UPPSALA
UNIVERSITET

Department of Law
Autumn Term 2020

Master's Thesis in European Law and Legal Informatics
30 ECTS

Artificial Adjudication and Fundamental Human Rights

A study of artificial intelligence as a judge in light of the right to a fair trial of Article 6 ECHR

Author: Emanuel Björn Bergqvist
Supervisor: Professor Torbjörn Andersson

"*In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move.*"

- Douglas Adams

Preface

This thesis marks the end of my studies at Uppsala and I would like to take the opportunity to express my gratitude to all the friends I have met along the way as well as my family for their encouragement and support. Special thanks to my supervisor Torbjörn for providing valuable feedback during the writing and Anton for giving excellent thoughts on the intricacies of artificial intelligence.

Abstract

The goal of this thesis is to analyse and discuss the use of artificial intelligence within the judiciary. By looking at the role of judges and courts from a perspective of a fair trial, in the form of Article 6 of the European Convention of Human Rights, what characterises a fair trial can be determined. The discussion focuses on the collision of different properties of artificial intelligence and the aforementioned characteristics of a fair trial in order to evaluate whether or not an artificial intelligence is appropriate as a replacement for human judges.

The discussion boils down to transparency, impartiality, independence, accountability and a human element. If an artificial intelligence lacks transparency then there can be legitimate fears that the courts lack impartiality and independence. This could possibly damage the confidence in the judicial system. Since an artificial intelligence cannot be held to the same standards of accountability as humans a problem lies in how we think about accountability. Another problem is that giving up sovereignty to adjudicate removes a human element that might not be so easily dismissed when justifying adjudication. A solution may be to establish special courts for artificial judges with the possibility to appeal to a court of humans until we have evaluated AI judges more thoroughly.

Table of Contents

Abbreviations and Acronyms

AI	Artificial Intelligence
CEPEJ	European Commission for the Efficiency of Justice
CoE	Council of Europe
ECHR	Convention for the Protection of Human Rights and Fundamental Freedoms
ECtHR	European Court of Human Rights, *sometimes also referred to as the Strasbourg Court*
EU	European Union

1 Introduction

1.1 Background

Does the future hold a dystopian nightmare or a prosperous judiciary when it comes to artificial intelligence? Can we guarantee that artificial intelligence will not lead to the structural failure of the judiciary or undermine fundamental human rights that have been hard fought for? These are some of the questions that comes to mind when discussing the implementation of artificial intelligence within the judiciary.

Since the technology underlying artificial intelligence currently has innate technical limitations and inadequacies, when looking at AI from the perspective of justice, it is interesting to analyse how these issues may affect the judiciary if artificial intelligence was put in an adjudicatory role. As the use of artificial intelligence within the judiciary will most likely grow from today's use in analysis and prediction it is also crucial that the implementation is made in a way that respects the rights, freedoms and principles that we regard as cornerstones in today's society.

While the thought of artificial intelligence might strike fear or concern in some people it is important to look at the benefits that can come from proper use of the technology. Should the technology be implemented in a way that upholds our current standards it might serve a valuable purpose regarding efficiency or cost effectiveness. However, the question still stands whether or not the concept of fairness can ever be upheld by other intelligences than humans.

A human has never previously been subject to adjudication in courts by other intelligences than humans. To be judged by one's peers is in some jurisdictions an important cornerstone of the judicial system. If little or no human element would be left within the judicial proceedings, how would this affect the nature of the proceedings and could this have detrimental effects on what we consider a *fair* trial to be comprised of? The concept of accountability may also change the way we think about adjudication since an artificial judge could not possibly be held accountable in the same ways a human could.

The aforementioned questions and issues will have to be answered by developers and legislators before an AI revolution can take place within the judiciary. This thesis attempts to shed light on some of the key points of replacing judges with artificial intelligence.

1.2 Purpose

The general purpose of this thesis is to analyse and discuss the use of artificial intelligence within the judiciary to evaluate its suitability in courts. The focus being the collision with human rights, particularly the aspect of the right to a fair trial, in a situation where the adjudicator is an artificial intelligence instead of a human being. Therefore, AI is analyzed in relation to fundamental human rights, particularly narrowed down to European law in the form of Article 6 ECHR.

