UMEÅ UNIVERSITY

Automated decision-making vs indirect discrimination

Solution or aggravation?

Emma Lundberg

Spring Semester 2019
Independent Written Essay within the Field of Constitutional Law and Human Rights, 15 Credits
[Master's Program in Constitutional Law and Human Rights, 60 credits]
Supervisor: Markus Naarttijärvi

Abstract

The usage of automated decision making-systems by public institutions letting the system decide on the approval, determination or denial of individuals benefits as an example, is an effective measure in making more amount of work done in a shorter time period and to a lower cost than if it would have been done by humans. But still, although the technology has developed into being able to help us in this way, so has also the potential problems that these systems can cause while they are operating. The ones primarily affected here will be the individuals that are denied their benefits, health care, or pensions.

The systems can maintain hidden, historical stigmatizations and prejudices, disproportionally affecting members of a certain historically marginalized group in a negative way through its decisions, simply because the systems have learned to do so. There is also a risk that the actual programmer includes her or his own bias, as well as incorrect translation of applicable legislations or policies causing the finalized system to make decisions on unknown bases, demanding more, less or completely other things than those requirements that are set up by the public and written laws. The language in which these systems works are in mathematical algorithms, which most ordinary individuals, public employees or courts will not understand. If suspecting that you could have been discriminated against by an automated decision, the requirements for successfully claim a violation of discrimination in US-, Canadian- and Swedish courts, ECtHR and ECJ demands you to show on which of your characteristics you were discriminated, and in comparison to which other group, a group that instead has been advantaged. Still, without any reasons or explanations to why the decision has been taken available for you as an applicant or for the court responsible, the inability to identify such comparator can lead to several cases of actual indirect discriminations being denied.

A solution to this could be to follow the advice of Sophia Moreau's theory, focusing on the actual harm that the individual claim to have suffered instead of on categorizing her or him due to certain traits, or on finding a suitable comparator. This is similar to a ruling of the Swedish Court of Appeal, where a comparator was not necessary in order to establish that the applicant had been indirectly discriminated by a public institution. Instead, the biggest focus in this case was on the harm that the applicant claimed to have suffered, and then on investigating whether this difference in treatment could be objectively justified.

In order for Swedish and European legislation to be able to meet the challenges that can arise through the usage of automated decision making-systems, this model of the Swedish Court of Appeal could be a better suited model to help individuals being affected by an automated decision of a public institution, being potentially indirectly discriminative.

Table of Contents

Abbreviations **4**
1. Introduction **5**
1.1 Aim 6
1.2 Delimitations 6
1.3 Method and materials 7
2. Technologies **8**
2.1 Big Data 8
2.2 Automated decision-making 9
2.3 Mixed automated system + automation bias = fully automated system 10
2.4 Automated decision-making in practice 10
2.5 Identified problems of automated decision making-systems 12
2.5.1 Lost in translation 12
2.5.2 Opacity and ability to understand 13
2.5.3 Concluding remarks 13
3. Legislation **14**
3.1 European Convention on Human Rights 15
3.1.1 ECHR, ECtHR and indirect discrimination 15
3.2 European Union 16
3.2.1 Primary law protecting non-discrimination 17
3.2.2 Secondary law and case-law on indirect discrimination 17
3.2.3 Concluding remarks 19
3.3 Sweden 19
3.3.1 Legislation 19
3.3.2 Case-law 20
3.4 Current legislation and technological challenges 21
3.4.1 Legal requirements of indirect discrimination 22
3.4.2 Bias 22
3.4.3 Opacity 23
3.5 No comparator, no discrimination? 23
Conclusion **26**
Bibliography **27**

Abbreviations

ECHR: European Convention on Human Rights

ECtHR: European Court on Human Rights

EU-Charter: Charter of Fundamental Rights of the European Union

TEU: Treaty on European Union

TFEU: Treaty on the Functioning of the European Union

ECJ: Court of Justice of the European Union

DO: the Swedish Ombudsman on Discrimination

AI: Artificial Intelligence

1. Introduction

The technological developments during the last decades have affected most areas of human lives. Both regarding how individuals e.g. find information but also how public institutions communicates, works and decides.[1] The developments have e.g. enabled using algorithms that rapidly find hidden patterns and connections in databases full of information to, based on previous decisions, be able to predict future events that are likely to happen.[2] The attraction of introducing new technologies to ease many of our tasks does of course seem very appealing, like a toolbox containing all tools possible, even those that no one can imagine yet.

Some years ago, in the US, loud protests and demonstrations against police forces use of violence and racial bias took place. This initiated the idea of introducing big data- and automated decision making-systems with the hopes of them leading to police officers doing their work more accurately as well as to calm down the protesters and their dissatisfactions with the police. Said and done, these types of systems were introduced to the police forces with the hopes of them helping to end the accusations of police officers' racial bias, and to enable a more efficient and functional police force.[3]

Today, public institutions in Sweden have also introduced automated decision making-systems to some extent.[4] The benefits of introducing these systems have been argued to be their ability to make decisions at a lower cost and much faster than human employees, while also increasing the quality of the decisions.[5] These systems are said to have the potential to make decisions that are more coherent and equal, where individuals in similar situations will be correctly treated alike.[6] Allowing computers to do more of the basic work, including making decisions independently, has been increasingly emphasized as something that will likely be common in every public institution soon enough.[7]

However, several researchers have accused the usage of machine learning- and data mining-algorithms and AI-technologies of automated decision making-systems in rather harsh manners. Namely, that these components of the systems are likely to be discriminative[8] and, that there is a risk that they will produce incorrect decisions based on incorrect data.[9]

Are there problems of automated decision making-systems making discriminative decisions?

---

[1] Johan Ledendal, Stefan Larsson, Joakim Wernberg, *Offentlighet i det Digitala Samhället: Vidareutnyttjande, Sekretess och Dataskydd* (Norstedts Juridik 2018) 40-41.

[2] Examples on usage of predictive analytics: insurance companies basing insurance conditions on automated profiling, or; police departments predicting where crimes are likely to occur (see more in chapter 2).

[3] Andrew Guthrie Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement* (New York University Press 2017) 19.

[4] Markus Naarttijärvi, 'Rättsstatlighet och Algoritmiska Svarta Lådor' in Örjan Edström, Johan Lindholm, Ruth Mannelqvist (eds) *Jubileumsskrift till Juridiska institutionen 40 år* (Juridiska Institutionen Umeå Universitet 2017) 246-247.

[5] Sveriges Kommuner och Landsting, 'Automatisering i Välfärden: Möjligheter och Utmaningar för Kommuner och Regioner' <https://rapporter.skl.se/automatisering-i-valfarden.html> accessed 23 May 2019.

[6] Danielle Keats Citron, 'Technological Due Process' (2008) 85 Washington University Law Review 1249, 1252-1253.

[7] Sveriges Kommuner och Landsting (n 5) accessed 19 May 2019.

[8] Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671, 677.

[9] Keats Citron (n 6) 1256.

1.1 Aim

The aim of this essay is to investigate the existing research on the functions, usage and results of automated decision making-systems and their possible discriminative effects, and to investigate how the legal rules and principles on discrimination in our Swedish context can meet such new technologies. This will be done by first investigating the US-context due to its wide usage of automated decision making-systems in the public sphere, and the extensive research on this topic in that context. The focus here will be to investigate what issues or challenges that these automated systems pose when being used by public institutions to make decisions affecting individuals.

After identifying possible challenges in the US-context, these will be translated into a European and Swedish context. The aim here is to see to what extent the ECHR, EU- and Swedish legislations can meet such new technological challenges in the Swedish context, and if the current legislations are even capable of doing so.

In order to reach this aim, two research questions will lead the discussion and investigation throughout this essay. These questions are:

1. What are the challenges of public institutions usage of automated decision making-systems in relation to discriminations against individuals in a US-context?
2. Can current legislation on discrimination of ECHR, EU and Sweden in force today handle the challenges of automated decision making-systems?

1.2 Delimitations

In order to make this discussion as efficient as possible, being able to answer my research questions while still offering a good overview of the situation that I aim describing, this requires quite extensive limitations. First, regarding the technologies, the focus will be on presenting and describing the usage of automated decision making-systems that are used by public institutions in decisions affecting individuals. Further, these will be two different types of systems, namely mixed and fully automated systems. The former is dependent on a human operator realizing the, by the systems', suggested decisions, while the latter is completely independent. However, due to so called 'automation bias' by these human operators, a mixed system will work as a fully automated one.[10] Therefore, all automated decision making-systems referred to will be considered as fully automated ones unless otherwise presented.

