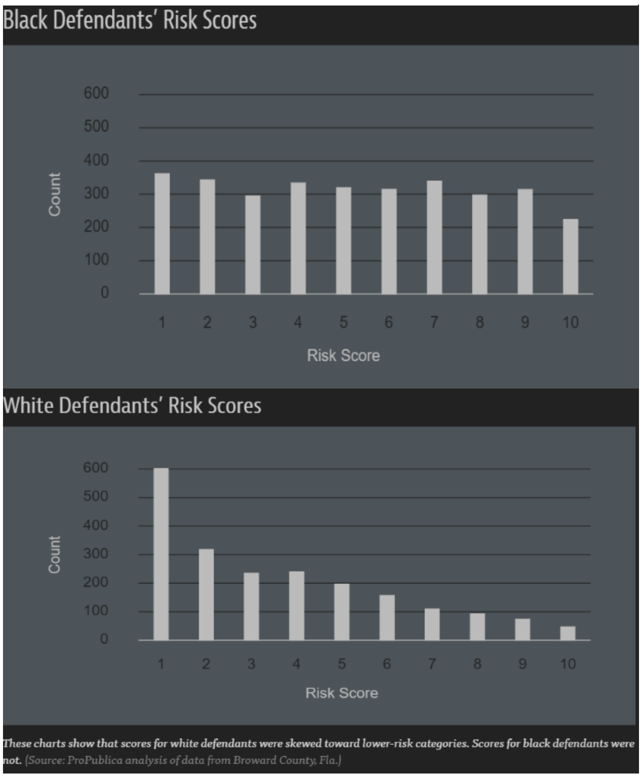
In addition to static discrimination grounds, some experts argue against the use of social and economic factors as input data to risk assessment systems. Because education, employment, and economic factors are the outcome of social and economic inequalities. Furthermore, if there is sufficient connection between socioeconomic factors and discrimination grounds, socioeconomic input data may produce proxy discrimination. Lastly, socioeconomic factors are unrelated to moral culpability; thus, experts argue that they should not be a part of risk assessment input data.[295]

Having discussed the theory, now it is time to focus on practice. In their thorough analysis of one of the most widely used risk assessment tools in the USA, COMPAS, the researchers compared actual recidivism rates and the tool's predicted recidivism rates for two years. The findings of this study suggest that the risk assessment tool in question poses many threats to human rights and may violate prohibition of discrimination.

[292] (Gramlich , 2018)
[293] (Jennifer & Christopher, 2016, p. 8)
[294] (Stevenson, 2017, p. 329)
[295] (Jennifer & Christopher, 2016, p. 7), (Stevenson, 2017, p. 328)

COMPAS generates scores for “risk of recidivism” and “risk of violent recidivism” based on answers of 137 questions that are replied by defendants or obtained from criminal records. However, race is not an input factor in risk assessment. Nevertheless, risk scores for White and Black people were significantly different. Firstly, the tool labeled Black defendants higher risk than they were. The percentage of defendants labeled higher risk yet did not re-offended was 23.5% for White defendants compared to 44.9% for Black defendants. Secondly, the tool labeled White defendants less risky than they were. The percentage of defendants labeled lower risk and re-offended was 47.7% for White defendants in comparison to 28% for Black defendants. Thirdly, Black defendants were disproportionally mislabeled for higher risk of violent recidivism. Black defendants were misclassified as two times more likely than White defendants as high risk of violent recidivism. Furthermore, by comparison with Black defendants, White defendants were 63% more misclassified as low risk of violent recidivism. The researchers controlled some inputs, prior crimes, future recidivism, age, and gender, and tested the risk assessment tool. In that case, Black defendants were 45% more likely to be labeled higher risk scores for recidivism and 77% more likely to be classified higher risk scores for violent recidivism compared to White defendants.[296] To conclude, it is plausible to argue that COMPAS (one of the most widely used risk assessment tools) exacerbates racial disparities.

Figure 7 - COMPAS’ risk score comparison:
White vs. Black defendants



These charts show that scores for white defendants were skewed toward lower-risk categories. Scores for black defendants were not. (Source: ProPublica analysis of data from Broward County, Fla.)

5.3.3. The underlying reasons of risk assessment tool discrimination

Thus far, this section has analyzed risk assessment tools with discrimination law perspective. In addition to this assessment, it is essential to touch upon interrelated problems, particularly lack of transparency and accountability.

[296] (Julia, Jeff, Surya, Lauren, & ProPublica, 2016), (Jeff, Surya, Lauren, & Julia, 2016). The figure is taken from (Julia, Jeff, Surya, Lauren, & ProPublica, 2016).

Generally, private companies develop artificial intelligence systems used in the judicial process. Research suggests that private companies refuse to disclose the functionality of their products. In other words, due to trade secrets, it becomes troublesome to gain insight into how artificial intelligence assists a court decision. This is particularly true for defendants. Defendants only receive risk scores, yet how input data converted into results is not revealed. Therefore, there is an increasing concern that it may be difficult to challenge artificial intelligent assisted court decisions. This may harm the legality of court decisions and equality of arms principle.[297]

Bringing cases against the use of risk assessment tools is easier said than done. The leading case in the USA illustrates this argument. In 2013, Eric Loomis was charged with driving a stolen car and fleeing from police. The court used a risk assessment tool in the judgment. The tool predicted that the chances of Loomis to re-offend is very high and found Loomis guilty.[298] Loomis argued that he is discriminated by the risk assessment tool, his sentence is not individualized, and not based on accurate information.[299] The Wisconsin Supreme Court rejected his due process claims.[300] The applicant tried to bring his case before the Supreme Court of the United States, yet the highest court denied the writ of certiorari.[301]

It may be strenuous to challenge risk assessment tools legally. Also, trade secrets may be a safe harbor or getaway for developers. As a result, risk assessment developers may not feel pressure to prevent disparate impacts of their tools. The overwhelming proportion of developers do not seek to find a solution to risk assessment discrimination.[302]

In Europe, it may be easier to challenge risk assessment tool decisions. As stated above, the use of risk assessment tools in Europe is exceptional, therefore not many cases brought before the courts. Experts argue that right to information and the General Data Protection Regulation provide a legal ground to challenge lack of transparency in the use of risk assessment tools.[303]

[297] (Julia, Jeff, Surya, Lauren, & ProPublica, 2016), (Rebecca, 2018), (Filippo, Hannah, Vivek, Christopher, & Levin, 2018, p. 20)

[298] (Kehl, 2017, p. 18)

[299] "Specifically, he argued that it violated due process for three reasons: (1) it violated his right to be sentenced based on accurate information because the proprietary nature of the COMPAS software prevented him from assessing the accuracy of the score; (2) it violated his right to an individualized sentence because it relied on information about the characteristics of a larger group to make an inference about his personal likelihood to commit future crimes; and (3) it improperly used "gendered assessments" in calculating the score." (Kehl, 2017, p. 18)

[300] (Harvard Law Review, 2019), (State v. Wisconsin, 2016)

[301] (Loomis v. Wisconsin, 2017)

[302] (Stevenson, 2017, p. 328)

[303] (The European Commission for the Efficiency of Justice of the Council of Europe, 2018, p. 54)

Lastly, it is noteworthy to mention that it is uncommon but possible to use machine learning to develop risk assessment tools. Developers can decide on inputs, outputs, and preferred machine learning method. After that, artificial intelligence can create a risk assessment tool by itself. Put differently, artificial intelligence can learn risk assessment process without any human intervention. It is highly likely that such method generates serious legal and moral problems. Because artificial intelligence's learning pattern may be greatly complicated (even too intricate for the designers of the artificial intelligence).[304] Hence, sometimes machine learning referred as to black box.[305] Even though authorities that use artificial intelligence are willing to explain the underlying reasoning of artificial intelligence assisted judgments, it may not be possible.

5.4. Conclusions

This chapter attempted to show artificial intelligence bias in public domain. In the light of the findings, the study proposes three hypothesizes.

1) The findings show that some of the most important artificial intelligence applications tend to discriminate the most discriminated groups (which confirms the findings of Chapter 4). As stated above, facial recognition systems perform perfectly on males and light skin colors, fail to identify people of color, particularly Black women. Search engines amplify gender stereotypes; a simple search query can result in sexist, racist, anti-Semitic, and Islamophobic search results. Risk assessment tools mislabel people of color as high-risk offenders.

2) Artificial intelligence is used in diverse domains, and many of them remain largely or entirely unregulated, which multiples problems. Lack of transparency, accountability, standards, audits, and procedures to address the problems in the development, deployment, and use of artificial intelligence is alarming. Arguably, they stem from lack of regulation.

3) Confusion around legal liability of artificial intelligence's actions and non-transparent use of AI encumber one's ability to bring discrimination claims before courts. Artificial intelligence developers and users are almost immune from violations of prohibition of discrimination. In addition to need for new regulation, progressive interpretation of existing laws is necessary to prevent AI discrimination and provide remedy for victims.

[304] (Alex, 2016), (Ian, 2017)

[305] Black box can be loosely described as "[a] model that is opaque to its user; although the model can produce correct results, its internal relationships are not known." (Negnevitsky, 2005, p. 367)

CHAPTER 6: ARTIFICIAL INTELLIGENCE AND REGULATION

This chapter serves as an introduction to artificial intelligence regulation. It firstly discusses the reasons to regulate AI, which is followed by an analysis of who should regulate it. Another objective of the chapter is to propose preliminary suggestions on how to regulate artificial intelligence to tackle machine discrimination.