The following questions are answered in order to accomplish the above mentioned purpose:

- Does adjudication by artificial intelligence undermine the right to a fair trial in article 6 ECHR?
- Can "artificial judges" live up to the requirements of a fair trial as set forth in Article 6 ECHR?
- How would "artificial adjudication" affect the public's view of justice?
- Are AI judges suitable from a perspective of the appearance of fairness?

1.3 Delimitations

Because the field of artificial intelligence and law spans endless questions and problems certain delimitations have been made to this thesis. This thesis does not cover questions related to data privacy and questions related to the General Data Protection Regulation of the EU. Nor does this thesis cover IT-security related issues in depth. To keep the thesis somewhat concrete and manageable for jurists the purely technical aspects of artificial intelligence is kept to a minimum or at the very least a low threshold of the technical intricacies is meant to be upheld. Additionally no aspects of copyright or patent law concerning AI is covered.

Regarding Article 6 certain delimitations are also made and therefore only the aspects of access to justice, a fair hearing, impartiality and independence and the definition of tribunal or court is discussed. These delimitations have been chosen since these aspects are of particular interest when discussing artificial intelligence as a judge whereas certain other aspects of the article such as immunities and legal aid as subsets of the access to justice doctrine are less relevant.

This thesis does not cover the autonomous meaning of civil rights or obligations, neither does it cover the meaning of a criminal charge. This is because these autonomous prerequisites are not essential in order to determine what institutional requirements are put on the courts. Otherwise the meaning of civil rights, obligations or criminal charge is naturally an important aspect of Article 6 and could not otherwise be disregarded.

1.4 Method and Sources

The methodology used in this thesis is what is known as a "legal dogmatic method" in order to determine what the established law regarding a fair trial is. The method dogmatically attempts to determine the established law by looking at relevant sources of law, such as legalisation, case law and legal doctrinal literature, from within a chosen legal system.[1] A legal dogmatic method in this thesis means to use the relevant sources of law to determine what the established law is in relation to Article 6 ECHR. As such this thesis is primarily a study of the case law from ECtHR as case law is the primary source material for interpreting the application and scope of the convention. When the established law of Article 6 is determined it is applied to different aspects, some technical and some theoretical, of artificial intelligence in the role as a judge to evaluate its adjudicatory suitability in terms of fairness and efficiency. This analytical application of the law serves as a solid foundation to identify legal issues as well as possibilities of artificial intelligence. As such this thesis does not apply a strictly legal dogmatic method throughout.

As stated the source material used is primarily case law although certain relevant legal doctrinal literature concerning the ECHR at large have been used. The cases have been selected based on their relevancy to Article 6 with certain regard to their authority, i.e. if they are cases from the ECtHR grand chamber they have been deemed authoritative. The doctrinal literature provide important summarization of relevant cases that serve as a starting point when looking at the different criteria of Article 6. Furthermore the literature may also provide good points for later analysis.

Regarding the source material for artificial intelligence certain considerations has been made. The field is mainly within the domain of computer science and as such the most

[1] Kleineman, pp. 21-29.

well-cited material that is found is rather technical in its nature. The sources have been chosen by looking at their citations and trying, from an outside perspective without prior in-depth knowledge, to determine their relevancy and authority in the field.

1.5 Disposition

This thesis is divided into multiple parts for structural purposes. The first part introduces artificial intelligence and lays the foundation for understanding its fundamentals such as a broad definition, the advantages of AI, machine learning, different biases, deep neural networks and finally a theoretical model of adjudicative AI that is presented to facilitate later discussion.

The second part is focused on the right to a fair trial and the prerequisites that needs to be met for a court in order to comply with Article 6 ECHR. The chapter delves into the case law of the ECtHR to try to determine what the requirements may be.

Thereafter, the collision between certain aspects of AI and the aforementioned human rights are discussed in the third chapter. This is done specifically to highlight the difficulties that stems from how AI works and the demands we put on the judiciary.

The fourth part broadens the discussion and focuses on the effect that AI has on the perceived fairness, transparency and if there is a human aspect of adjudication that may be lost by implementing AI judges.