When translating these challenges into ECHR, EU and Swedish legal contexts, the focus here will be whether the presented issues of the systems can be considered as causing indirect discrimination according to these legal frameworks or not. This is due to, as will be presented in the following chapter, the hidden and neutral adjudications of the technologies. Here, non-discrimination will be the focus while this is a part of the bigger right to equality.[11] However, the right to equality will not be investigated or further presented. Instead, this is viewed as the aim and goal to be achieved through the realization of non-discrimination. Equality and non-discrimination will therefore be considered as parts of the same right, while the 'operative part' of it, namely discrimination, is what will be investigated further.

Lastly, to enable better comparisons between the different contexts of presented examples from case-law and applications of technologies, there will be a limitation of focusing mostly

[10] ibid 1271. This is further developed in chapter 2.
[11] Elise Muir, *EU Equality Law: The First Fundamental Rights Policy of the EU* (OUP 2018) 3.

on public institutions usage of automated decision making-systems in cases concerning social welfare benefits and health care. The same focus applies to the selected case-law from the ECHR-, EU- and Swedish contexts of examples of indirect discrimination primarily in relation to such benefits or medical care. However, additional case-law regarding other situations will also be presented shortly, in order to illustrate the basic principles of the different courts' adjudications in different cases concerning discrimination.

1.3 Method and materials

To be able to answer my first research question and to explain the functions of the automated systems, mainly technological or semi-technological materials will be used. This will entail using books, articles and newspapers to be able to, in a general way, explain the basic functions of the systems while giving real-life examples of when they have been used and what issues they have caused. This will be sources from mainly a US-context since the most research and analyzes are available here.

For my second question, I will be focusing on the indirect discrimination-legislation of ECHR, EU and Sweden, once again regarding the national public institutions applications, systems and requirements for social welfare benefits and health care. Here, the legal sources will be considered due to the legal hierarchy of norms. This means that, since the aim is to translate the identified challenges of the US-context into a Swedish context, both ECHR and EU-law on indirect discrimination will be considered. As regional sources, their most fundamental norms will first be examined, being the ECHR and the primary EU-law such as the treaties and EU-charter[12] followed by secondary laws such as Directives.[13] Case-law of both the ECtHR and ECJ will be presented to show the adjudications and deliberations in cases of discriminations. The same will be done in the Swedish context, where first presenting the hierarchically highest source of law, namely the Constitution.[14] After this, ordinary law, followed by case law and the work of the DO will be presented in order to lay out the possibilities and challenges of current national and regional legislations to combat possibly new issues caused by the usage of automated decision making-systems.

After presenting this, the essay will end with a suggestion on how courts could better encounter the challenges of automated decision making-systems in cases of alleged indirect discriminations. Here, Swedish case-law will be tied together with a legal theory on equality rights from a US- and Canadian context. This theory will, together with the Swedish case-law, be translated into the Swedish context in order to present a model that could potentially be better suited for the cases of indirect discriminations caused by automated decision making-systems before Swedish courts.

[12] EU Commission, 'Types of EU Law' < https://ec.europa.eu/info/law/law-making-process/types-eu-law_en> accessed 20 May 2019.
[13] EUR-Lex, 'Sources of European Union Law' < https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Al14534> accessed 20 May 2019.
[14] András Sajó and Renata Uitz, *The Constitution of Freedom: An Introduction to Legal Constitutionalism* (OUP 2017) 7.

2. Technologies

Big data or AI is today often used to describe many different kinds of modern technologies, but the concepts do actually encompass several different components, making the sloppy use of 'big data' or 'AI' in all situations, kind of misleading. As an example, big data consists of more technologies than to simply collect large amounts of information. These different technologies do therefore require more detailed definitions, in order to sort out what they do and how they function. A short description of the technologies that will be the foundations of the discussion in this essay will be presented here.

2.1 Big Data

The actual usage of *big data* has primarily been described, analyzed and researched by scholars in the US-context, where it is often used by public institutions like police forces.[15] Big data collects tremendous amounts of information, in a speed that no human would ever be able to compete with. This collection is then analyzed by *data mining* and *machine learning algorithms*, that are created by human programmers and which are given the task to solve a particular issue. Data mining-algorithms analyses the huge amounts of data collected to identify connections and patterns hidden therein. These identified patterns are made into categories or classifications by the algorithms, to make predictive analytics of potential events possible. With the help of machine-learning algorithms, the systems will constantly learn what categories or classifications that risks resulting in events that the programmer has decided to be unwanted, based on the information that is given to the algorithms and based on its previous decisions.[16]

An example on the usage of big data might help to illustrate this even further. In the US, the retail company Walmart uses big data-systems in collecting massive amounts of data each day about their customers shopping habits. Through this collection, their systems identified an unexpected correlation. It was found that the sale of Strawberry Pop-Tarts, a sweet ready-to-eat snack, increased before hurricanes. However, even though big data together with its algorithms and machine-learning could find and show such correlation, it could not explain it. The potential reasons or explanations for this surprising discovery was therefore left unanswered.[17]

So, in the definition of 'big data-systems', several technologies are included. To sum up, these do together collect large amounts of information, being analyzed by data mining and machine-learning algorithms to find and reveal concealed correlations or patterns therein, as well as analytics that predicts future events or developments.[18]

The original idea of big-data systems was primarily to be a tool for private companies that would enable them to better understand their customers' needs and preferences in order to create better and more individualized advertisements to increase their sales and incomes. However, in the US, many police departments are today using these big data-systems to a wide extent after realizing the benefits of these systems. The polices' wide usage of the systems today shows their beliefs in that these can be equally helpful for the public sphere.

[15] Naarttijärvi (n 4) 247-248.
[16] Barocas and Selbst (n 8) 677-678.
[17] Guthrie Ferguson (n 3) 8-9.
[18] ibid (n 3) 8.

The usage of big data-systems by the police departments has been motivated by claims that it will be cost-effective and will improve the quality of policing.[19]

2.2 Automated decision-making

The different technologies presented above, that together form a big data-system, can also serve as a basis for another type of system. A system which is capable of independently making its own decisions based on various facts and circumstances. Such system is called an automated decision making-system which is, similar to big data, also a type of an algorithmic system[20], based on data mining- and machine-learning algorithms.[21] The algorithms collects data, analyzes it and predicts possible needs or actions in the future. These findings are then used as a basis for the automated decisions that are made by the system.[22]

An 'automated decision' is a technological process where human-created algorithms are being faced with a question or an issue, on which the system possesses the power to independently make a decision.[23] To better explain what an automated decision making-system is, we can explain this in practical steps. The first step is that a programmer translates e.g. a policy to the language of the computer, in other words into computer code. The second step is that a collection of algorithms, selected by the human programmer, compares data from several databases to find as much information as possible regarding the task that it has been given. The third and last step is then that data-mining-algorithms searches through the massive amounts of information identified in order to find correlations and patterns therein that will be bases for its decision.[24] The machine-learning algorithms incorporated into the systems are what enables the systems to develop their accuracy and quality of its decisions, in accordance with what the programmer has chosen to be accurate and qualitative decisions.[25]

The learning process of machine-learning algorithms are steered by so called 'training-data'. This is data that the system is given to practice on, in order to improve its quality of analyzation and decision-making. Through this practice, the automated decision making-system, which the machine-learning algorithms are a component of, is able to make more correct decisions. It is commonly known in technology science that the precision of the algorithms in an automated decision making-system improves with a bigger quantity of training-data. Due to this, it is crucial that the algorithms are given comprehensive, qualitative, but also quantitative, training-data, in order for the automated decision making-systems to be able to make qualitative decisions.[26]

---

[19] ibid (n 3) 16-17.

[20] Sandra Wachter, Brent Mittelstadt, Chris Russell 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31 Harvard Journal of Law & Technology 841, 844.

[21] James Larus and Chris Hankin, 'Regulating Automated Decision Making' (2018) 61(8) Communications of the ACM 5, 5.

[22] Scott Monteith and Tasha Glenn, 'Automated Decision-Making and Big Data: Concerns for People With Mental Illness' (2016) 18(12) Current Psychiatry Reports 1, 1.

[23] Matthias Gallé, 'What Can I Do Now? Guiding Users in a World of Automated Decisions' (2018) Xerox Research Centre Europe <https://arxiv.org/pdf/1701.03755.pdf> accessed 1 May 2019, 1.

[24] Keats Citron (n 6) 1260.

[25] Mireille Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics' (2018) 68 University of Toronto Law Journal 12, 25.

[26] Jenna Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) Big Data & Society <https://journals.sagepub.com/doi/pdf/10.1177/2053951715622512> accessed 13 May 2019 1, 5.