6.1. To Regulate Artificial Intelligence Before It Regulates Everything

> We shape artificial intelligence and afterwards it shapes us.[306]

> "Where AI is discussed in such a broad way, there is a tendency to assume that the technology poses challenges that are so radically new that all existing laws, regulations and standards are no longer applicable or appropriate. The 'flipside' of that discourse is to demand regulation of the technology itself, regardless of how and where it is applied.[307]"

This section focuses the debate around why artificial intelligence should be regulated and who is responsible for its regulation. Firstly, the section will discuss possible adverse impacts of artificial intelligence on equality and democracy. Subsequently, the spotlight will be on whether lawmakers or artificial intelligence industry should regulate AI.

To understand what is at stake: As eluded to, there may be many good reasons to regulate artificial intelligence. Due to limitations, only two of them will be addressed in the following paragraphs, artificial intelligence's potential adverse impacts on equality and democracy.

With respect to equality, the discussion starts with a historical perspective. The Enlightenment is a cornerstone in the history of humanity that established philosophical foundations for equality. Another turning point for equality was the Industrial Revolution because factories depended on a large number of healthy workers. The next decisive moment for equality was the 20th century, which humanity witnessed two tragedies, the first and second world war. Governments relied on millions of soldiers and workers in front and assembly line. Owing to the need for a great number of workers and soldiers, governments had to invest in masses to keep their soldiers loyal and workers healthy. In other words, in the 20th century, masses became critical due to military and industrial reasons. As a result of pragmatic choice, governments started to invest in health, education, and welfare of masses, which reduced inequality. It is argued that artificial intelligence will change the course of history by eliminating the dependency on masses. Stated differently, armed forces may not rely on

[306] This sentence is an adjusted form of Winston Churchill's famous quote "We shape our buildings and afterwards our buildings shape us." The original quote is taken from (Churchill and the Commons Chamber, 2019)
[307] (ARTICLE 19, 2018, p. 20)

millions of soldiers as a result of autonomous warfare; production may not depend on a large number of workers by dint of automated manufacturing.[308]

Let us take a closer look at artificial intelligence's possible influence on the job market. There is a good chance that the Fourth Industrial Revolution will affect some of the most common jobs. In near future, there may not be a need for a cashier because customers can shop at automated supermarkets without checkout. Fast food restaurants started to replace food preparation and cooking workers with ordering kiosks and burger-flipping robots.[309] It is highly likely that driverless cars and trucks will substitute taxi drivers and truck drivers.[310] An Oxford University study found that in ten to twenty years, 47% of jobs in the USA can be automated. The study claims that most of transportation, logistics, office and administrative support, and labor in production occupations are under risk owing to job automation.[311]

It is true that artificial intelligence will create new jobs. Most likely, these jobs will demand specific skills. The World Economic Forum's report observes that technology-related and non-cognitive soft skills will become more important in the Industry 4.0's job market and addresses the importance of lifelong learning.[312] Acquiring new skills and adaptation can be easier for new generations. However, it may be tough for a blue-collar middle-aged person to learn these advanced skills, particularly after losing his/her job to a robot.

Both governments and companies may not depend on masses in future. Arguably, a small group of highly trained people will be able to build up an autonomous army and operate a mass production line. Put differently, institutions that were traditionally dependent on masses will not rely on them. There is a good chance that AI will take over the most common jobs, which will remove many from their profession. Also, it is possible that many unemployed people will not be able to develop necessary skills that the future requires. It is argued that it may create a useless class. On the other hand, only a handful of elites may reap the harvest of artificial intelligence. This may shake the foundations of equality.

> "Once the masses lose their economic importance and political power, the state loses at least some of the incentive to invest in their health, education and welfare. It's very dangerous to be redundant. Your future

[308] (Yuval, Are we about to witness the most unequal societies in history?, 2017)
[309] (John, 2017)
[310] (Olivia, 2016)
[311] (Carl & Michael, 2013, pp. 47-48)
[312] (Centre for the New Economy and Society, 2018, pp. 22-23)

> depends on the goodwill of a small elite. Maybe there is goodwill for a few decades. But in a time of crisis – like climate catastrophe – it would be very tempting, and easy, to toss you overboard.[313]”

To conclude, artificial intelligence may eliminate some practical foundations of equality. Bearing the mind that equality and non-discrimination are interdependent and interrelated, these possible developments may have drastic effects on prohibition of discrimination.

Having outlined artificial intelligence’s potential adverse effects on equality, now the section turns to artificial intelligence’s possible harmful impact on democracy. The first observations cover freedom of speech, an indispensable foundation of democratic society. The adverse impacts of artificial intelligence on freedom of speech are already visible. Artificial intelligence became a tool to disseminate fake news owing to its unique ability to propagate online content. On macro level, it can use numerous bots to circulate fake news or hate speech to a broad audience; on micro level AI can profile and target specific individuals or groups.

As machine learning advances, it becomes harder to distinguish between AI and human-produced content. Researchers raised concerns on the use of artificial intelligence to write fake news.[314] It is highly likely that artificial intelligence will be able to write and propagate fake news by itself in near future. Bearing the mind that significant proportion of society is exposed to information overload, receives news from social media, and lacks media and digital literacy, fake news can undermine democracy. In addition to fake news, the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression raised his concerns about other adverse impacts of artificial intelligence on freedom of speech. The Rapporteur claimed that filter bubble and artificial intelligence-driven online content moderation may impair human rights and democracy.[315]

Artificial intelligence may provide the ability to counteract democracy. It is true that the way humanity uses technology may turn it into a destructive or good tool. On another note, technology may bring competences that can be useful for good purposes and more beneficial for evil intentions. That may be the case of how democracy or authoritarian power can use artificial intelligence. In other words, it is argued that artificial intelligence tends to favor authoritarian regimes over democracy.

[313] (Yuval, Are we about to witness the most unequal societies in history?, 2017)
[314] A leading AI research center claimed that it developed an extraordinarily successful AI text generator. Due to its potential to generate fake news, the researchers refused to release the research. (Alec, et al., 2019)
[315] (David K. , 2018)

Authoritarian regimes aim to centralize information and decision-making power, which raises the issue of ineffectiveness. As such, limited number of decision makers had difficulties in processing a great volume of data. Thus, authoritarian regimes based their decisions on missing or incorrect information. On the other hand, democracies tend to decentralize decision-making process, which enables them to process sheer volume of data. It increases the quality of decisions and effectiveness.[316] The study proposes the hypothesis that artificial may turn the tide.

Artificial intelligence offers a solution to an essential problem of authoritarian regimes. It enables to process a great volume of data centrally. AI encourages the centralization of data because more extensive training databases can improve the outcome of data processing. To illustrate, if an authoritarian regime like China collects its 1.4 billion citizens medical data and process it by virtue of artificial intelligence, it can gain serious advantage in medical research. On the other hand, a democracy based on separation of powers and respects human rights (especially right to privacy) is arguably hesitant to collect sensitive personal data (such as medical data) and centrally process it. To conclude, "[t]he main handicap of authoritarian regimes in the 20th century—the desire to concentrate all information and power in one place—may become their decisive advantage in the 21st century.[317]"

Beyond doubt, artificial intelligence provides unique opportunities. Also, it may pose serious challenges that society cannot foresee at present. The findings of minimal research showed some of these issues. Artificial intelligence is capable of eroding democracy and equality, two foundations of prohibition of discrimination. The thesis argues that the gravity of possible problems is too critical to be left unregulated. Possible developments in AI may create more questions than answers, and it very challenging to find legal solutions. Therefore, policymakers should be involved in AI regulation process, cogitate, and devote resources to prepare humanity for the 21st century.

Ethics vs. laws: This section has analyzed the possible adverse causes of artificial intelligence and argued that AI should be regulated. The next part examines who should regulate artificial intelligence.

Generally, opinions divide into two on regulating newly emerging technologies. Many believe that brand-new technologies can create unique challenges that may be difficult to bring within

[316] (Yuval, Why Technology Favors Tyranny, 2018)
[317] (Yuval, Why Technology Favors Tyranny, 2018)

scope of law. There may be other flipsides of regulation. It may decelerate the advancement of technology, quickly become outdated, and be challenging to implement. To tackle these issues, some support the idea of self-regulation by the technology developers or users and object to lawmakers' intervention. On the other hand, to leave advancing technologies beyond the scope of law, one of the foundations of society, may generate serious issues: lack of order, misuse of technology, and (unpunished) human rights violations. Thus, despite the difficulties, lawmakers tend to regulate new technologies. The rapid advancement of artificial intelligence in the last decade provoked this old discussion with a new object: who should regulate artificial intelligence?

Sundar Pichai, Google's CEO, observed that policymakers follow developments from behind and still try to comprehend the effects of artificial intelligence.[318] This may be true due to expeditious development, complexity, and diversity of artificial intelligence. The lack of qualified technology literacy, legal conservatism, narrow-minded politicians, and shallow politics also becloud artificial intelligence regulation. In addition, lawyers and policymakers have paid too little attention to artificial intelligence. Consequently, despite diversified and wide use artificial intelligence, most of the fields artificial intelligence involved remain largely or entirely unregulated.

Artificial intelligence developers noticed the absence of policymakers and have acted accordingly. There has been a growing interest in non-binding guidelines and ethical frameworks in the last years.[319] In other words, nowadays leading multinational corporations (Apple, Amazon, Google, Facebook, IBM, and Microsoft) "study and formulate best practices on AI[320]" so that they can fulfill the space. A researcher noticed a critical issue and raised concerns on this movement.