Finally the concluding chapter covers the summarized thoughts regarding the use of AI adjudicators.

2 Artificial Intelligence

2.1 General

This chapter explores the basics of artificial intelligence that is needed to discuss its advantages and disadvantages as well as conceptualizing it within the judiciary later on.

Artificial intelligence is predicted to have a profound impact on society as a whole where the possibilities are held back only by our imagination, or possibly by our lack of imagination and knowledge.[2] However, as of right now the technology is not at the level where we have *thinking* computers that can solve any abstract problem we throw at it.[3] Additionally, the technological advances have been theorized to be exponential; a phenomenon that is known as Moore's Law.[4] Initially Moore's Law was sprung from the development pace of the number of transistors in integrated circuits which roughly doubled every other year. Since then Moore's Law has been argued to apply to technology in a much wider sense than hardware thus encompassing for example software as well.[5] As such we are moving closer and closer to a reality where algorithms and AI will become inseparable from human lives even more so than today. In parallel with this development the efficiency and sophistication of AI will continue to advance further. To some people the goal, or the consequence if one is so inclined, of this exponential development is a so called technological singularity where AI eventually supersedes human intelligence on a general level.[6]

Prominent experts in the use of artificial intelligence have voiced concern over the implications of an emerging artificial intelligence in an open letter and article. The open letter has been signed by over 8,000 signatories from different fields all over the world.[7] The conclusion of the article is clear; more research needs to be done in order to ensure that AI is secure, controllable and its uses are morally justifiable. We need to make AI beneficial while avoiding many pitfalls before we have taken on more than we can

---

[2] Russell & Norvig, pp. 1051-1052.
[3] Ibid, pp. 28-29.
[4] Collins, p. 100.
[5] Ibid, p. 101.
[6] Kurzweil, p. 35.
[7] See Russell, Dietterich et al, *Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter.*

manage.[8] Criticism of this dystopian view take different forms and some oppose the technological singularity as a concept while others argue that the limitations of software will hinder artificial intelligence to ever become as complex as a human intelligence.[9] No matter which view one sides with further research and discussion is important and will be much needed in the near future.

The following chapters investigates how artificial intelligence is defined, what artificial intelligence really is and how the technology behind it makes it work.

2.2 Defining Artificial Intelligence

There is no single definition of what artificial intelligence really is. The Council of Europe has defined AI as:

> "*A set of sciences, theories and techniques dedicated to improving the ability of machines to do things requiring intelligence. An AI system is a machine-based system that makes recommendations, predictions or decisions for a given set of objectives.*"[10]

Another proposed way to broadly define AI is that:

> "*Artificial Intelligence involves using methods based on the intelligent behaviour of humans and other animals to solve complex problems*"[11]

Neither of the definitions holds authority over the other. Regardless of the semantic or technical definition when discussing artificial intelligence the imagination can usually lead to a so-called *artificial general intelligence* or *strong artificial intelligence* that usually takes the form of an ominous super computer, murderous robot or machine;

---

[8] See Russell, Dewey & Tegmark, *Research Priorities for Robust and Beneficial Artificial Intelligence*, p. 112.
[9] See e.g. Kurzweil, pp. 309-310.
[10] CEPEJ, *Ethical Charter*, p. 5.
[11] Coppin, p. 4, for more definitions and some differences between definitions see also Russell & Norvig pp. 1-2.

something that also has been popularized in TV and film.[12] This type of artificial intelligence is not restricted to operating in a specific field or with a specific task but can abstractly apply and adapt its intelligence like a human being, essentially being conscious.[13] This kind of artificial intelligence is not yet a reality and for the purpose of this thesis such an AI is not of primary concern and will therefore not be discussed in more detail.

Instead, when discussing artificial intelligence from a more practical standpoint, one is usually more interested in an AI that operates within a specific field or with a specific task. For example an AI that autonomously drives a car. This is sometimes called a *narrow artificial intelligence* or *weak artificial intelligence*. The narrow AI can apply and adapt its intelligence only to the specific task or narrow field it was designed to work within and can oftentimes exceed human capabilities at this task.[14] This is an important distinction as it restricts the AI from operating in any field to only being able to operate with a defined set of tasks. When presented with a task outside of its field it will ultimately fail. This however, does not preclude the possibility that the narrow task(s) the AI is instructed to perform is of a very complex nature that would be excessively laborious for a human to complete within a reasonable time.