2.3 Mixed automated system + automation bias = fully automated system

A fully automated decision making-system has no human intervention and produces decisions which are being directly enforced, producing effects for the individual or group in question. A mixed system means that it makes its own decisions as well, but with the difference that such will not be automatically enforced. The realization of the suggested decisions by a mixed system is dependent on specifically authorized personnel to review and enforce them. This enables that personnel to realize the systems decision, but also to adjust it.[27] However, even though this possibility of changes exists, there have been situations where such personnel have failed at large to identify incorrect decisions by the system as well as incorrect requirements or data that such decisions have been based on.[28] This phenomenon has been the topic of researches investigating the reasons for certain automated decision making-systems many incorrect decisions. The findings are that the persons assigned to operate these systems tends to overestimate them, even in situations where these persons actually suspect that the system might not work correctly. In other words, they had what can be called an over-belief in the system's ability to make correct decisions since considering them to be 'error-resistant'. These findings have led to a new term for such personnel ignoring their own suspicions regarding the system's correctness. The term for this is 'automation bias'.[29]

In such situations, when automation bias leads to that the reviewing-personnel approves and realizes close to every suggested decision by the system without second thoughts, the line between a mixed system and a fully automated decision making-system has completely disappeared. This over-belief in automated decision making-systems can be due to a general belief that computer systems are more competent than humans to take correct and objective decisions. With factors like increasing needs of public services and expectations by the public to get help faster and faster, the existing difference between mixed and fully automated decision making-systems, might be completely erased during the 21th century. This, due to the human over-belief in the systems, and the notion that they make more correct and competent decision than us.[30]

2.4 Automated decision-making in practice

Different types of computer systems have been used by state and federal institutions in the US as assistance to the employees in their decision-makings since the 1970s. As with all other types of technology, these systems have developed into becoming the automated decision making-systems that we know today with their underlying machine-learning and data-mining algorithms and predictive analytics. However, several incidents in the US-context have occurred where these automated decision making-systems have failed by making decisions based on incorrect data and affecting those concerned in a negative and sometimes unjust way, often being the result of automation bias.[31] Another issue has been that the systems have identified individuals incorrectly and mistaken them for being someone else due to having similar names.[32]

[27] Keats Citron (n 6) 1263-1267.
[28] Examples are the Colorado Benefit Management System and the No Fly-system, both which will be presented in part 2.4.
[29] Juergen Sauer, Alain Chavaillaz, David Wastell, 'Experience of Automation Failures in Training: Effects on Trust, Automation Bias, Complacency and Performance' (2016) 59 Ergonomics 767, 767-768.
[30] Keats Citron (n 6) 1271-1272.
[31] ibid 1267.
[32] ibid (n 6) 1273.

This was the case with a system analyzing data from both state and federal levels in the search for parents who owed child support.[33] The issue with such system was that it was eventually found to incorrectly identify individuals as owing child support, while they in fact did not since the system had mixed them up with the correct parents due to them having similar names. In one such case, a man who had been incorrectly identified in such way needed to spend two months in proceedings with his attorney before they had convinced the court that this was simply a mistake and that he in fact did not owe such child support. These incorrect decisions, affecting individuals in such a way, are of course very serious. The fact that it in this case demanded two months of proceedings to correct the situation, makes it hard for everyone to achieve such redress if being affected in the same way. For many, such amount of time and money for similar proceedings might not be available. At the same time, the parent that is to be paid child support is affected negatively as well, while the responsible individual being in debts is not being held accountable.[34]

A similar example is the "No Fly"-system in the US where, due to the usage of an automated decision making-system by airports, around 1,500 travelers are incorrectly being labeled as potential terrorists every week. Behind the "No Fly"-system are data-mining algorithms that goes through databases containing massive amounts of information of millions of American citizens like e.g. e-mails and telephone records. The information about individuals is then compared to pre-programmed profiles which according to the system represents a terrorist. The individuals with information matching these profiles are therefore the ones that are labeled as potential terrorists by the automated system. However, an explanation of the reasons for why these individuals have been labeled as such is not available, even though they because of this automated decision risks being detained and exposed to long interrogations. Due to this, it is hard to effectively defend yourself against such decision. As a result, the system has gotten massive criticism mainly due to the algorithms errors and the often inaccurate data that its decisions and identifications are based upon. It has also been emphasized that this can lead to stigmatizing categorizations of individuals.[35]

Another example is the Colorado Benefits Management System which has produced automated decisions affecting individuals. This benefit-system did due to wrongfully translated rules and laws into computer codes, produce around 100,000 incorrect automated decisions. These automated decisions were considered as incorrect since they, due to erroneous translations of the rules and laws by the programmers of the system, denied individuals access to social welfare benefits based on requirements that were not in accordance with the law.[36] As many as 900 rules where incorrectly translated into the codes that the automated decision-system was built upon and one of these incorrectly translated rules affected several patients suffering cancer by denying them the medical aid that they in fact were entitled to. Others incorrectly discontinued food stamps based solely on that these persons had *previously* had problems with drugs, decisions that were in violation of state law.[37] These are only two examples of the many mistakes that this automated benefit-system made, and since they were so very many and affected so many individuals, these incorrect decisions was eventually observed by society which lead to actions to correct this incorrect

[33] Keats Citron (n 6) 1265.
[34] ibid 1273.
[35] ibid 1256-1257.
[36] ibid 1256.
[37] ibid 1268.

decision-making. However, if the errors by the system had not been so many or so evidently incorrect, these errors might not have been discovered.[38]

2.5 Identified problems of automated decision making-systems

As has been presented above, wrong or incorrect data equals wrong or incorrect decisions. This is what has given rise to the famous expression regarding these technologies, namely "garbage in, garbage out".[39] In order for an automated decision making-system to be trained well and to be able to make the most appropriate decisions, this will require qualitative, quantitative and non-biased training data.[40] However, if a programmer includes her/his own bias into the data or algorithms or gives it biased training-data which e.g. prefers one group over another, this will be the truth and the basis for the automated decision making-system which in turn will reproduce the same patterns over and over again in its future decisions.[41] As an example, if a system is given records that contains a higher amount of criminal convictions against black people than white people, the following automated decisions will reproduce this pattern. The result will be that the system would target black people as potential criminals to a higher extent than white people, due to its disproportional training-data. Such decisions will be discriminative, if people are being targeted primarily due to their skin color or ethnicity.[42]

2.5.1 Lost in translation

Translating a written legislation into computer code, with all of the legislation's nuances intact, could be a very complex task when legislation, case-law and human legal reasoning are to be converted into mathematics. In other words, the language of a computer compared to our human language is not equivalent.[43] A risk with trying to translate legislations and policies, that requires both complex investigations and to strike a fair balance between different interests, is that their nuances and complex actions are at risk of being removed to simpler 'yes or no'-questions.[44] This could of course enable the automated decision-making to be done much faster and to result in cost-savings for the state, at least in a short-term perspective. However, this will indeed cause distortions of those policies which will likely be to the cost of the institution's expertise of deciding in complex situations.[45]

As could be seen in the case of the Colorado benefit-system, programmers, even though unintentionally or without any intentions, engaged in rulemaking through altering the legal requirements identifying whether individuals were entitled to the benefits or not. This entails a type of rulemaking where programmers can create systems that requires more of, or imposes completely different requirements on, individuals than what the actual policies or legislations does.[46] Are the programmers the only ones to blame for risking the possible distortion of the policies being translated into computer code? Not really. The public employees working at the institutions also risks crossing the line from individual adjudications and to enter into

[38] ibid 1256.
[39] Barocas and Selbst (n 8) 683.
[40] Hildebrandt (n 25) 25.
[41] Barocas and Selbst (n 8) 683-684.
[42] Sina Aghaei, Mohammad Javad Azizi and Phebe Vayanos, 'Learning Optimal and Fair Decision Trees for Non-Discriminative Decision-Making' (2019) Center for Artificial Intelligence in Society <https://arxiv.org/pdf/1903.10598.pdf> accessed 14 May 2019, 1.
[43] Burrell (n 26) 4.
[44] Hildebrandt (n25) 23.
[45] Keats Citron (n 6) 1297.
[46] ibid 1279-1280.

rulemaking in the same way as the programmers, when being influenced by automation bias.[47] They will also risk enforcing new rules creating stigmatizing categories of individuals which will be the basis for future decisions of the automated decision making-systems.[48]

2.5.2 Opacity and ability to understand

Another identified problem with automated systems is that not all save the basis for their decisions. If you are affected by a negative decision which you do not agree with and wishes to know the reasons for, that can be problematic since it would not be possible to go back and see *what* has caused the system to decide in the way that it has. If you cannot see *why* the decision against you have been decided in the way that it has and what it is based on, it will make it very hard, if not impossible, to appeal it or to control whether it has been based one proper or on incorrect information.[49] Even if you strongly suspect that an incorrect decision by an automated decision-system has been made, it can still be hard to appeal it. To decipher the many computer codes behind a system will be a very costly process for simply one case. The chances of a public institution having the ability of doing so in all appeals of automated decisions, will be small.[50] Where the affected individual cannot understand why the decision has been taken, due to the opacity of the system, knowing or estimating your possible abilities to appeal the decision would be almost impossible.[51] This could lead to a majority of the decisions being based on incorrect or false data to never be appealed.