> "A strange confusion among technology policy makers can be witnessed at present. While almost all are able to agree on the common chorus of voices chanting 'something must be done,' it is very difficult to identify what exactly must be done and how. In this confused environment it is perhaps unsurprising that the idea of 'ethics' is presented as a concrete policy option. Striving for ethics and ethical decision-making it is argued, will make technologies better. While this may be true in many cases, much of the debate about ethics seems increasingly focused on private companies avoiding regulation. Unable or unwilling to properly provide regulatory solutions, ethics is seen as the 'easy' or 'soft' option which can

[318] (Romm, Timberg, & Harwell, Google CEO Sundar Pichai: Fears about artificial intelligence are 'very legitimate,' he says in Post interview, 2018)
[319] (ARTICLE 19, 2018, p. 12) (Borgesius, 2018, p. 27)
[320] (partnershiponai, 2019)

> help structure and give meaning to existing self-regulatory initiatives. In this world, 'ethics' is the new 'industry self-regulation.'[321]"

This raises the question of why developers prefer regulating themselves and oppose traditional regulation. There may be several reasons behind it. The risky nature of artificial intelligence development may be one of the causes. Artificial intelligence developers devote serious resources to development: a long, risky, demanding process that may generate fruitful outcomes. Also, venture capitalists see high potential profit and invest significant capital in artificial intelligence industry. This raises the pressure on developers to produce results swiftly. In the absence of regulation, under-pressured developers look ways to cut corners, which may accelerate development process. It comes at a price, violating human rights, which can be illustrated by facial recognition industry. It is not coincident that some of the leading companies in facial recognition industry are from China and Israel.[322] In China, right to privacy and data protection are not respected, particularly in Xinjiang, the homeland of Uighur Muslim minority. Consequently, Chinese companies test facial recognition systems on the Uighur Muslim minority as they please.[323] In a similar fashion, Israeli companies freely try facial recognition systems in occupied Palestinian territories.[324] These actions may create advantage to developers at the cost of severe human rights violations.

It could be argued that today's artificial intelligence research, to a certain degree, is driven by economic interests and dehumanizing algorithmic efficiency. As a result, research suggests that artificial intelligence is, to some extent, progressed without transparency, accountability, and respect for human rights.[325] It should be noted that accelerated development can increase costs, inaccurate artificial intelligence decisions, and public distrust in the long term (as exemplified in Section 5.1. and 5.3.). On the other hand, research argues that regulating artificial intelligence and integrating human rights-based approach to artificial intelligence research can mitigate these effects. Thereby, human rights-based approach to AI development may increase profitability in the medium to long term.[326]

Having discussed the problems of unregulated development procedure, now this study focuses on market dynamic because regulation can dramatically influence it. Some developers are

[321] (Wagner, 2018, p. 1)
[322] (Xiang, 2018), (Hagar & Jonathan, 2018), (Blake & Venus, 2019)
[323] See Section 4.2.1. for details.
[324] (David, 2018), (Hagar & Jonathan, 2018), (Alex K. , 2016)
[325] (ARTICLE 19, 2018, p. 15)
[326] (Mantelero, 2019, p. 6)

against traditional regulation because it may narrow artificial intelligence market down. Let us consider autonomous weapons which can create a large market. Autonomous weapons also pose serious threats to human rights and can violate main principles of humanitarian law.[327] Therefore, leading academics, many Member States of the United Nations, and the United Nations Secretary-General called for a ban on autonomous weapons.[328] If this movement succeeds, regulation may terminate a large possible market.

Regulation brings standards, oversight, and accountability, which may influence customer behavior. This can be exemplified by largely unregulated and profitable facial recognition market in the USA. Law enforcement in the USA is one of the primary customers of facial recognition. They also do not require any quality assurance checks (accuracy threshold for facial recognition systems) before or after purchase.[329] This takes the pressure off developers to refine products. Possibly, the key factor of this market dynamic is lack of regulation because there are no standards on purchase and use of facial recognition.

Let us address the controversy between self-regulating and regulating artificial intelligence from another angle. It is not the first time that technology users and developers desire to deregulate a transformative new technology. This is evident in the case of the Internet. The debates in the course of the rise of the Internet divided public opinion. Many supported to regulate the Internet and others disagreed. Unregulated Internet supporters believed that traditional forms of government and laws did not fit this newly emerging technology. They suggested that the online community should establish its rules (which should not violate vital interests of non-Internet users). And the Internet was not regulated as it should have been for a long time due to various complex reasons.[330]

Humanity expected the unregulated Internet to be a global democratizing force and to a certain degree, it did so. On the other hand, through the progress of the Internet, developers were able to put technology before democracy and human rights owing to lack of regulation. The motto was “move fast break things” which gave rise to surveillance capitalism (an economic system founded on monetizing personal data) and unpunished large data-breaches. Due to lack of regulation, Silicon Valley asked forgiveness than permission for its fiascos like Cambridge Analytica data scandal. Developers were not the only actor misused the Internet; populists also

[327] (Maya, 2017)
[328] (stopkillerrobots.org, 2019), (researchers, 2019)
[329] (Garvie, Frankle, & Bedoya, 2016, p. 47)
[330] (Nemitz, 2018, pp. 1-8)

enjoyed it. They managed to turn the Internet into a platform to spread hate speech and populist ideology. The governments also took advantage of unregulated Internet and executed illegal mass surveillance.[331] The online community's or developers' rules were not sufficient to prevent these problems. It is also highly questionable whether artificial intelligence developers' self-regulation will be adequate to protect human rights from adverse impacts of AI. The effectiveness of self-regulation becomes more dubious bearing the great potential of artificial intelligence. One should also address that many frontrunner artificial intelligence developers are profit-driven large tech companies that dominated unregulated Internet and systematically violate human rights (especially right to privacy).

Another issue concerning self-regulation is it is neither forcible nor democratic. On the other hand, law is enforceable and a product of democratic process. To exemplify, a leading regulation related to AI is General Data Protection Regulation. Four thousand stakeholders contributed to the preparation process of it, and selected representatives passed the bill. In case of severe violations of General Data Protection Regulation, authorities can impose serious penalties such as 4% of annual global turnover of violator, which makes it forcible against responsible actors. Ethics, on the other hand, is neither forcible nor a product of democratic process.[332]

6.2. How to Regulate Artificial Intelligence to Prevent Machine Discrimination

This section touches upon recent attempts to regulate AI and makes preliminary suggestions on how to regulate artificial intelligence to prevent machine discrimination.

Some policymakers started to bring AI into scope of laws despite the difficulty of the task. General Data Protection Regulation, Guidelines on Artificial Intelligence and Data Protection by Convention 108, and New York City's law on automated decisions (Local Law 49 of 2018[333]) are notable examples in the field. Some important bills concerning artificial intelligence such as AI JOBS Act[334] (aims to assess AI's impact on workforce), Innovation Corps Act[335] (intends to assist workers replaced by AI), FUTURE of Artificial Intelligence Act[336] (establishes a committee to advise issues relating to the development of artificial intelligence) introduced to the United States Congress in the last few years.

[331] (Nemitz, 2018, pp. 1-8), (ARTICLE 19, 2018, p. 13)
[332] (Nemitz, 2018, pp. 17-18)
[333] (The New York City Council, 2019)
[334] (Darren S. , 2019)
[335] (Doris, 2019)
[336] (John K. D., 2019)

With regard to how to regulate AI, unfortunately, a full discussion lies beyond the scope of this study. Due to practical constraints, the study only makes five preliminary suggestions on how to regulate AI to prevent machine discrimination.

Sector-specific regulation and oversight: Artificial intelligence is used in various domains such as healthcare, finance, automotive, marketing, government, transportation, and military. Artificial intelligence involved fields have distinctive characteristics and require expertise. A regulation that fits an AI-involved field (such as automotive) most likely may not be appropriate for another AI domain (e.g., medicine). Due to the same reason, a central artificial institution may lack expertise and face difficulties to oversee different sectors. Sector-specific regulation and oversight may be more appropriate to prevent artificial intelligence discrimination. SELF DRIVE ACT,[337] a law that solely concerns autonomous vehicle regulation in the USA, may set an example to sector-specific AI regulation.

Multi-layered and technology neutral regulation: AI regulation should be technology neutral and focus on the results of the technology instead of technology itself and leave room for maneuver for responsible actors to achieve desired goals. Regulation may remain in force longer, and implementation of it can be easier thanks to this approach. Multi-layered rules include laws, regulations, self-regulation, resolutions, guidelines, and ethics may be required to create technology neutral regulation. Rules higher in the hierarchy should be broader, and others can fill the gaps and answer practical needs of daily life.[338]

Holistic legal reform: Research demonstrates that the USA discrimination law has many shortcomings to prevent artificial intelligence discrimination.[339] Although the EU discrimination laws are more prepared compared to United States law, it also falls short to prevent AI discrimination.[340] The issue at hand is discrimination law by itself may be inadequate to prevent AI discrimination. Collaboration of different areas of law (particularly discrimination, intellectual property, and data protection laws) may be needed to tackle AI discrimination.[341] Therefore, a holistic approach is needed to reform various fields of law.

Algorithmic transparency: Unrevealing artificial intelligence discrimination is already a challenging task owing to AI's black box characteristics. In addition, AI developers do not

[337] (Robert, 2019)
[338] (Borgesius, 2018, p. 33)
[339] (Barocas & Selbst, 2016, pp. 694-714)
[340] (Hacker, 2018, pp. 12-24)
[341] (Hacker, 2018, pp. 24-34)

provide any information on functionality of their products due to trade secrets. Developers can dismiss information requests as a result of intellectual property law. In some cases, this provides a safe harbor to AI developers and removes transparency. Striking a balance between intellectual property law and right to access to information is necessary to establish algorithmic transparency. This is particularly true for AI made or assisted decisions in justice, welfare, and healthcare.[342]

Providing effective remedies: The confusion around legal liability of AI obstructs providing effective legal remedies to victims. Policymakers should establish clear lines of responsibility for every phase of AI lifecycle: the development, deployment, and usage. Different branches of law, including but not limited to civil, administrative, and criminal law may be used to provide effective remedies.[343]

6.3. Conclusions

The main concerns of this chapter were to show the need to bring AI into scope of laws and how to regulate AI. The chapter started with a discussion on AI's potential harmful effects on equality and democracy. The study argued that AI may eliminate some practical foundations of equality. It is also claimed that artificial intelligence may be more useful for authoritarian regimes in comparison to democracies, and it may erode democracy. The chapter concludes that AI may put democracy and equality (two main pillars of prohibition of discrimination) in peril.