In summary there is no set definition of AI that fits all purposes of artificial intelligence, nor is there a consensus what 'intelligence' is or how to define it for an algorithm. As there are significant differences between a strong AI and a weak AI there are also difficulties to cover both kinds in one definition. As a result the definitions that are put forth, as seen above, are often much too vague to discuss concretely. In chapter 2.6 a proposed model of a narrow AI will be presented that is used in order to have a meaningful discussion further on.

2.3 The Fundamentals of Artificial Intelligence

2.3.1 Machine Learning

In order to lay a foundation for better understanding and further discussion some key elements of artificial intelligence are explained below.

---

[12] See for example the popular Matrix trilogy or Terminator franchise.
[13] See Coppin, p. 5, and Russell & Norvig, pp. 1051-1052.
[14] See Kurzweil, pp. 213-214 and Coppin, p. 5.

The basic technology that needs to be understood which also can be called the core of artificial intelligence is the learning process. This technology is known as machine learning.[15] There are three main categories of machine learning: supervised, semi-supervised and unsupervised.[16] A supervised learning process is where a human helps the AI through every step of the way. A semi-supervised process is where a human helps the AI initially but after that it is on its own. Unsupervised learning on the other hand is where the AI is left on its own from the start.

The common denominator between the different learning processes is that any machine learning process requires a significant amount of data to learn from. What this data *is* depends on the purpose of the AI. If the purpose is, for example, facial recognition in images then the data might be comprised of millions of images that the AI will analyse. One of the most famous databases used to develop image recognition AIs, ImageNet, has over 15 million images in its dataset.[17] The training data is important in order for the AI to develop and recognise patterns. As mentioned previously a supervised AI requires help from a human. It needs help in order to categorise and interpret the data that initially has no meaning by itself to the AI. This means that before the AI can recognize a face in an image or video, the human may have to point out to the AI where the face is on the image. From there on the AI will learn after enough analysed images to develop patterns in order to 'see' where the faces are by itself. Unsupervised learning on the other hand requires no initial help from a human to categorize or sort the training data but it will itself develop the necessary categorization.[18]

Semi-supervised learning is a hybrid between supervised and unsupervised learning and has been proposed by some to be close to human learning as the AI is given some categorization and labels but is then left on its own to categorize the rest like when humans are children.[19]

What becomes evident from the above description of machine learning is that the data the AI is trained on is what constitutes its knowledge. The *quality* of the data is equally

---

[15] Coppin, pp. 267-268.
[16] Ibid, pp. 284-285.
[17] See Krizhevsky et al, *ImageNet Classification with Deep Convolutional Neural Networks*.
[18] Coppin, pp. 285-286.
[19] See McCarty, p. 66.

important as the *amount* of data when training an AI. It is also crucial how well the data can be labelled or categorized. If the data is flawed as in mislabelled or insufficient[20], or in the case of adjudication one can imagine training the AI on erroneous cases, then this will have adverse effect on the AIs ability to reach a correct and accurate outcome. As stated in the previous chapter, a narrow AI in its most simple form cannot in a meaningful way process data that it was not instructed, or trained, to process. For example, if the AI is a supervised intelligence trained to recognise faces it cannot be asked to drive a car because there is no framework in place and no data to support its decisions. In the next chapter we will move further up the technological ladder and investigate the technology that has advanced AI to the next level and taken AI into our modern day lives.