Another problem that will arise if automated systems work secretly without tracking or saving each data that they include in their decisions is to constitute a real issue for the judicial review of such decisions. The US Supreme Court has demanded public institutions to keep a contemporaneous record of their decision-making in order to be able to review not only the actual decisions but also the reasons behind them. Thus, with automated decision making-systems working in secret, such reasons will not exist. Still, a US-court can still find that a negative decision is a violation of the public institution's discretion simply due to a lack of not offering the adequate reasons for why e.g. an application was denied. However, what will be problematic regarding the non-existing reasons will be that the court cannot identify *what* has caused the violation of the institution's discretion. It cannot identify whether this is due to incorrect interpretations of a policy or if it due to incorrect information.[52] If not knowing the reasons behind such decisions, hidden patterns in the automated decision making-systems of differential and possibly discriminative bias due to incorrect information can continue to be used as basis for future decisions.

2.5.3 Concluding remarks

In this chapter, two primary issues with the usage of automated decision making-systems has been identified. First in the process of translating policies and legislation into computer codes, there can be risks of the programmer's own bias or disproportional training-data being included into the systems. This can lead to distortions of the policies and legislations being translated, where they can lose their nuances and important balancing acts of different interests.[53]

[47] ibid 1262-1263.
[48] ibid 1272.
[49] Hildebrandt (n 25) 26-28.
[50] Ledendal, Larsson, Wernberg (n 1) 106.
[51] Hildebrandt (n 25) 28.
[52] Keats Citron (n 6) 1298-1299.
[53] ibid 1297.

Second, regarding the opacity of automated decision making-systems, the possibilities of obtaining the benefits and other support by the state which you are entitled to might depend on your ability to have the time and economy to appeal, and on your knowledge regarding legislations, policies and public institutions.[54] Many of those who does not possess sophisticated knowledge about legislations and policies will likely not question a negative decision due to them trusting the institutions abilities of making competent and correct decisions.[55] Therefore, such incorrect decisions do risk to adversely affect individuals that are perhaps already socially marginalized, having less financial resources, worse literacy and less means and opportunities to challenge decisions by public institutions.

To conclude, the challenges presented of the issues that the usage of automated decision making-systems by public institutions may cause are the following. The possible rulemaking by the programmers can create new requirements and therefore demand more from some, and less from others, e.g. in situations of several individuals applying for the same benefits. Possible bias of programmers, mistakes in translation, negligence in application by public employees, historical prejudices being maintained in the system through biased training-data, or data containing a skewed and incorrect picture that is not corresponding with the real world, will all have dire consequences in decision-makings which public institutions must prepare themselves to combat and prevent.

3. Legislation

In this chapter, the focus will turn to the European and Swedish contexts, and their respective legislations on discrimination. As has been presented, the challenges of the usage of automated decision making-systems can be the disproportional effects on certain groups or individuals in cases of applying for benefits or as simple as trying to board a plane. Due to this, the focus here will be on so called indirect discrimination.

Indirect discrimination is a type of discrimination that can exist even in situations where the legislation applicable to the situation appears to be neutral and to treat individuals alike. However, if the application of such legislation results in a disproportionately negative impact on a specific group of individuals only, without any legitimate reasons justifying such impacts, this would amount to an indirect discrimination. An example can be where it is almost impossible for only one group of society to reach the requirements of a legislation while no other group have the same problem. Here, the first group is being disproportionately affected in a negative way since they have less possibilities of fulfilling the requirements and therefore suffers a disadvantage by not being able to enjoy the benefits of the legislation.[56] This comparison between two groups where one is treated worse than the other, is often the legal requirement for being able to establish whether an indirect discrimination has occurred or not.[57] In the US- and Canadian contexts, such groups are often called stereotypes, being generalized assumptions of common traits that members of a certain group shares.[58]

[54] See example of incorrectly identified parents owing child support in part 2.4, Citron (n 6) 1273.
[55] Keats Citron (n 6) 1272. Also, see example of Colorado Benefit-system, 1256.
[56] Ilias Bantekas and Lutz Oette, *International Human Rights: Law and Practice* (2nd edn CUP 2016) 76-77.
[57] European Court of Human Rights and European Union Agency for Fundamental Rights, *Handbook on European Non-Discrimination Law: 2018 Edition* (Publications Office of the European Union 2018) 58.
[58] Sophia Moreau, 'Equality Rights and Stereotypes' in David Dyzenhaus and Malcolm Thorburn (eds), *Philosophical Foundations of Constitutional Law* (OUP 2016) 286.

In the following parts, European and Swedish legislations and case-law concerning indirect discriminations will be presented.

3.1 European Convention on Human Rights

In accordance with article 14 of the ECHR, Contracting States are bound to respect the right of non-discrimination, which entails not to make any discriminative measures towards individuals based on grounds of "sex, race, color, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status[59]." Such measures would be illegitimate and cannot take place when states fulfill their obligations under the ECHR, securing the rights therein, unless it can be properly justified.[60]

Article 14 of the ECHR is directed towards, and imposes responsibilities on, the Contracting States and their respective public institutions. Such responsibilities are e.g. positive and procedural obligations to primarily ensure that the right to non-discrimination, together with the other rights of ECHR, is not violated by state actors.[61]

In order for an individual to properly claim a violation of discrimination before ECtHR, article 14 only comes into effect when it is claimed in conjunction with the enjoyment of the other rights of ECHR.[62] This rather limited protection of the principle of equality and the right to non-discrimination was the reason for the establishment and adoption of Protocol 12 which was motivated through that an independent protection of discrimination was needed in the ECHR to better comply with other international human rights instruments' level of protection of non-discrimination.[63] However, since Sweden will be of focus in this essay and since Sweden has not signed nor ratified Protocol 12[64], this protocol will not be further dealt with.

3.1.1 ECHR, ECtHR and indirect discrimination

Through the case-law of the ECtHR, indirect discrimination has been included into article 14's prohibition of discrimination. The fundaments of indirect discrimination in the context of ECHR is: first, to have what appears to be a neutral rule or policy. Second, that such rule or policy affects a specific group of people, defined by one of the protected grounds in article 14, in a *significantly* more negative way than to those in a similar situation who are not part of this affected group even though there are no intentions of such effects.[65] The different grounds of illegitimate discriminations established in article 14 is not an exhaustive list of prohibited grounds. This is why the ground 'other status' has been included and given a wide meaning[66] since it is applicable to several different situations.[67]

---

[59] ECHR art. 14.
[60] Rory O'Connell, 'Cinderella comes to the Ball: Art 14 and the right to non-discrimination in the ECHR' (2009) 29 Legal Studies 211.
[61] Samantha Besson, 'Evolutions in Non-Discrimination Law within the ECHR and ESC Systems: "It Takes Two to Tango in the Council of Europe" (2012) 60 The American Journal of Comparative Law 147, 161.
[62] *Carson and Others v. the United Kingdom* App no 42184/05 (ECtHR 16 March 2010) para. 63.
[63] Council of Europe (COE) 'Explanatory Report to the Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms' (4 November 2000) European Treaty Series no. 177, para. 1.
[64] Council of Europe, 'Search on Treaties: Simplified Chart of Signatures and Ratifications' <https://www.coe.int/en/web/conventions/search-on-treaties/-/conventions/chartSignature/3> accessed 11 May 2019.
[65] Besson (n 61) 164.
[66] *Engel and Others v. the Netherlands* (1976) Series A no. 22, para. 72.
[67] *Biao v. Denmark* App no 38590/10 (ECtHR 24 May 2016) para. 89.