The chapter continued with the debate on whether AI industry or lawmakers should regulate AI. The findings show that policymakers have paid too little attention to AI and artificial intelligence industry tries to fill emptiness by proposing non-binding guidelines and ethical frameworks. Moreover, in the absence of regulation, AI research progresses without transparency, accountability, and respect for human rights. Another observed problem is the leading AI developers are profit-driven multinational companies that systematically violate human rights, particularly right to privacy. And considering regulation may narrow artificial intelligence market down, it is understandable that major AI developers prefer self-regulation (an undemocratic and non-forcible form of regulation) over lawmakers' involvement in regulation process. However, deregulating transformative technologies create serious issues which illustrated by the rise of unregulated Internet. The research concludes that without strong

[342] (Council of Europe Commissioner for Human Rights, 2019, pp. 9-10)
[343] (Council of Europe Commissioner for Human Rights, 2019, pp. 13-14)

laws and institutional backing, there is fair chance that AI may not serve to public good[344] and notes that that self-regulation may not be sufficient to protect human rights.[345]

The last section briefly mentioned some AI laws and bills, also made preliminary suggestions on regulation to prevent AI discrimination. The research recommends that AI regulation should be sector specific, multi-layered, technology neutral, include various fields of law in harmony, and aim to provide algorithmic transparency and effective remedies for victims.

[344] (Nemitz, 2018, pp. 1-8), (ARTICLE 19, 2018, p. 13)

[345] (Borgesius, 2018, p. 27)

CONCLUSIONS

One of the main pillars of this thesis is artificial intelligence, and a generally accepted definition of it is lacking. Because as a set of sciences and systems, AI is used in different domains and doctrine adopts different approaches it. Additionally, it is burdensome to define intelligence, which makes it harder to describe an artificial version of it. Turning to its development, AI is a young and multidisciplinary field of study that rapidly advanced, particularly in the last decade. At present, it eases daily life and deployed in different areas such as healthcare, finance, transportation, marketing, government, and military. The future of AI, most particularly strong AI, is a disputed topic. That being said, it is plausible to say that AI will continue to develop and increase its influence in more areas.

The other central pillar of the thesis is discrimination. Although it remains a poorly defined term, there is a consensus on the importance of the prohibition of discrimination owing to its close connection to equality and enjoyment of a wide range of human rights. The prohibition of discrimination is protected under international human rights law, and its scope becomes more inclusive thanks to progressive interpretation of laws. Nevertheless, the effectiveness of discrimination laws is very much depended on economic, historical, cultural, and political factors.

As regards to artificial intelligence discrimination, the future and the drawbacks of artificial intelligence are already here and require further attention, especially from legal researchers. As shown in Chapter 4, AI can discriminate in various ways. In the wild, AI tends to discriminate the most vulnerable and discriminated groups, such as women, people of color, the LGBTQI community, and ethnic minorities. Also, AI's advanced ability to identify and profile masses can turn it into a very harmful tool in the wrong hands.

Democracy and global freedom are in decline for more than a decade.[346] Populist rhetoric and far-right politicians becoming more influential actors at universal level, and they magnify discrimination. As unfolded in this thesis, AI can take discrimination into a higher tier. It can amplify stereotypes, mask systemic discrimination, profile and target minority individuals and groups. Society needs precautions to prevent AI's adverse effects, especially in the era of rising populism.

[346] (Freedom House, 2019, pp. 4-9)

Society and decision makers tend to over trust and over-rely on artificial intelligence. AI is far from being faultless because it is product of human design and data, and neither of these is perfect. Additionally, the complexity of AI and rapid advancement in technology make it challenging to understand it. In some cases, artificial intelligence developers cannot explain AI due to its “black box” characteristics. Therefore, unfolding AI-driven human rights violations may be an uphill task. Society and decision makers should increase their awareness of the flip sides of AI. Increasing AI literacy can be useful to overcome this problem.

Artificial intelligence remains mostly unregulated at present, and it escalates problems. Lack of regulation gives rise to lack of transparency, accountability, standards, audits, and procedures in the development, deployment, and usage of AI. Owing to lack of regulation, the goodwill of AI developers and users and advancing human rights in business are critical to prevent AI-driven discrimination in the short term.

The research asserts that artificial intelligence may eliminate practical foundations of equality and erode democracy, and AI developers’ self-regulation is not capable of averting it. Recalling two cornerstones of prohibition of discrimination are in peril, this research invites lawmakers to regulate artificial intelligence. This study makes preliminary suggestions on regulation to prevent AI discrimination, and advocates holistic, sector-specific, multi-layered, technology-neutral laws that can provide algorithmic transparency and effective remedies for victims.

Recommendations for Future Research

Artificial intelligence is a very dynamic domain, which makes it difficult to foresee future developments. Seeking solutions for artificial intelligence’s adverse impacts, including AI discrimination, is a challenging, vital, and continuous task. This thesis, as an exploratory study, searches for answers to tackle machine discrimination, and it raises more question than answers. Reaffirming the knowledge gap on artificial intelligence effects on human rights, future studies with human right perspective on AI discrimination are recommended.

Due to dynamic character of artificial intelligence, it may be necessary to follow developments of AI and its surrounding fields. In future investigations, it may be useful to take account of some developing technologies, especially in the areas of information technologies and communications. Some technologies such as 5G (the fifth-generation cellular network technology), the Internet of things, and quantum computers may significantly influence AI and requires the attention of researchers.

This research could not examine how policymakers should regulate AI in detail. At present, there is abundant room for further progress in AI regulation, a complex and debated topic. Following AI regulations, bills, the Council of Europe’s policies and publications on AI, and national AI development strategies may be useful for future research.

Bibliography

- (2019, February 12). Retrieved from partnershiponai: https://www.partnershiponai.org/
- (2019, March 27). Retrieved from gendershades.com: http://gendershades.org/overview.html
- *24th Council of Europe Conference of Directors of Prison and Probation Services (CDPPS)*. (2019, May 17). Retrieved from Council of Europe: https://www.coe.int/en/web/artificial-intelligence/-/24th-council-of-europe-conference-of-directors-of-prison-and-probation-services-cdpps-
- A Gorup of Human Rights NGOs. (2018, April 26). *Letter to Axon AI Ethics Board regarding Ethical Product Development and Law Enforcement*. Retrieved from civilrights.org: http://civilrightsdocs.info/pdf/policy/letters/2018/Axon%20AI%20Ethics%20Board%20Letter%20FINAL.pdf
- AI Index. (2018). *AI Index 2018 Report.* Stanford: AI Index Steering Committee,Human-Centered AI Initiative, Stanford University.
- Alec, R., Jeffrey, W., Dario, A., Daniela, A., Jack, C., Miles, B., & Ilya, S. (2019, February 14). *Better Language Models and Their Implications*. Retrieved from OpenAI: https://openai.com/blog/better-language-models/
- Alec, T. (2018, November 8). *The 2018 midterm vote: Divisions by race, gender, education*. Retrieved from Pew Research Center: https://www.pewresearch.org/fact-tank/2018/11/08/the-2018-midterm-vote-divisions-by-race-gender-education/
- Alex, H. (2016, June 28). *Google says machine learning is the future. So I tried it myself.* Retrieved from Guardian: https://www.theguardian.com/technology/2016/jun/28/google-says-machine-learning-is-the-future-so-i-tried-it-myself
- Alex, K. (2016, October 17). *HOW ISRAEL BECAME A HUB FOR SURVEILLANCE TECHNOLOGY*. Retrieved from theintercept: https://theintercept.com/2016/10/17/how-israel-became-a-hub-for-surveillance-technology/
- Alexander, L. (2016, April 8). *Do Google's 'unprofessional hair' results show it is racist?* Retrieved from Guardian: https://www.theguardian.com/technology/2016/apr/08/does-google-unprofessional-hair-results-prove-algorithms-racist-
- Allen, A. (2016, July 10). *The 'three black teenagers' search shows it is society, not Google, that is racist*. Retrieved from Guardian: https://www.theguardian.com/commentisfree/2016/jun/10/three-black-teenagers-google-racist-tweet
- Amazon. (2018, November 21). *Amazon Rekognition announces updates to its face detection, analysis, and recognition capabilities*. Retrieved from Amazon: https://aws.amazon.com/tr/about-aws/whats-new/2018/11/amazon-rekognition-announces-updates-to-its-face-detection-analysis-and-recognition-capabilities/