2.3.2 Neural Networks and Deep Learning

In contrast to the early days of AI where machine learning algorithms were computationally very expensive[21] we now have phones that match or outperform computers from just 10 years ago. This technological advance has made computationally expensive tasks much more available than they were in the early 2000s. Because of this development AI is not as restricted by physical hardware limitations anymore and the neural networks that were previously deemed computationally too expensive can now be run on consumer hardware.[22] The technology that makes the current level of AI possible is called a neural network which itself is a subset of the machine learning introduced in the previous chapter. However, the technology is not as recently developed as one may think but was conceptualized as early as the 1950s.[23] The reason that the technology has seen a sudden surge in use during the last ten years is a combination of the capabilities of recent hardware, the development in the theory behind AI as well as a reawakened belief that neural networks may be the future of AI.[24] A neural network, or when discussing AI one usually refers to it as an *artificial neural network* (ANN)[25], is a network of virtual nodes that run as a program on a computer. Between the nodes are links that connect the

---

[20] Cortes, C. et al, pp. 57-58.
[21] Meaning that they needed a lot of computing power and was a very expensive field for research and development.
[22] E.g. almost every smartphone can run games like chess that have a simple AI built in.
[23] Russell & Norvig, pp. 16-19.
[24] For example Coppin, pp. 8-10 and Goodfellow et al, p. 24.
[25] Goodfellow et al, p. 13.

nodes to each other. The links are simply mathematical functions that transform values from the previous node and feed into the next one based on weights and biases that are initially determined by the programmer in the algorithm. The numerical value of a link by itself or in a series of nodes, called layers, is what ultimately determines the output of the network.[26] To go back to the previous example with the facial recognition; when an image is presented to the AI it tries to determine where the face is by looking for patterns. The network then adjusts itself, which means that the nodes in the network are given new weights and the connections between the nodes *learn* new more complex concepts about the images. How the network adjusts itself after being presented with a new image is determined by the programming of the algorithm.

The technology can also be illustrated by the architecture of the human brain. Each virtual node is a representation of a neuron in the brain. The vast amount of connections between the neurons is what gives the brain its complex functionality.[27] One can therefore think of the neural network as the brain of the artificial intelligence. When the brain learns it develop new connections or alters already existing connections between neurons. When the AI learns it adjusts existing weights and biases by altering the mathematical function between the nodes, i.e. the link.[28]

However, a simple neural network is nowhere near the complexity of the human brain. The closest we have come is by the technology called deep learning, or a deep neural network as it is a subset of neural networks. Deep learning is just a further development, or subset, of machine learning and neural networks that has taken artificial intelligence to another level in a revolutionary way.[29] Together with the excessive amount of data[30] that is available training neural networks have never been easier. What constitutes 'deep learning' when talking about artificial intelligence is, very simplified, the amount of layers of nodes in between the input and output of the network. As mentioned briefly earlier, the nodes can be imagined as structure in layers with many connections in between

---

[26] Collins, p. 103.
[27] Coppin, pp. 292-293.
[28] Ibid, pp. 293-294.
[29] Samek & Müller, p. 6.
[30] See Smith, Reuters 2013. In 2013 Facebook uploaded an estimated 350 million pictures each day. In 2020 when this thesis was written that number has most likely more than doubled. See also infra chapter 2.3.4.

each layer and each node. On one side there is an input layer that receives data and feeds into the next layer or back into itself before reaching the output layer which yields a result.[31] A very simple neural network may have two or three layers of nodes that connect to each other while a deep neural network may have any number of layers between the input and output layer, known as the *hidden layer(s)* of nodes.[32] The illustrations below are meant to show how a simple neural network and a deep neural network may be structured. Each line is a representation of a mathematical function, i.e a weighted link, between the nodes.

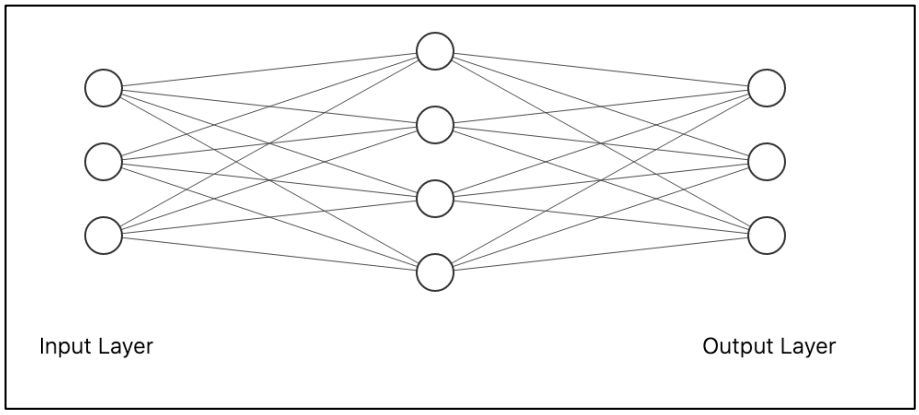

Figure 1. A visualization of a simple neural network .