Under article 14, Contracting States enjoys a varied margin of appreciation depending on the situation. The margin of appreciation is wide in questions of e.g. economic strategies. This is due to the ECtHR's claims that the State's own public institutions knows better than an international judge what their citizens need and that they are also better equipped to respect the different interests in their societies.[68] In contrast to that, the margin of appreciation in cases concerning racial discrimination is instead non-existing. The Court has, in several cases, established that where a difference in treatment are based solely, or to a *decisive extent*, on a person's ethnic origin, such measure can never be justified within a democratic society.[69]

The ECtHR has established, as in the *Hoogendijk*-case, that statistics cannot prove discriminative practices by themselves. However, where significant differences can be shown by official statistics, the Court will not ignore such evidences. If the applicant can show that such differences is the result of the application of what at first appeared to be a neutral policy or measure, the responsibility will be turned over to the respondent, the Government, to provide proof that such differences are based on objective and reasonable justifications.[70] Such justifications could be if the policy or measure fulfills a legitimate aim or if the relationship between the measures used and the aim to be achieved is proportional.[71] In other words, when individuals are being treated in a way that is less advantageous than others, and such treatment is not pursuing a legitimate aim or where the measures imposed to reach such aim are not proportional, this would amount to indirect discrimination under article 14.[72]

In the case of *Carson and others v. the UK*, a case concerning a national pension system, the ECtHR established its position as an international court deciding only on the cases before it in principle. The Court meant that it will focus on whether a national legislation do discriminate people of similar or analogous situations rather than to focus on individual facts of the applicant in question. The only focus in the proceedings will be on whether article 14 is compatible to the situation and claimed violations before the Court or not.[73]

3.2 European Union

The principle of non-discrimination, together with the principle of equality, is considered to be an EU-principle with constitutional status and a cornerstone of EU-law, with a multi-layered protection. The high importance and protection of non-discrimination by EU-law can be explained as being due to the aim of advancing towards an 'ever closer Union'. The multi-layered protection of non-discrimination is meant to offer a wide and comprehensive protection but has been accused of being confusing in cases concerning the protection by both primary and secondary EU-law.[74] For the sake of this essay, the possible confusions, strengths or weaknesses of the multi-layered protection of non-discrimination will not be further investigated. Instead, a presentation of some of the different legislative layers will follow.

[68] *Carson and Others* (n 62) para. 61.
[69] *Biao* (n 67) para. 94. Also: *Timishev v. Russia* App nos 55762/00 and 55974/00 (ECtHR 13 December 2005) para. 58; *D.H and Others v. the Czech Republic* App no 57325/00 (ECtHR 13 November 2007) para. 176.
[70] *Hoogendijk v. the Netherlands* App no 58641/00 (ECtHR 6 January 2005).
[71] *Carson and Others* (n 62) para. 61.
[72] *Biao* (n 67) para. 90
[73] *Carson and Others* (n 62) para. 62.
[74] Muir (n 11) 2-4.

3.2.1 Primary law protecting non-discrimination

The principle of non-discrimination is established as a fundamental value upon which the Union is built, as well as a value that is common to all Member States, and which should be given the highest protection on the internal market of the EU.[75]

The Union is generally obliged to combat discrimination when implementing its polices.[76] EU-institutions are given the task to, within their conferred powers, combat and prevent discriminations based on age, disability, racial or ethnic origin, religion or belief, sex or sexual orientation.[77] In the Union's protection of non-discrimination, social progresses are included such as the Union's aim of ensuring equal treatments of men and women. The constitutional status of non-discrimination, applying equally to both EU- and member states public institutions, is driven by the aim of assuring that all of these institutions adhere to the principle of non-discrimination in their interactions with individuals and that bias in the public structures are eliminated.[78]

In the EU-Charter, the prohibited grounds for any discrimination are similar to the treaties but are more extensive. These prohibited grounds are, 'sex, race, color, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation'. It does also, like the treaties, have a special emphasis on prohibiting discrimination based on nationality.[79] The list of prohibited grounds for discrimination, similar to that of ECHR's article 14, is not exhaustive in order to enable it to develop and to encompass all possible situations of illegitimate discrimination.[80]

Due to the general character of primary EU-law, the provisions prohibiting discrimination refers to *any* type of discrimination without specifying more in detail the obligations of prohibiting direct or indirect discrimination. However, the primary EU-law does apply and offer protection in cases of indirect discrimination, since such violations are included under the EU-Charter's right to a judicial recourse[81]. This right is, due to the scope of the EU-Charter, attributed to EU-citizens only.[82] Here, indirect discrimination requires that the effects of legislation that seemed to be neutral, significantly disadvantages a group or an individual more than others. This needs to be shown through that the ones being disadvantages can show the similar characteristics they share, in order to compare them to the others that was not disadvantaged due to that they shared other similar characteristics.[83]

3.2.2 Secondary law and case-law on indirect discrimination

Apart from the EU primary law, with their non-exhaustive lists of prohibited grounds for discriminations, this is not the case with secondary EU-law due to that these directives are assigned to certain, narrow areas and does therefore include specific prohibited grounds. The ECJ has in their case-law established that the prohibited grounds for discrimination of

[75] TEU art. 2-3.
[76] TFEU art. 10.
[77] ibid art. 19.
[78] Muir (n 11) 4-7.
[79] Charter of Fundamental Rights of the European Union (adopted 7 December 2000, entered into force 26 October 2012) Official Journal of the European Communities 2012/C 326/02 (EU-Charter) art. 21.
[80] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 62.
[81] Charter of Fundamental Rights of the European Union (n 79) art. 47.
[82] EUR-Lex, 'Glossary of Summary: Non-Discrimination (the principle of)' < https://eur-lex.europa.eu/summary/glossary/nondiscrimination_principle.html> accessed 19 May 2019.
[83] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 53.

secondary law cannot be extended to encompass other grounds than those being established in the written directives in question.[84]

Several directives explain indirect discrimination similarly to the primary EU-law and definition by ECHR, with the difference that each Directive specifies the common traits that individuals needs to share in order to be victims of indirect discrimination. This, due to that they apply to a specific area of law and therefore have a much more limited scope.[85] Examples of these common traits on which people that are disadvantages will likely be found to have been victims of indirect discriminations are e.g. ethnic or racial origin[86] and sex[87].[88]

The ECJ has on several occasions found national policies of Members States to be indirectly discriminative towards specific individuals in accordance with directives. One example is a Spanish case concerning whether the Spanish national pension system could be considered as discriminative or not in relation to a directive on the equality between men and women concerning social security. The Spanish Social Court of Barcelona referred the case to the ECJ asking for a preliminary ruling concerning the applicant's, a female part-time worker, claim that the Spanish Governmental pension system breached EU-secondary law regarding the prohibition of discrimination based on sex[89]. The Spanish provision in question here, demanded part-time workers to pay pension-fees for a longer period of time than those who worked full-time, while the former would still be granted a lower amount of pension funds than the amount granted to the latter. The applicant claimed that such provision had an indirect discriminative effect on her, referring to statistics showing that a majority, around 80 %, of Spanish women had these types of part-time contracts.[90] The comparator group here was men, who according to the same statistics instead had full time contracts to a much higher extent than women.[91]

The statistics were undisputed by the Spanish Social Court[92], and the ECJ established that the national pension system was contrary to the Directive in question, but that these differences in demands could be justified by objective reasons not having any connections to discrimination based on sex. Such would e.g. be if the regulations of the pension system would reflect a legitimate social policy aim of the Spanish institution, and that these different requirements in the system were appropriate and necessary in order to reach that aim. However, the ECJ could not find that the differences between part-time and full-time workers were necessary in order to, as the Spanish Government claimed, maintain the balance of the pension system or that they were genuinely necessary to protect the system's existence. Therefore, the Spanish pension system was found to violate EU-secondary law and to indirectly discriminate women

[84] ibid 62.
[85] ibid 53.
[86] Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/22, article 2(2)(b).
[87] Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation [2006] OJ L204/23, art. 2(1)(b), and; Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L373/37, art. 2(b).
[88] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 53.
[89] Council Directive 79/7/EEC of 19 December 1978 on the progressive implementation of the principle of equal treatment for men and women in matters of social security [1979] L 6/24, art. 4.
[90] Case C-385/11 *Isabel Elbal Moreno v. Instituto Nacional de la Seguridad Social (INSS) and Tesorería General de la Seguridad Social (TGSS)* ECJ (22 November 2012) para. 12.
[91] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 55
[92] Case C-385/11 *Isabel Elbal Moreno* (n 90) para. 31.

on grounds of sex since the Spanish Government failed to show that the difference in treatments were objectively justified.[93]

In this case, the facts presented before the ECJ existed and showed the actual disadvantages between men and women regarding the allowed benefits by the national pension system. However, the ECJ made an exception in an earlier case from 1996 regarding the requirement of showing an actual comparator not being disadvantaged regarding indirect discriminations. In a preliminary ruling regarding a system of social benefits in the United Kingdom (UK) for funeral expenses of employees being buried within the UK, the question before the ECJ was if UK legitimately could make a difference in treatment, approving benefits or not, based only on the deceased employee's nationality.[94] The ECJ established that it, based on its own previous case-law, can be sufficient for the applicant to only show that there is a *risk* of disadvantages towards a group based on nationality which will not apply to those being UK-citizens.[95] In other words, an actual disadvantage in the situation at hand was not needed to be shown.