- Amazon. (2019, April 4). *Amazon Rekognition Developer Guide Searching Faces in a Collection Use Cases that Involve Public Safety*. Retrieved from Amazon: https://docs.aws.amazon.com/en_us/rekognition/latest/dg/considerations-public-safety-use-cases.html
- American Civil Liberties Union. (2013). *The War on Marijuana in Black and White.* New York: American Civil Liberties Union.
- American Civil Liberties Union. (2019, January 15). *PRESSURE MOUNTS ON AMAZON, MICROSOFT, AND GOOGLE AGAINST SELLING FACIAL RECOGNITION TO GOVERNMENT*. Retrieved from American Civil Liberties Union: https://www.aclu.org/news/pressure-mounts-amazon-microsoft-and-google-against-selling-facial-recognition-government
- Amit, D., M. C., & Anupam, D. (2015). Automated Experiments on Ad Privacy Settings A Tale of Opacity, Choice, and Discrimination. *Proceedings on Privacy Enhancing Technologies*, 92-112.
- Andre, E., Brett, K., Roberto A., N., Justin, K., Susan M., S., Helen M., B., & Sebastian, T. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature International Journey of Science*, 115-118.
- Andrew, P. (2019, May 9). *Gerrymandering, explained*. Retrieved from Vox: https://www.vox.com/2014/8/5/17991934/gerrymandering-explained
- Angwin, J., & Terry, P. J. (2016, October 28). *Facebook Lets Advertisers Exclude Users by Race*. Retrieved from Propublica: https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race
- ARTICLE 19. (2018). *Privacy and Freedom of Expression in the Age of Artificial Intelligence.* ARTICLE 19.
- Bailey, S. P. (2015, July 24). *Skipping church? Facial recognition software could be tracking you*. Retrieved from The Washington Post: https://www.washingtonpost.com/news/acts-of-faith/wp/2015/07/24/skipping-church-facial-recognition-software-could-be-tracking-you/?noredirect=on&utm_term=.4129bb5b8e35
- Barker, V. (2016, January 18). *Policing Difference.* Retrieved from SSRN: https://ssrn.com/abstract=2717138
- Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review, Volume 104*, 671-732.
- Biao v. Denmark, 38590/10 (The European Court of Human Rights (Grand Chamber) May 24, 2016).
- Big Brother Watch. (2018). *Face Off The lawless growth of facial.* Big Brother Watch.

- Blaise, A. y., Alexander, T., & Margaret, M. (2018, January 11). *Do algorithms reveal sexual orientation or just expose our stereotypes?* Retrieved from Medium.com: https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477
- Blake, S., & Venus, F. (2019, February 21). *The Companies Behind China's High-Tech Surveillance State*. Retrieved from Bloomberg: https://www.bloomberg.com/news/articles/2019-02-21/the-companies-behind-china-s-high-tech-surveillance-state
- Bloch, D. A. (2018, September 22). *Recipe for Quantitative Trading with Machine Learning.* Retrieved from SSRN: https://ssrn.com/abstract=3232143
- Borgesius, F. Z. (2018). *Discrimination, artificial intelligence, and algorithmic decision-making.* Strasbourg: Directorate General of Democracy, Council of Europe.
- Bosker, B. (2013, January 22). *SIRI RISING: The Inside Story Of Siri's Origins — And Why She Could Overshadow The iPhone*. Retrieved from Huffingtonpost: https://www.huffingtonpost.com/2013/01/22/siri-do-engine-apple-iphone_n_2499165.html
- Bostrom, N., & Müller, V. C. (2016). Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In N. Bostrom, & V. C. Müller, *Fundamental Issues of Artificial Intelligence* (pp. 553-571). Berlin: Springer.
- Box, G., & Draper, N. (1987). *Empirical Model-Building and Response Surfaces.* Wiley.
- Brandom, R. (2018, August 29). *How Should We Regulate Facial Recognation?* Retrieved from The Verge: https://www.theverge.com/2018/8/29/17792976/facial-recognition-regulation-rules
- Buck v. Davis, 15-8049 (Supreme Court of the United States February 22, 2017).
- Buolamwini, J. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability, and Transparency* (pp. 1-15). New York: Proceedings of Machine Learning Research (81).
- Cadwalladr, C. (2016, December 4). *Google, democracy and the truth about internet search*. Retrieved from Guardian: https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook
- Caliskan, A., Narayanan, A., & Bryson, J. J. (2017, April 1). *Semantics derived automatically from language corpora contain human-like biases*. Retrieved from Researchgate: https://www.researchgate.net/publication/316973825_Semantics_derived_automatically_from_language_corpora_contain_human-like_biases
- Cambridge Dictionary Online. (2019, January 23). *Artificial Intelligence*. Retrieved from Cambridge Dictionary Online:

https://dictionary.cambridge.org/us/dictionary/english/artificial-intelligence?q=artificial+intelligence+

- Canon, G. (2019, February 2019). *How Taylor Swift showed us the scary future of facial recognition*. Retrieved from Guardian: https://www.theguardian.com/technology/2019/feb/15/how-taylor-swift-showed-us-the-scary-future-of-facial-recognition
- Carl, B. F., & Michael, A. O. (2013). *The Future of Employment: How susceptible are jobs to computerisation?* Oxford: Oxford University.
- Carty, S. S. (2011, May 31). *Many Cars Tone Deaf to Women's Voices*. Retrieved from Auto Blog: https://www.autoblog.com/2011/05/31/women-voice-command-systems/?guce_referrer=aHR0cHM6Ly93d3cuY2F0YWx5c3Qub3JnL3Jlc2VhcmNoL3RvcGljLWJyaWVmLWdlbmRlci1iaWFzLWluLWFpLw&guce_referrer_sig=AQAAACWnqbMftD7fJTcfzF0jGcJQjFyjJe6LlBRA3bJA63X5zy1G3loHUVt7dIJsEEdlGYc
- Castellanos, S. (2019, March 14). *HR Departments Turn to AI-Enabled Recruiting in Race for Talent*. Retrieved from The Wall Street Journal : https://www.wsj.com/articles/hr-departments-turn-to-ai-enabled-recruiting-in-race-for-talent-11552600459
- Catherine, D., & Lauren, K. (2019). *Data Feminism.* Cambridge: MIT Press. .
- Centre for the New Economy and Society. (2018). *World EconomicForum The Future of Jobs Report 2018.* Geneva: World Economic Forum.
- Chan, T. F. (2018, May 21). *A school in China is monitoring students with facial recognition technology that scans the classroom every 30 seconds*. Retrieved from Businessinsider: https://nordic.businessinsider.com/china-school-facial-recognition-technology-2018-5?r=US&IR=T
- Chen, B. X. (2009, December 22). *HP Investigates Claims of 'Racist' Computers*. Retrieved from Wired: https://www.wired.com/2009/12/hp-notebooks-racist/
- CHEZ Razpredelenie Bulgaria" AD v. Komisia za zashtita ot diskriminatsia (GC), C-83/14 (Court of Justice of the European Union July 16, 2015).
- Christopher, I. (2015, March 1). *This is the best explanation of gerrymandering you will ever see*. Retrieved from Washington Post: https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see/?noredirect=on&utm_term=.d2e9a0e16b34
- *Churchill and the Commons Chamber*. (2019, February 26). Retrieved from parliament.uk: https://www.parliament.uk/about/living-heritage/building/palace/architecture/palacestructure/churchill/
- *churchix.com*. (2019, May 7). Retrieved from churchix.com: https://churchix.com/

- Clifford, C. (2018, February 2). *You can pay for your burger with your face at this fast food restaurant, thanks to A.I.* Retrieved from CNBC: https://www.cnbc.com/2018/02/02/pay-with-facial-recognition-a-i-at-caliburger-in-pasadena-california.html
- Committee, U. N. (1989). *CCPR General Comment No. 18: Non-discrimination.* United Nations Human Rights Committee.
- Cooper v. Harris, 581 US _ (2017) (Supreme Court of the United States May 22, 2017).
- Copeland, B. (2019, January 23). *Artificial intelligence*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/technology/artificial-intelligence
- Coppin, B. (2004). *Artificial Intelligence Illuminated* (First ed.). Sudbury: Jones and Bartlett Publishers, Inc.
- Council of Europe Commissioner for Human Rights. (2019). *Unboxing artificial intelligence: 10 steps to protect human rights.* Strasbourg: Council of Europe .
- Darren, B. (2019, April 11). *China's hi-tech war on its Muslim minority*. Retrieved from Guardian: https://www.theguardian.com/news/2019/apr/11/china-hi-tech-war-on-muslim-minority-xinjiang-uighurs-surveillance-face-recognition
- Darren, S. (2019, May 22). *H.R.827 - AI JOBS Act of 2019*. Retrieved from United States Congress: https://www.congress.gov/bill/116th-congress/house-bill/827/text
- Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Retrieved from Reuters: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
- David, K. (2018). *Report of the Special Rapporteur to the General Assembly on Artificial Intelligence technologies and implications for the information environment (A/73/348).* Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.
- David, K. (2018, October 18). *Watch Out Workers, Algorithms Are Coming to Replace You — Maybe*. Retrieved from New York Times: https://www.nytimes.com/2018/10/18/business/q-and-a-yuval-harari.html
- deepmind.com. (2019, February 6). *AlphaGo Zero: Learning from scratch*. Retrieved from deepmind.com: https://deepmind.com/blog/alphago-zero-learning-scratch/
- Doffman, Z. (2019, January 28). *Amazon Refuses To Quit Selling 'Flawed' And 'Racially Biased' Facial Recognition*. Retrieved from Forbes: https://www.forbes.com/sites/zakdoffman/2019/01/28/amazon-hits-out-at-attackers-and-claims-were-not-racist/#f488f5446e78
- Doris, O. M. (2019, May 22). *H.R.1576 - Innovation Corps Act of 2017*. Retrieved from United States Congress: https://www.congress.gov/bill/115th-congress/house-bill/1576