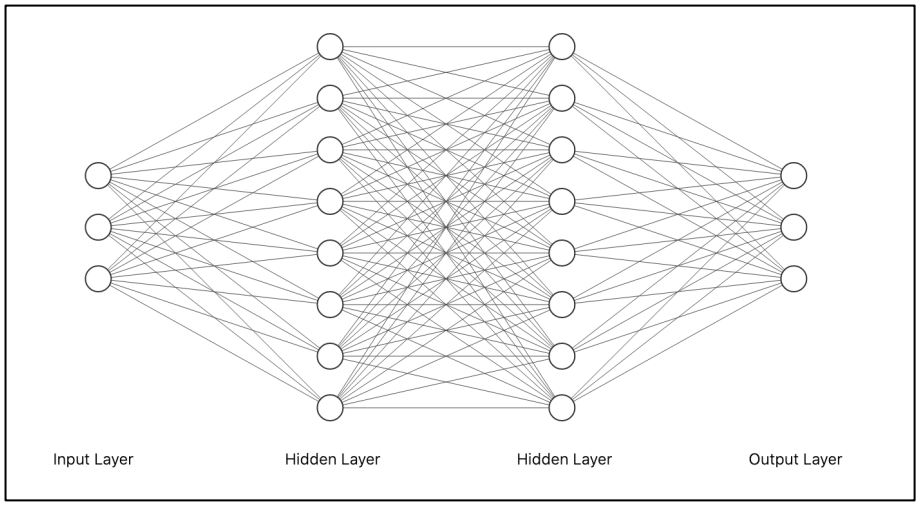

Figure 2. A visualization of a deep neural network with hidden layers between the input and output.

[31] Coppin, pp. 300-302.
[32] Skansi, p. 79.

An important aspect of deep neural networks is that due to the vast amount of connections between the layers of nodes what happens between the input layer and output layer is impossible to know for certain from an outside perspective. The network itself lacks transparency between input and output which has not been a problem in early adoptions of neural networks for consumer use since no expectation of transparency existed. Since the network adjusts itself after each iteration of learning reverse engineering[33] the deep neural network would essentially mean undoing every step of trained data in order to map out how the numerical values and connections change between the nodes which may prove very difficult. However, work is being done in order to create interpretable and explainable deep learning algorithms.[34] Furthermore, as artificial intelligence based on neural networks become more prevalent in, for example, driverless cars the need for transparency in *how* decisions are made is becoming increasingly important in order to put public trust in its usage.[35] Additionally, when looking at autonomous decision making at large it becomes evident that transparency is of higher importance now than ever before and is even required to certain extent when involving for example personal data processing within the EU.[36] This development means that when developing AI judges transparency will be a key issue that need to be addressed.

2.3.3 Bias in Machine Learning

As stated in the previous chapter one intrinsic property of AI is that the data it is trained on is what constitutes its knowledge. At first glance this might not seem peculiar, such is also the knowledge of every human, is it not? However, since the training data may be biased in terms of gender, race, sexuality or other aspects the output or decisions of the AI will also reflect such bias which can have adverse effects when neutrality is of importance.