3.2.3 Concluding remarks

To conclude, the ECHR and EU-law entails the same requirements of indirect discrimination. First, the responsible rule, legislation or policy needs to be formulated neutrally.[96] Second, that such neutral policy causes a specific group to be disadvantaged, where the focus here is on the effects that the policy causes. Also, this specific group needs to be considered as a 'protected group' being discriminated on one of the prohibited grounds, to be able to establish that indirect discrimination has taken place. This can be demonstrated through statistical evidence showing that a large number of the ones affected are part of such group, as in the Spanish case showing that a big number of the ones suffering disadvantages were women. This also includes having a comparator, namely the opposite group of those *not* being disadvantaged. This is crucial since showing that one group is being disadvantaged under the other in the actual situation at hand, is necessary for being able to establish indirect discrimination.[97] Still, the ECJ has made an exception from the requirement of showing a comparator not being disadvantaged in the case at hand, where it can be shown that the applicable rule *risks* disadvantaging a specific group.[98] Third and last, disadvantages and differences by a neutral rule can still be legitimate if they can be objectively justified. This would require the state responsible to show that it has a legitimate aim and that the measures imposed to reach such aim are both necessary and proportionate.[99]

3.3 Sweden

3.3.1 Legislation

In the Swedish Constitution, it is established that all public power and public institutions are responsible for working for everyone's opportunity of equality and participation in society, through combating and eliminating discriminations based on grounds of sex, color, ethnical or

[93] ibid paras. 28-38.
[94] C-237/94 *John O'Flynn v Adjudication Officer* [1996] ECR I-02617 para. 12.
[95] ibid paras. 18-21.
[96] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 54.
[97] ibid 56-58.
[98] C-237/94 *John O'Flynn* (n 94) para. 21.
[99] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 93.

national origin, linguistic or religious affiliation, disability, sexual orientation, age, or any other circumstances affecting individuals.[100]

Further, the Swedish Discrimination Act is established to combat discrimination and to promote equal opportunities and rights.[101] This Act is applicable to situations where public employees meets individuals, e.g. in the cases of applications for social insurances or benefits, placing obligations on the public employees to which they must abide.[102] Indirect discrimination is explicitly prohibited through the Act but includes more prohibited grounds on which indirect discrimination is unfair than ECHR and EU-law does. Beyond the prohibited grounds of EU-law, ECHR and the Swedish Constitution, the Swedish Discrimination Act also prohibits grounds based on transgender identity and expression.[103]

Similar to the requirements of ECHR and EU-law regarding indirect discrimination, Swedish law contains similar ones.[104] The requirements that needs to be shown to be able to establish an indirect discrimination, is a neutral rule or policy that disadvantages an individual on a prohibited ground, unless it can be objectively justified through a demonstrated legitimate aim and proportional, necessary measures to reach such aim. A comparator of the actual situation which, in opposition to the victim in question, has not been disadvantaged, must be shown.[105]

However, a few years ago, in a case before the Swedish Court of Appeal, an exception from the need of showing a comparator was made. Here, the Court did not establish such a comparator but still managed to find that there had been an indirect discrimination against the applicant by the respondent, a public institution.[106] This exception might seem insignificant, but the adjudication of this case will be presented below, and further evaluated in part 3.5.

3.3.2 Case-law

The Swedish Discrimination Act does, through an amendment in 2012[107], also contains provisions explicitly prohibiting discriminations based on sexual orientation within health and medical care, social services and in relation to the national social insurance system.[108] As a contracting and member state to both ECHR[109] and the EU[110], Sweden is responsible of respecting the obligations arising from these legislations. However, this amendment was driven by the fact that Sweden considered the protection of EU-law regarding discrimination on grounds of sexual orientation in health or medical care to not be sufficient. Therefore, the

[100] Swedish Instrument of Government (1974:152) chapter 1 art. 2.
[101] Swedish Discrimination Act (2008:567) chapter 1 art. 1.
[102] Diskrimineringsombudsmannen, 'Inom Vilka Områden Gäller Diskrimineringslagen?' (1 June 2016) <http://www.do.se/om-do/vanliga-fragor/fragor-om-diskriminering/inom-vilka-omraden-galler-diskrimineringslagen/> accessed 19 May 2019.
[103] Swedish Discrimination Act (n 101) chapter 1 art. 4(2).
[104] ibid.
[105] Birgitta Nyström, 'Grundläggande Rättigheter i EU-rätten – Påverkan på Svensk Lagstiftning och Rättspraxis' in Annika Staaf and Lars Zanderin (eds), *Mänskliga Rättigheter i Svensk Belysning*' (2nd edn, Liber 2011) 69.
[106] Göta Hovrätt Case no T 2134-14.
[107] Law on the Amendment of the Swedish Discrimination Act 2012:673.
[108] Swedish Discrimination Act (n 101) chapter 2 arts. 13-14.
[109] Council of Europe, '47 Member States' < https://www.coe.int/en/web/portal/47-members-states> accessed 17 May 2019.
[110] European Union, 'The 28 Member Countries of EU' < https://europa.eu/european-union/about-eu/countries_en> accessed 17 May 2019.

provisions of the amendment were seen as a complement to EU-law, creating a greater, national protection of non-discrimination.[111]

The Swedish Court of Appeal found an indirect discrimination based on grounds of sexual orientation by a Swedish Health Care-center in a case concerning a homosexual woman who had paid a ten times larger fee than a heterosexual woman for a similar insemination attempt. The applicant did therefore, represented by DO, claim that the Health Care-institution's difference in treatment due to a patient's sexual orientation was discriminative.[112]

The Swedish Court of Appeal did not find that the applicant's situation could be compared with that of a heterosexual woman, since the latter's need of insemination assistance is considered to be due to injuries or sickness and not to "biological lacking's" as in the case of homosexual couples. As the provision that the Health Care-institution relied on regarding the fees appeared to be neutral, the Court moved on to investigate whether there were any objective reasons for the institution to have these different fees that could legitimize this difference, if they were necessary and proportional to reach that aim[113]. The Court found that the reason behind the more expensive fee for homosexual women was due to a choice by the regional health care-center to offer this treatment to a higher cost, due to that such treatment is for non-medical reasons, while the same treatment for heterosexual couples is due to medical reasons which includes such treatments under the medical responsibility of the institution. The Court accepted this as an objective reason for demanding the different fees.[114]

However, the Court then moved on to investigate whether this was the least measure possible, or if there could be any less intrusive measures available. Here, there were no sufficient evidences supporting the view that such high fees were the least measure that the institution could use. Therefore, the Court came to the conclusion that even though the Health Care-institution had legitimately established the different fees, it could not be shown that the profits of these measures overweighed the harm that they caused. In other words, the measures imposed were not proportional to the aim. Therefore, the Court ruled that the institution had indirectly discriminated the applicant.[115]

In a later case before the Swedish Labor Court, the DO representing the applicant of this case, referred back to the exception of the ECJ from 1996 claiming that even though this ECJ case-law had not been included into the preparatory works regarding indirect discrimination of the Swedish Discrimination Act, it should still be respected by the Courts.[116] The Court followed DO's claim by investigating the case without evidences of a comparator, but did conclude that the disputed measure did not amount to an indirect discrimination of the applicant.[117]

3.4 Current legislation and technological challenges

In this part, a first summary and conclusion of what has been presented up until now will follow. Can the ECHR, EU-laws and Swedish laws on indirect discrimination handle the challenges that automated decision making-systems can create?

[111] Göta Hovrätt (n 106).
[112] ibid.
[113] Swedish Discrimination Act (n 101) chapter 1 art. 4(2).
[114] Göta Hovrätt (n 106).
[115] ibid.
[116] AD (Arbetsdomstolen) 2017 nr 65.
[117] ibid.