- Drew, A. (2017, September 8). *GLAAD and HRC call on Stanford University & responsible media to debunk dangerous & flawed report claiming to identify LGBTQ people through facial recognition technology*. Retrieved from glaad.org: https://www.glaad.org/blog/glaad-and-hrc-call-stanford-university-responsible-media-debunk-dangerous-flawed-report
- Dunja, M. (2019, May 9). *Ethnic profiling: a persisting practice in Europe*. Retrieved from Council of Europe: 2019
- Eidelson, B. (2015). *Discrimination and Disrespect.* Oxford: Oxford University Press.
- Electronic Frontier Foundation. (2019, April 1). *Face Recognition*. Retrieved from Electronic Frontier Foundation: https://www.eff.org/tr/pages/face-recognition
- Eric, U. (2016, November 3). *Lawmakers to Facebook: Don't Let Advertisers Exclude by Race*. Retrieved from Propublica: https://www.propublica.org/article/lawmakers-to-facebook-dont-let-advertisers-exclude-by-race
- Europe, E. U. (2018). *Handbook on European non-discrimination law.* Luxembourg: Publications Office of the European Union.
- European Court of Human Rights. (2018). *Freedom of thought, conscience and religion.* European Court of Human Rights.
- European Union Agency for Fundamental Rights. (2015). *Protection against discrimination on grounds of sexual orientation, gender identity and sex characteristics in the EU.* Luxembourg: Publications Office of the European Union.
- Evelyn, E., & Watson, P. (2012). *EU Anti-Discrimination Law.* Oxford: Oxford University Press.
- Ewert v. Canada, 2018 SCC 30 (Supreme Court of Canada June 13, 2018).
- Falcon, W. (2018, October 25). *What Happens Now That An AI-Generated Painting Sold For $432,500?* Retrieved from Forbes: https://www.forbes.com/sites/williamfalcon/2018/10/25/what-happens-now-that-an-ai-generated-painting-sold-for-432500/#2f4e8d35a41c
- Fellesforbundet for Sjøfolk (FFFS) v. Norway, 74/2011 (European Committee of Social Rights July 2, 2013).
- Filippo, R., Hannah, H., Vivek, K., Christopher, B., & Levin, K. (2018). *Artificial Intelligence & Human Rights: Opportunities & Risks.* Massachusetts: Berkman Klein Center for Internet & Society at Harvard University.
- Fredman, S. (2001). *Discrimination and Human Rights.* Oxford: Oxford University Press.
- Freedom House. (2019). *Freedom in the World 2019.* Washington, D.C.: Freedom House.
- Gabbatt, A. (2011, February 17). *IBM computer Watson wins Jeopardy clash*. Retrieved from The Guardian: https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy

- Garvie, C., Frankle, J., & Bedoya, A. M. (2016). *Unregulated Police Face Recognition in America - Perpetual Line Up.* New Jersey: GeorgeTown Law Center on Privacy & Technology.
- Gayathri, M., K. R., & Midori, A. (2014). Investigating the Periocular-Based Face Recognition Across Gender Transformation. *IEEE Transactions on Information Forensics and Security 9(12)*, 2180-2192.
- Geere, D. (2013, August 17). *Google goes down for a few minutes, web traffic drops 40 percent*. Retrieved from Wired: https://www.wired.co.uk/article/googledip
- Gill v. Whitford, 585 U.S. ___ (Supreme Court of the United States June 18, 2018).
- Gillespie, E. (2019, February 24). *Are you being scanned? How facial recognition technology follows you, even as you shop*. Retrieved from Guardian: https://www.theguardian.com/technology/2019/feb/24/are-you-being-scanned-how-facial-recognition-technology-follows-you-even-as-you-shop
- Gina, N. P. (2016). Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication 10*, 4915–493.
- Google. (2018). *Search Quality Rater Guidelines.* Google.
- Google. (2019, May 3). *Get gender-specific translations*. Retrieved from Google Translate Help: https://support.google.com/translate/answer/9179237?p=gendered_translations&hl=en&visit_id=636924762002229493-3572379412&rd=1#
- Google. (2019, April 13). *Get Search results faster*. Retrieved from Google support: https://support.google.com/websearch/answer/106230?co=GENIE.Platform%3DAndroid&hl=en
- Google. (2019). *Google Diversity Annual Report 2019.* Google.
- Google. (2019, April 18). *How Google autocomplete works in Search*. Retrieved from Google Blog: https://www.blog.google/products/search/how-google-autocomplete-works-search/
- Google. (2019, April 12). *How Google Search Works*. Retrieved from support.google.com: https://support.google.com/webmasters/answer/70897
- Google. (2019, April 12). *How Search algorithms work*. Retrieved from Google.com: https://www.google.com/search/howsearchworks/algorithms/
- Google Spain SL, Google Inc. v Agencia Española de Protección de Datos (AEPD), Mario Costeja González, Case C-131/12 (European Court of Justice, Grand Chamber May 13, 2014).
- Gramlich , J. (2018, January 12). *The gap between the number of blacks and whites in prison is shrinking*. Retrieved from Pew Research Center: https://www.pewresearch.org/fact-tank/2018/01/12/shrinking-gap-between-number-of-blacks-and-whites-in-prison/

- Griggs et al. v. Duke Power Co., Co 401 US 424 (Supreme Court of the United States 1971).
- Hacker, P. (2018, May 5). *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law.* Retrieved from SSRN: https://ssrn.com/abstract=3164973
- Hagar, S., & Jonathan, J. (2018, October 20). *Revealed: Israel's Cyber-spy Industry Helps World Dictators Hunt Dissidents and Gays*. Retrieved from Haaretz: https://www.haaretz.com/israel-news/.premium.MAGAZINE-israel-s-cyber-spy-industry-aids-dictators-hunt-dissidents-and-gays-1.6573027
- Harcourt, B. E. (2010). Risk as a Proxy For Race. *University of Chicago Public Law Working Paper No. 323*, 1-14.
- Hardt, M. (2016, October 7). *Equality of Opportunity in Machine Learning*. Retrieved from Google AI Blog: https://ai.googleblog.com/2016/10/equality-of-opportunity-in-machine.html
- Harvard Law Review. (2019, May 14). *State v. Loomis*. Retrieved from Harvard Law Review: https://harvardlawreview.org/2017/03/state-v-loomis/
- Hawking, S., Tegmark, M., Russell, S., & Wilczek, F. (2014, June 19). *Transcending Complacency on Superintelligent Machines*. Retrieved from Huffingtonpost: https://www.huffingtonpost.com/stephen-hawking/artificial-intelligence_b_5174265.html?guccounter=1
- Hawkins, A. J. (2018, December 5). *RIDING IN WAYMO ONE, THE GOOGLE SPINOFF'S FIRST SELF-DRIVING TAXI SERVICE*. Retrieved from theverge.com: https://www.theverge.com/2018/12/5/18126103/waymo-one-self-driving-taxi-service-ride-safety-alphabet-cost-app
- Henry, K. (2018, June 1). *How the Enlightenment Ends*. Retrieved from The Atlantic: https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/
- Hern, A. (2015, May 20). *Flickr faces complaints over 'offensive' auto-tagging for photos*. Retrieved from Guardian: https://www.theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos
- Hernández, J. C. (2017, March 20). *China's High-Tech Tool to Fight Toilet Paper Bandits*. Retrieved from New York Times: https://www.nytimes.com/2017/03/20/world/asia/china-toilet-paper-theft.html?smid=tw-share&_r=0
- Hill, K. (2015, June 23). *You're being secretly tracked with facial recognition, even in church*. Retrieved from Splinternews: https://splinternews.com/youre-being-secretly-tracked-with-facial-recognition-e-1793848585

- Human Rights NGOs coalition. (2018, June 18). *Letter to Amazon*. Retrieved from American Civil Liberties Union: https://www.aclu.org/letter-nationwide-coalition-amazon-ceo-jeff-bezos-regarding-rekognition
- Human Rights Watch. (2018). *"Eradicating Ideological Viruses" China's Campaign of Repression Against Xinjiang's Muslims.* Human Rights Watch.
- Human Rights Watch. (2019). *China's Algorithms of Repression Reverse Engineering a Xinjiang Police Mass Surveillance App* . Human Rights Watch.
- Ian, S. (2017, January 27). *AI watchdog needed to regulate automated decision-making, say experts*. Retrieved from Guardian: https://www.theguardian.com/technology/2017/jan/27/ai-artificial-intelligence-watchdog-needed-to-prevent-discriminatory-automated-decisions
- IBM. (2019, March 21). *IBM Watson® Recruitment.* Retrieved from IBM: https://www.ibm.com/downloads/cas/1AKXPKVV
- Identoba and Others v. Georgia, 73235/12 (European Court of Human Rights May 12, 2015).
- Ingold, D., & Soper, S. (2016, April 21). *Amazon Doesn't Consider the Race of Its Customers. Should It?* Retrieved from Bloomberg: https://www.bloomberg.com/graphics/2016-amazon-same-day/
- James, B. (2018, August 1). *Outnumbered: From Facebook and Google to Fake News and Filter-bubbles by David Sumpter – review*. Retrieved from Guardian: https://www.theguardian.com/books/2018/aug/01/outnumbered-facebook-google-algorithms
- James, V. (2016, March 24). *Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day*. Retrieved from TheVerge: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist
- Jeff, L., Surya, M., Lauren, K., & Julia, A. (2016, May 23). *How We Analyzed the COMPAS Recidivism Algorithm*. Retrieved from Propublica: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- Jennifer, S., & Christopher, L. (2016, July 15). *Risk, Race, & Recidivism: Predictive Bias and Disparate Impact*. Retrieved from SSRN: https://ssrn.com/abstract=2687339
- John, K. (2017, June 26). *Smart Robots Put 10.5 Million US Jobs At High Risk, New Report Says*. Retrieved from Forbes: https://www.forbes.com/sites/johnkoetsier/2017/06/26/smart-robots-put-10-5m-us-jobs-at-high-risk-new-report-says/#2a683fd59277
- John, K. D. (2019, May 22). *H.R.4625 - FUTURE of Artificial Intelligence Act of 2017*. Retrieved from United States Congress: https://www.congress.gov/bill/115th-congress/house-bill/4625/text
- Jordan, E. (2017, October 6). *How Computers Turned Gerrymandering Into a Science*. Retrieved from New York Times:

https://www.nytimes.com/2017/10/06/opinion/sunday/computers-gerrymandering-wisconsin.html