3.4.1 Legal requirements of indirect discrimination

In the case-law presented concerning indirect discriminations, the ECtHR and ECJ did first identify the measure, policy or law as being neutral, then systematically go through the applicants claims of alleged harm, consider the respondents claimed justifications and reasons, and compare the applicant's situation with an appropriate comparator in order to see if the alleged harm had caused disadvantages to the applicant.[118] In the Swedish context, similar systematic steps were followed. The regulation causing the different treatment in question was formulated neutrally, the applicant's alleged harm was evaluated, and the respondents claimed objective justifications was investigated, compared to the measures they had used to reach the claimed aim.[119] Also, in the EU- and Swedish legal contexts, there seems to exist an exception where the applicant can only show a risk of disadvantages only affecting the applicant and not the presumed comparator.[120]

3.4.2 Bias

As was identified as a challenge of the usage of automated decision making-systems, bias from biased training-data, the programmer's own bias, or, improper translations of legislations and policies into computer code can lead to decisions that, e.g. in the case of social benefits, demands more information from some or immediately denies the application of others. As was the case with the No Fly-system, where individuals are incorrectly presumed to belong to a certain category based on incorrect data, such errors can have severe impacts on individuals, e.g. when possibly being denied benefits that they are entitled to. If such incorrect or biased training-data is given to the automated decision making-system to train on, it will reproduce the same discriminative patterns in its future decisions. Where the system's algorithms analyze data that contains incorrect facts, or where the proportions of the information do not match reality, this will still be the systems own reality.[121] When incorrect information is encoded, even though unintentionally, the algorithms of the automated decision-system will learn to uphold patterns where one group is being benefited over another and to maintain detrimental historical prejudices against already marginalized groups.[122]

Will the ECHR, EU- and Swedish legislation be able to combat such bias and arbitrary new rules by the automated systems? As was seen in the examples from the US-context, the ability of proper scrutiny of the public institutions' decisions was impaired by the automated decision making-systems since it was not possible to see what data the system had used and based its decisions on. These systems would risk causing the same issues in the European and Swedish contexts as well. However, without knowing the reasons, the courts could still find that an action by public institutions are arbitrary due to lack of presented justifications for discriminative effects against the applicant.[123] This could lead to that some affected individuals would be awarded compensation for incorrect decisions, based on the harm the applicant claims to have been a victim of and to the lack of justifications for such effects, since those reasons does not exist. However, knowing what the rule, law or policy behind such case has in fact required, will be a complex task for the courts to figure out. As was mentioned regarding judicial review, if the courts cannot see the actual rules that were applied to a case, their adjudications and ability to combat possibly discriminative ones, will be

[118] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 54-58.
[119] Göta Hovrätt (n 106).
[120] E.g. C-237/94 *John O'Flynn* (n 94) and AD (n 116).
[121] Barocas and Selbst, (n 8) 683-684.
[122] Moreau (n 58) 291.
[123] *Martzaklis and Others v. Greece* App no 20378/13 (ECHR 9 July 2015) paras. 71-75.

severely wounded.[124] If the courts are not able to identify the laws, rules or policies applied in decisions, they will not be able to identify the incorrect, distorted or arbitrarily new ones either. Such situation could lead to the courts only helping a few of those affected through court proceedings, while the systems continues to spit out several new and equally incorrect decisions, with the risk of reinforcing historical and detrimental prejudice and bias each day.

3.4.3 Opacity

The identification of the indirect discriminations as described above in the ECtHR, ECJ and Swedish case-law, mainly requires that there is an appropriate comparator available to each situation in order for an event of indirect discrimination to be established.[125]

What would happen then when the alleged discrimination is a result of the usage of an automated decision making-system, whose reasons for the decisions are not presented, the data that has been used as the basis for the decision are not saved, and the tracks of the system's work are secret and non-available? An example here is the one of the Colorado benefit-system presented above in the US-context, where today after more than a decade ago since the systems was corrected to function better, individuals cannot know whether the decisions by the system are correct or if they are based on incorrect facts once again.[126]

Going back to the basis of showing comparators of ECHR, EU- and Swedish-law, if you do not know why a decision has been taken, on what basis, due to which factors or due to lack of what information, how can you be sure of which comparator is appropriate for your case or even on what characteristics you might have been discriminated on? It will be equally hard for these courts to establish on which grounds an individual might have been discriminated and which comparator is suitable to properly compare that situation with, without knowing the reasons or even the applied rules of a decision.[127] The examples from the Swedish courts, and the exception from the ECJ in 1996, will therefore be further evaluated below.

3.5 No comparator, no discrimination?

If our legislation in force today will struggle in combating the issues arising from public institution's usage of automated decision making-systems, should we instead encounter the issues of possible indirect discriminations in a different way?

Sophia Moreau[128] has, in her theory on equality rights, argued that demanding victims of alleged discriminations to show which group, stereotype, they belong to in order to be able to show that a discrimination has taken place through comparing one stereotype to another, can entail negative effects on those individuals. Moreau argues that we should instead view unfair discrimination from a harm-based perspective. Her theory is anchored in the US- and Canadian contexts and she argues that such a harm-based perspective will lead to that the main focus in discrimination-proceedings will not be on demanding the victims to show which stereotypes he or she does and does not belong to[129]. Instead, focus would be on the harm that the victims claims to have suffered and to try to understand that claim and the

[124] Keats Citron (n 6) 1298.
[125] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 57-58.
[126] Keats Citron (n 6) 1256.
[127] ibid 1298.
[128] Associate Professor at the University of Toronto.
[129] Which is a common demand in discrimination-cases in the American and Canadian contexts. See examples in: Moreau (n 58) 286-288.

situation leading to these allegations. This could instead be the basis for establishing whether there has been an unfair discrimination or not.[130]

In the US- and Canadian contexts, stereotypes can be explained as a generalization of a group of people that are treated as if everyone were identical. The US-courts have articulated that being treated differently based on belonging to a certain stereotype is what *makes* a discrimination unfair. Without being able to show that you belong to a stereotype sharing similar characteristics, the courts argues that discrimination cannot be proven since it would not be possible to show that the claimed alleged discrimination has caused lesser treatment than to other groups of the same situation.[131] The same requirements are generally established for the ECHR, EU- and Swedish law[132], the generalizations instead being called 'comparators' and 'characteristics'.[133] The issues of generalizing too much, is that these assumptions can be false and based on pure myths, over-inclusive and deprive situations of appropriate individual assessments.[134]

In cases of discriminations, a victim needs to show affiliation to one presumed group or stereotype. However, what if the only stereotype available falsely attributes capacities which are incorrect according to the victim in question? In a Canadian case about homosexual couples not having the legal right to marriage as heterosexual couples had, a false stereotype was found. In the Canadian court's investigation, the justified exception from the constitutional principle of equality, excluding homosexual couples from being able to marry, was in fact found to be based on incorrect myths. These myths claimed incorrect assumptions about same-sex couples' inability to both raise children and to have long relationships. From this perspective of the law, every same-sex couple that was not married was considered as belonging to this stereotype that attributed them with relationship- and nursing incapacities. These laws did therefore uphold old and completely incorrect assumptions of same-sex couples, and the Canadian court did therefore find this exclusion from marriage as being directly discriminative towards same-sex couples.[135]

Even though this case was established to be a direct discrimination which is not the topic for this essay, the point made here is interesting. As Moreau argues her harm-based theory to be used, the case-law of the Swedish Court of Appeal, and the exception made by the ECJ in 1996, appears to be corresponding. As has been established, Swedish legislation demands a victim of indirect discrimination to show the harms in relation to a comparator not sharing the same characteristics as the victim.[136] However, the model of the Court of Appeal shows a possibility of Swedish courts ignoring this demand.[137]

Perhaps this could be a more suitable solution in order to meet the challenge with opaque systems not always presenting their reasons or the data it bases its decisions on. Instead of cases being denied due to not being able to show a comparator, or the incapacity of showing on which traits a difference in treatment depended on, the courts could focus more on the actual harm that the applicant claims to have suffered, e.g. whether freedom or autonomy was

[130] Moreau, (n 58) 299-303.
[131] ibid 286-287.
[132] Nyström (n 105) 69.
[133] European Court of Human Rights and European Union Agency for Fundamental Rights (n 57) 57-58.
[134] Moreau (n 58) 289-292.
[135] ibid 289.
[136] Nyström (n 105) 69.
[137] Göta Hovrätt (n 106).

limited, or if a measure has demeaned the applicant.[138] The model of the Swedish Court of Appeal, even though not being from the highest legal instance, shows the possibilities of finding indirect discrimination based on the actual claimed harm of the individual and the lack of objective justifications by the Government. If Swedish courts would follow this model with these steps in adjudicating cases regarding indirect discriminations, our Swedish legislation could be better equipped to meet the challenges arising from public institutions usage of automated decision making-systems, then the ECHR and the current EU-laws.

To some extent, most societies accepts that their governments and public institutions will generalize in order to be effective. However, false generalizations that are based on historical stigmatizations and disadvantages must be both identified and eliminated by the responsible government. [139] This is more important now than ever, since democratic societies must be maintained even in times of increasingly new usages of automated decision making-systems by public institutions.

[138] Moreau (n 58) 299-300.
[139] ibid 291.

Conclusion

The original hope of automated decision making-systems to be almost error resistant has in this essay been rebutted. However, this hope seems to have survived in our minds through the automation bias, affecting the human operators to almost be paralyzed by their over-belief of these systems' capacities. It can therefore be important to note here that a created automated system is not automatically a better decision-maker than a human just because of its being. It is a compilation of codes and algorithms which needs to be programmed into using correct information and to reinforce those patterns and results which are just and correct.

As has been presented, two main challenges of the automated decision making-systems have been identified. First, the possible inclusion of both a programmer's own personal bias or of biased training-data, which risks resulting in a biased decision-system which reproduces the bias, historical stigmatizations and discriminative categorizations that it has learned. There is also a risk of new rulemaking when legislation is being incorrectly translated or potentially changed into being more easily converted into codes. This can lead to that these systems both denies or approves applications on grounds that no one can expect, which are difficult to identify and therefore based on reasons that are not accessible. These non-accessible, complex and sometimes hidden reasons or bases that a decision is made upon constitutes the second challenge being these systems' opacity. A system working incorrectly by e.g. basing its decisions on incorrect or improper data, risks not being revealed due to the impossibility of finding out the reasons for why the system has decided in the way that it has.

In relation to these challenges, the current EU-, ECHR and Swedish legislations could be able to combat the bias and opacity of the systems to some extent. By an ordinary systematical adjudication of an alleged indirect discrimination, the courts could find a decision by an automated system to be indirectly discriminative based on the inability of the public institution to provide objective justifications of the disadvantage suffered by the applicant. However, even though such possibility exists, there can be a real issue of establishing on which grounds the applicant was actually discriminated on and compared to which others that has not been disadvantaged in the same way. Since the legislations presented generally has these requirements of indirect discriminations, several cases of real indirect discriminations could, due to hidden bias in the opaque systems, falsely be denied access to the courts.

As suggested through Moreau's theory, and through the model of the Swedish Court of Appeal, this could possibly be better encountered if the legal focus in fact were on the alleged harm of the applicant, instead of on her/his specific traits and on the correct comparator of the case. The potential bias and incorrect decisions will most likely have indirect discriminative effects on the applicant, while the opacity of the systems will make it impossible to identify a suitable comparator. By focusing on the claimed harm of the applicant, more cases of indirect discrimination can be enabled access to courts, where individuals would not need to attribute false traits to themselves just to make up a comparator simply for getting access to courts, a procedure that might be indirectly discriminative in itself in relation to the applicant.

To conclude, it might seem typical for a legal essay to present negative features of new technologies, which might come across as simple fears of lawyers eventually being replaced by automated decision making-systems. On the contrary, the challenges presented *must* be discussed in a legal context in order for them to be met in the best way possible. Automated decision making-systems are probably here both to stay and to be used to a further extent by public institutions. We must then make sure that this can be done while fundamental human rights are still being respected.

Bibliography

Legislation and Public Materials

Charter of Fundamental Rights of the European Union (adopted 7 December 2000, entered into force 26 October 2012) Official Journal of the European Communities 2012/C 326/02 (EU-Charter)

Consolidated Version of the Treaty on European Union [2012] OJ C326/13 (TEU)

Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C326/47 (TFEU)

Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin [2000] OJ L180/22

Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L373/37

Council Directive 79/7/EEC of 19 December 1978 on the progressive implementation of the principle of equal treatment for men and women in matters of social security [1979] L 6/24

Council of Europe (COE) 'Explanatory Report to the Protocol No. 12 to the Convention for the Protection of Human Rights and Fundamental Freedoms' (4 November 2000) European Treaty Series no. 177

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

Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation [2006] OJ L204/23

Law on the Amendment of the Swedish Discrimination Act 2012:673

Swedish Discrimination Act 2008:567

Swedish Instrument of Government (1974:152)

Cases and Awards

AD (Arbetsdomstolen) 2017 nr 65

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

Carson and Others v. the United Kingdom App no 42184/05 (ECtHR 16 March 2010)

Case C-385/11 *Isabel Elbal Moreno v. Instituto Nacional de la Seguridad Social (INSS) and Tesorería General de la Seguridad Social (TGSS)* ECJ (22 November 2012)

C-237/94 *John O'Flynn v Adjudication Officer* [1996] ECR I-02617

D.H and Others v. the Czech Republic App no 57325/00 (ECtHR 13 November 2007)

Engel and Others v. the Netherlands (1976) Series A no. 22

Göta Hovrätt mål nr T 2134-14

Hoogendijk v. the Netherlands App no 58641/00 (ECtHR 6 January 2005)

Martzaklis and Others v. Greece App no 20378/13 (ECHR 9 July 2015)

Timishev v. Russia App nos 55762/00 and 55974/00 (ECtHR 13 December 2005)

Literature and articles

Aghaei S, Javad Azizi M, Vayanos P, 'Learning Optimal and Fair Decision Trees for Non-Discriminative Decision-Making' (2019) Center for Artificial Intelligence in Society <https://arxiv.org/pdf/1903.10598.pdf>

Bantekas I, and Oette L, *International Human Rights: Law and Practice* (2nd edn CUP 2016)

Barocas S, and D. Selbst A, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671

Besson S, 'Evolutions in Non-Discrimination Law within the ECHR and ESC Systems: "It Takes Two to Tango in the Council of Europe" (2012) 60 The American Journal of Comparative Law 147

Burrell J, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) Big Data & Society <https://journals.sagepub.com/doi/pdf/10.1177/2053951715622512>

Gallé M, 'What Can I Do Now? Guiding Users in a World of Automated Decisions' (2018) Xerox Research Centre Europe <https://arxiv.org/pdf/1701.03755.pdf>

Guthrie Ferguson A, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement* (New York University Press 2017)

Hildebrandt M, 'Law as Computation in the Era of Artificial Legal Intelligence: Speaking Law to the Power of Statistics' (2018) 68 University of Toronto Law Journal 12

Keats Citron D, 'Technological Due Process' (2008) 85 Washington University Law Review 1249

Larus J, and Hankin C, 'Regulating Automated Decision Making' (2018) 61(8) Communications of the ACM 5

Ledendal J, Larsson S, Wernberg J, *Offentlighet i det Digitala Samhället: Vidareutnyttjande, Sekretess och Dataskydd* (Norstedts Juridik 2018)

Moreau S, 'Equality Rights and Stereotypes' in Dyzenhaus D, Thorburn M (eds) *Philosophical Foundations of Constitutional Law* (OUP 2016)

Monteith S, and Glenn T, ’Automated Decision-Making and Big Data: Concerns for People With Mental Illness’ (2016) 18(12) Current Psychiatry Reports 1

Muir E, *EU Equality Law: The First Fundamental Rights Policy of the EU* (OUP 2018)

Naarttijärvi M, ’Rättsstatlighet och Algoritmiska Svarta Lådor’ in Edström Ö, Lindholm J, Mannelqvist R (eds), *Jubileumsskrift till Juridiska institutionen 40 år* (Juridiska Institutionen Umeå Universitet 2017)

Nyström B, ’Grundläggande Rättigheter i EU-rätten – Påverkan på Svensk Lagstiftning och Rättspraxis’ in Staaf A and Zanderin L (eds) *Mänskliga Rättigheter i Svensk Belysning*’ (2nd edn, Liber 2011)

O’Connell R, ‘Cinderella comes to the Ball: Art 14 and the right to non-discrimination in the ECHR’ (2009) 29 Legal Studies 211

Sajó A, and Uitz R, *The Constitution of Freedom: An Introduction to Legal Constitutionalism* (OUP 2017)

Sauer J, Chavaillaz A, Wastell D, ‘Experience of Automation Failures in Training: Effects on Trust, Automation Bias, Complacency and Performance’ (2016) 59 Ergonomics 767

Wachter S, Mittelstadt B, Russell C, ’Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR’ (2018) 31 Harvard Journal of Law & Technology 841

Other Sources

Council of Europe, ’47 Member States’ < https://www.coe.int/en/web/portal/47-members-states>

Council of Europe, ’Search on Treaties: Simplified Chart of Signatures and Ratifications’ <https://www.coe.int/en/web/conventions/search-on-treaties/-/conventions/chartSignature/3>

Diskrimineringsombudsmannen, ‘Inom Vilka Områden Gäller Diskrimineringslagen?’ (1 June 2016) <http://www.do.se/om-do/vanliga-fragor/fragor-om-diskriminering/inom-vilka-omraden-galler-diskrimineringslagen/>

EU Commission, ‘Types of EU Law’ < https://ec.europa.eu/info/law/law-making-process/types-eu-law_en>

EUR-Lex, ‘Glossary of Summary: Non-Discrimination (the principle of)’ < https://eur-lex.europa.eu/summary/glossary/nondiscrimination_principle.html>

EUR-Lex, ‘Sources of European Union Law’ < https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM%3Al14534>

European Court of Human Rights and European Union Agency for Fundamental Rights, *Handbook on European Non-Discrimination Law: 2018 Edition* (Publications Office of the European Union 2018)

European Union, ’The 28 Member Countries of EU’ < https://europa.eu/european-union/about-eu/countries_en>

Sveriges Kommuner och Landsting, ‘Automatisering i Välfärden: Möjligheter och Utmaningar för Kommuner och Regioner’ <https://rapporter.skl.se/automatisering-i-valfarden.html>