- Julia, A. (2016, November 11). *Facebook Says it Will Stop Allowing Some Advertisers to Exclude Users by Race*. Retrieved from Propublica: https://www.propublica.org/article/facebook-to-stop-allowing-some-advertisers-to-exclude-users-by-race
- Julia, A., Jeff, L., Surya, M., Lauren, K., & ProPublica. (2016, May 23). *Machine Bias*. Retrieved from ProPublica: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Julia, A., Madeleine, V., & Ariana, T. (2017, September 14). *Facebook Enabled Advertisers to Reach 'Jew Haters'*. Retrieved from ProPublica: https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters
- Kay, M., Matuszek, C., & Munson , S. (2015). Unequal representation and gender stereotypes in image search results for occupations. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems ACM.* Seoul.
- Keane, T. (2018, January 24). *Federal agencies continue to advance capabilities with Azure Government*. Retrieved from Microsoft Azure Government Blog: https://devblogs.microsoft.com/azuregov/federal-agencies-continue-to-advance-capabilities-with-azure-government/
- Kehl, D. P. (2017). *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing.* Massachusetts: Berkman Klein Center for Internet & Society, Harvard Law School.
- Kemp, S. (2019). *Global Digital Report 2019.* We Are Social .
- Khaitan, T. (2015). *A Theory of Discrimination Law.* Oxford: Oxford Scholarship.
- Khullar, D. (2019, January 31). *A.I. Could Worsen Health Disparities*. Retrieved from New York Times: https://www.nytimes.com/2019/01/31/opinion/ai-bias-healthcare.html
- Kimberle, C. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum,* 139-167.
- Klare, B. F., Burge, M. J., Klontz, J. C., Bruegge, R. W., & Jain, A. K. (2012). Face Recognition Performance: Role of Demographic Information. *IEEE Transactions on Information Forensics and Security, Volume: 7, Issue: 6,* 1789 - 1801.
- Kofman, A. (2016, October 13). *LOSING FACE How a Facial Recognation Mismatch can Ruin Your Life*. Retrieved from The Intercept: https://theintercept.com/2016/10/13/how-a-facial-recognition-mismatch-can-ruin-your-life/

- Kosinski, M., Graepel, T., & Stillwell, D. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences 110(15)* , 5802-5805.
- Kumar, V., & Minz, S. (2014). Feature Selection: A literature Review. *Smart Computing Review, vol. 4, no. 3, June 2014,* 214-229.
- Latanya, S. (2013). Discrimination in Online Ad Delivery. *Communications of the Association of Computing Machinery (CACM), Vol. 56 No. 5,*, 44-54.
- Leviathan, Y. (2018, May 8). *Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*. Retrieved from Google AI Blog: https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html
- Liao, S. (2018, November 22). *Chinese facial recognition system mistakes a face on a bus for a jaywalker*. Retrieved from The Verge: https://www.theverge.com/2018/11/22/18107885/china-facial-recognition-mistaken-jaywalker
- Linn, A. (2018, March 14). *Microsoft reaches a historic milestone, using AI to match human performance in translating news from Chinese to English*. Retrieved from Microsoft The AI Blog: https://blogs.microsoft.com/ai/machine-translation-news-test-set-human-parity/
- Liptak, A. (2017, May 1). *Sent to Prison by a Software Program's Secret Algorithms*. Retrieved from New York Times: https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html
- Loomis v. Wisconsin, 16-6387 (Supreme Court of the United States June 26, 2017).
- Louise, L. (2018, January 15). *Alibaba and Microsoft AI beat humans in Stanford reading test*. Retrieved from Financial Times: https://www.ft.com/content/8763219a-f9bc-11e7-9b32-d7d59aace167
- Manning, A. (2017, May 1). *A False Facial Recognition Match Cost This Man Everything*. Retrieved from vocativ: https://www.vocativ.com/418052/false-facial-recognition-cost-denver-steve-talley-everything/index.html
- Mantelero, A. (2019). *Artificial Intelligence and Data Protection: Challenges and Possible Remedies.* Strasbourg: The Council of Europe Directorate General of Human Rights and Rule of Law.
- Maya, B. (2017). *AUTONOMOUS WEAPON SYSTEMS UNDER INTERNATIONAL HUMANITARIAN AND HUMAN RIGHTS LAW.* Geneva: Geneva Academy.
- Micah, A., & Michael, P. M. (2010). The Promise and Perils of Computers in Redistricting. *5 Duke Journal of Constitutional Law & Public Policy,* 69-111.

- Michal, K., & Yilun, W. (2018). Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images. *Journal of Personality and Social Psychology. Vol. 114, Issue 2*, 246-257.
- Microsoft Corporate Blogs. (2018, June 18). *Microsoft statement on separating families at the southern border*. Retrieved from Microsoft Blog: https://blogs.microsoft.com/on-the-issues/2018/06/18/microsoft-statement-on-separating-families-at-the-southern-border/
- Moeckli, D., Shah, S., & Sivakumaran, S. (2014). *International Human Rights Law.* Oxford: Oxford University Press.
- Monahan, J., & Skeem, J. L. (2015). Risk Assessment in Criminal Sentencing. *Virginia Public Law and Legal Theory Research Paper, No. 53*, 1-53.
- *mookkie.com*. (2019, March 27). Retrieved from https://www.mookkie.com/
- Negnevitsky, M. (2005). *Artificial Intelligence: A Guide to Intelligent Systems* (2nd ed.). Essex: Pearson Education.
- Nemitz, P. (2018). Constitutional Democracy and Technology in the age of Artificial Intelligence. *Royal Society Philosophical Transactions A*, 1-14.
- Noble, S. U. (2018). *Algorithms of Oppression How Search Engines Reinforce Racism.* New York: New York University Press.
- Norvig, P., & Russell, S. (2010). *Artificial Intelligence: A Modern Approach* (Third ed.). New Jersey: Prentice Hall.
- Nottebohm (Liechtenstein v. Guatemala) (International Court of Justice April 6, 1955).
- Novruk and Others v. Russia, 31039/11, 48511/11, 76810/12, 14618/13, 13817/14 (European Court of Human Rights March 15, 2016).
- Olivia, S. (2016, June 17). *Self-driving trucks: what's the future for America's 3.5 million truckers?* Retrieved from Guardian: https://www.theguardian.com/technology/2016/jun/17/self-driving-trucks-impact-on-drivers-jobs-us
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: 2016.
- Patrick, W. H. (1992). *Artificial Intelligence* (Third ed.). Boston: Addison-Wesley Publishing Company.
- Paul, M. (2019, April 14). *One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority*. Retrieved from New York Times: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html
- Pennsylvania Commission on Sentencing. (2015). *Impact of Removing Demographic Factors.* Pennsylvania: Pennsylvania Commission on Sentencing.

- Pennsylvania Commission on Sentencing. (2018). *Validation of a Risk Assessment Instrument by Offense Gravity Score for All Offenders.* Pennsylvania: Pennsylvania Commission on Sentencing.
- Pennsylvania Commission on Sentencing. (2018). *What are the implications of removing gender as a risk factor?* Pennsylvania: Pennsylvania Commission on Sentencing.
- Purcell, B. (2018, February 27). *The Ethics Of AI: How To Avoid Harmful Bias And Discrimination Build Machine Learning Models That Are Fundamentally Sound, Assessable, Inclusive, And Reversible.* Retrieved from ibm.com: https://www.ibm.com/downloads/cas/6ZYRPXRJ
- Purcell, K., Brenner, J., & Rainie, L. (2012). *Search Engine Use 2012.* Pew Research Center.
- Rachel, L. S. (2015, April 16). *Event Attendance with Facial Recognition Software*. Retrieved from CHURCHMAG : https://churchm.ag/facial-recognition-software/
- Rachel, M. (2019, May 14). *San Francisco just banned facial-recognition technology*. Retrieved from CNN: https://edition.cnn.com/2019/05/14/tech/san-francisco-facial-recognition-ban/index.html
- Ratcliffe, S. (2018). *Oxford Essential Quotations* (6th ed.). Oxford: Oxford University Press.
- Rebecca, W. (2018). Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System. *70 Stanford Law Review 1343 (2018)*, 1-87.
- Regan, J. (2016, December 7). *New Zealand passport robot tells applicant of Asian descent to open eyes*. Retrieved from Reuters: https://www.reuters.com/article/us-newzealand-passport-error/new-zealand-passport-robot-tells-applicant-of-asian-descent-to-open-eyes-idUSKBN13W0RL
- researchers, A. g. (2019, May 16). *AUTONOMOUS WEAPONS: AN OPEN LETTER FROM AI & ROBOTICS RESEARCHERS*. Retrieved from futureoflife.org: https://futureoflife.org/open-letter-autonomous-weapons/
- Reynolds, E. (2018, June 1). *The agony of Sophia, the world's first robot citizen condemned to a lifeless career in marketing*. Retrieved from Wired: https://www.wired.co.uk/article/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics
- Roach, J. (2019, June 26). *Microsoft improves facial recognition technology to perform well across all skin tones, genders*. Retrieved from Microsoft AI Blog: https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/
- Robert, E. L. (2019, May 22). *H.R.3388 - SELF DRIVE Act*. Retrieved from United States Congress: https://www.congress.gov/bill/115th-congress/house-bill/3388
- Romm, T., Timberg, C., & Harwell, D. (2018, December 12). *Google CEO Sundar Pichai: Fears about artificial intelligence are 'very legitimate,' he says in Post interview*. Retrieved

from Washington Post: https://www.washingtonpost.com/technology/2018/12/12/google-ceo-sundar-pichai-fears-about-artificial-intelligence-are-very-legitimate-he-says-post-interview/?utm_term=.6654f276b469

- Romm, T., Timberg, C., & Romm, T. (2018, December 12). *Google CEO Sundar Pichai: Fears about artificial intelligence are 'very legitimate,' he says in Post interview*. Retrieved from Washington Post: https://www.washingtonpost.com/technology/2018/12/12/google-ceo-sundar-pichai-fears-about-artificial-intelligence-are-very-legitimate-he-says-post-interview/?noredirect=on&utm_term=.205f9e250264
- Sam, L. (2016, September 8). *A beauty contest was judged by AI and the robots didn't like dark skin*. Retrieved from Guardian: https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people
- Sandholm, T., & Noam, B. (2018). Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *American Association for the Advancement of Science*, 418-424.
- Sasha Costanza, C. (2018, July 27). *Design Justice, A.I., and Escape from the Matrix of Domination*. Retrieved from MIT: https://jods.mitpress.mit.edu/pub/costanza-chock?version=fced98f6-af9b-496c-9518-7053122037bd
- Schiek, D., Waddington, L., & Bell, M. (2007). *Cases, Materials and Text on National, Supranational and International Non-Discrimination Law.* Oxford: Hart Publishing.
- *Search Engine Market Share Worldwide*. (2019, April 17). Retrieved from gs.statcounter.com: http://gs.statcounter.com/search-engine-market-share
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 417-457.
- Simonite, T. (2018, November 1). *When It Comes to Gorillas, Google Photos Remains Blind*. Retrieved from Wired: https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/
- Singh, R. (2016). *World Facial Recognition Market - Opportunities and Forecasts, 2015 - 2022.* Allied Market Research.
- Sini, R. (2016, June 9). *'Three black teenagers' Google search sparks Twitter row*. Retrieved from BBC: https://www.bbc.com/news/world-us-canada-36487495
- Smith, B. (2018, July 13). *Facial recognition technology: The need for public regulation and corporate responsibility*. Retrieved from Microsoft Blog: https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/
- Smith, B. (2018, December 6). *Facial recognition: It's time for action*. Retrieved from The Official Microsoft Blog: https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/

- Snow , J. (2018, July 26). *Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots*. Retrieved from American Civil Liberties Union Foundation of Northern California: https://www.aclunc.org/blog/amazon-s-face-recognition-falsely-matched-28-members-congress-mugshots
- Society of Automotive Engineers International. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems.* Society of Automotive Engineers International.
- Springer , A., & Döpfner, M. (2018, October 21). *Author and historian Yuval Noah Harari discusses the battle against fake news, the challenges facing democracy worldwide, and the biggest threat facing humanity in the next 100 years*. Retrieved from Business Insider: https://www.businessinsider.com/yuval-noah-harari-interview-21-lessons-forthe-21stcentury-author-2018-10?r=US&IR=T&IR=T
- State v. Wisconsin, 881 N.W.2d 749 (Wisconsin Supreme Court July 13, 2016).
- Stevenson, M. T. (2017). Assessing Risk Assessment in Action. *103 Minnesota Law Review 303*, 303-384.
- stopkillerrobots.org. (2019, May 16). *UN head calls for a ban*. Retrieved from stopkillerrobots.org: https://www.stopkillerrobots.org/2018/11/unban/
- Stowe, B. (2018, December 10). *The Integral Accident*. Retrieved from medium.com: https://medium.com/work-futures/the-integral-accident-d22636632cf9
- Stumpe, M., & Mermel, C. (2018, November 16). *Improved Grading of Prostate Cancer Using Deep Learning*. Retrieved from Google AI Blog: https://ai.googleblog.com/2018/11/improved-grading-of-prostate-cancer.html
- Technology, E. O. (2016). *Preparing for the Future of Artificial Intelligence.* Washington, DC: National Science and Technology Council.
- The Council of Europe. (2019, January 28). *Glossary*. Retrieved from The Council of Europe: https://www.coe.int/en/web/artificial-intelligence/glossary
- The Council of Europe. (2019, February 5). *History of Artificial Intelligence*. Retrieved from The Council of Europe: https://www.coe.int/en/web/artificial-intelligence/history-of-ai
- The European Commission for the Efficiency of Justice of the Council of Europe. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment.* Strasbourg: Council of Europe.
- The Guardian. (2017, May 23). *World's best Go player flummoxed by Google's 'godlike' AlphaGo AI*. Retrieved from The Guardian: https://www.theguardian.com/technology/2017/may/23/alphago-google-ai-beats-ke-jie-china-go
- The Local. (2013, May 6). *No probe into Stockholm police's 'racial profiling'*. Retrieved from The Local: https://www.thelocal.se/20130506/47748

- The Local. (2013, February 21). *Outrage over Stockholm cops' 'racial profiling'*. Retrieved from The Local: https://www.thelocal.se/20130221/46330
- The New York City Council. (2019, May 22). *Local Law 49 of 2018*. Retrieved from The New York City Council: https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0
- Timishev v. Russia, 55762/00, 55974/00 (European Court of Human Rights December 13, 2005).
- Toor, A. (2017, April 25). *Google takes steps to limit offensive and inaccurate search results*. Retrieved from The Verge: https://www.theverge.com/2017/4/25/15418490/google-search-snippets-changes-fake-news-offensive-results
- UN Women. (2013, October 21). *UN Women ad series reveals widespread sexism*. Retrieved from UN Women: http://www.unwomen.org/en/news/stories/2013/10/women-should-ads
- United Nations. (2019, February 22). *15. Convention on the Rights of Persons with Disabilities*. Retrieved from United Nations Treaty Collection: https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=IV-15&chapter=4
- United Nations Committee on Economic, S. a. (1991). *General Comment No. 4: The Right to Adequate Housing (Art. 11 (1) of the Covenant).* United Nations Committee on Economic, Social and Cultural Rights.
- United Nations Committee on Economic, S. a. (2003). *General Comment No. 15: The Right to Water (Arts. 11 and 12 of the Covenant).* United Nations Committee on Economic, Social and Cultural Rights.
- United Nations Committee on Economic, S. a. (2009). *General comment No. 20: Non-discrimination in economic, social and cultural rights (art. 2, para. 2, of the International Covenant on Economic, Social and Cultural Rights).* Geneva: United Nations Committee on Economic, Social and Cultural Rights.
- Wagner, B. (2018). Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping? *BEING PROFILED:COGITAS ERGO SUM*. Retrieved from https://www.privacylab.at/wp-content/uploads/2018/07/Ben_Wagner_Ethics-as-an-Escape-from-Regulation_2018_BW9.pdf
- Walker, K. (2018, December 13). *AI for Social Good in Asia Pacific*. Retrieved from Google Blog: https://www.blog.google/around-the-globe/google-asia/ai-social-good-asia-pacific/amp/
- Waymo Team. (2018, December 5). *Riding with Waymo One today*. Retrieved from medium.com: https://medium.com/waymo/riding-with-waymo-one-today-9ac8164c5c0e
- Wichert, A. (2014). *Principles of Quantum Artificial Intelligence* (First ed.). Singapore: World Scientific Publishing Co. Pte. Ltd.

- Wiener, N., & Rosenblueth, A. (1945). The Role of Models in Science. *Philosophy of Science, Vol. 12, No. 4,* 316-321.
- Williams v. Saxbe, 413 F Supp 654 (U.S. District Court for the District of Columbia, 1976).
- Wingfield, N. (2018, May 22). *Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.* Retrieved from New York Times: https://www.nytimes.com/2018/05/22/technology/amazon-facial-recognition.html?module=inline
- Xiang, N. (2018, October 5). *China's AI Industry Has Given Birth To 14 Unicorns: Is It A Bubble Waiting To Burst?* Retrieved from Forbes: https://www.forbes.com/sites/ninaxiang/2018/10/05/chinas-ai-industry-has-given-birth-to-14-unicorns-is-it-a-bubble-waiting-to-pop/#5019b7f146c3
- Yadron, D., & Tynan, D. (2016, June 1). *Tesla driver dies in first fatal crash while using autopilot mode*. Retrieved from Guardian: https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk
- (2007). *Yogyakarta Principles.* Yogyakarta: March.
- York, C. (2016, June 8). *Three Black Teenagers: Is Google Racist?* Retrieved from HuffPost UK: https://www.huffingtonpost.co.uk/entry/three-black-teenagers-google-racism_uk_575811f5e4b014b4f2530bb5?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAIBFhBhv4aTrweThh71IZo_GW2pxwo7IYuttdTTkl6P7fMw6cOHiligTQq85dbqX-iKx5zC
- Yuval, N. H. (2017, May 24). *Are we about to witness the most unequal societies in history?* Retrieved from Guardian: https://www.theguardian.com/inequality/2017/may/24/are-we-about-to-witness-the-most-unequal-societies-in-history-yuval-noah-harari?CMP=share_btn_tw
- Yuval, N. H. (2018, October). *Why Technology Favors Tyranny*. Retrieved from The Atlantic: https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/
- Zhang, M. (2015, July 1). *Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software*. Retrieved from Forbes: https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software/#18058db713d8

Index