A distinction between three different varieties of AI bias can be made; input bias, training bias and programming bias. These three different biases can affect the output of an AI in a negative manner. Input bias stems from the source data being incomplete,

---

[33] Meaning that the network is deconstructed in order to see what makes it work and how it works.
[34] Montavon et al , p. 193.
[35] Ibid, p. 7.
[36] See for example the General Data Protection Regulation EU(2016/679), Article 13(2)(f) and 14(2)(g).

lacking certain variables or not being representative. Training bias comes from either the categorization of data or the determination whether the AIs output matches the desired result. Programming bias stems either from the programmer of the AI carrying bias when designing the algorithm or if the algorithm can be modified or re-programmed after interaction or contact with new data and therefore is impressionable to bias.[37]

Bias in AI can cause unwanted or unforeseen effects such as de facto discrimination based on gender, race or any characteristic. For example, an AI employed to help with screening new candidates for recruitment proved after some time to systematically favour male candidates, without being instructed to do so, just by looking at the candidates' resumes.[38] Another example concerns Google's ad delivery network showing results that have racist tendencies based on searching for names that are 'black-sounding'.[39] In a more recent case Facebooks advertisement algorithm has also shown these problematic tendencies based on race and gender when serving ads to its users.[40] Another Google service that tries to identify objects in images has been recently criticized to connect dark skin tone with labels such as 'gun' and 'firearm' while not making the same connections for lighter skin tones when the object in question, which was a handheld thermometer, is in fact the same.[41] What kind of bias or biases are affecting the outcome here is hard to guess from the outside but it is not hard to conclude that a biased AI can have negative consequences for minorities or groups that already are subject to certain systematic discrimination.

The risk of bias is well known in the field of artificial intelligence and efforts are made to create viable ways of working around and preventing bias from corrupting AI.[42] It is especially important to avoid systematic bias in institutions that require the public's trust such as the law enforcement and judiciary. If the public cannot trust the institution to uphold impartiality then rule of law will inevitably be undermined.

---

[37] McCarty, p. 96.
[38] See Dastin, Reuters 2018.
[39] *Racism is Poisoning Online Ad Delivery, Says Harvard Professor* , MIT Technology Review 2013.
[40] Hao, MIT Technology Review 2019.
[41] Kayser-Bril, AlgorithmWatch 2020.
[42] See for example Bolukbasi et al, pp. 1-9.

2.3.4 The Importance of Data and its Impact on Accuracy

Since the area of algorithms, AI and machine learning essentially relies purely on mathematics it is difficult to explain many concepts without getting too technical. So far we have covered the technology that lets AI arrive at conclusions or make predictions. Some additional information regarding the importance of the data must also be said.

The current state of AI development has a causal relationship with the emergence of the colossal amount of data that is available nowadays as mentioned earlier. And while the AI models from as far back as the 1980s are close to those we have today, the kind of and the amount of data that we can train the neural networks on has changed drastically.[43] According to Goodfellow et al in 2016 in order to successfully train a neural network to achieve human or superhuman capabilities for a specific task required $10x10^6$ labelled examples in a dataset. With time these large datasets have become more and more accessible and will, in the not so distant future, become less of a hurdle for AI researchers to overcome as more and more data is produced each year.[44]

A significant problem is also the accuracy of the neural network. How can we know that the decision or prediction is correct or accurate? One way to describe this accuracy is that "the accuracy is the proportion of examples for which the model produces the correct output"[45]. The accuracy is in turn dependent on the datasets which the neural network is trained and subsequently tested on. If the dataset the AI is trained on is sufficiently large and representative there will likely be more accurate outcomes. Take for example an AI trained mostly on cases of homicide and then tasking it to determine a case of a disputed parking ticket. The accuracy will most likely be low as the AI is much less familiar with parking than homicides. On the other hand, if the AI has been trained on too many cases including those leading to wrongful convictions then the outcome may lead to what is known as overfitting.[46] The AI will then have too wide of a scope of what data is relevant and will include wrong data in its weights, leading to lesser accuracy. The opposite problem, where the AI is trained on a too limited dataset, may instead lead to underfitting which in turn also results in lesser accuracy. This means that the training data

---

[43] Goodfellow et al, p. 19.
[44] Ibid, p. 21.
[45] Ibid, p. 103.
[46] Ibid, pp. 110-113.

has to be sufficiently representative while eliminating erroneous data and not removing valuable data at the same time.

2.4 Applications of Artificial Intelligence

As much as negative consequences of biased or inaccurate AI exist there are still clear advantages of developing artificial intelligence and many recent events or breakthroughs can testify that the field of artificial intelligence is getting more and more impactful. For example, in 2011 the IBM super computer "Watson" for the first time beat humans in the game Jeopardy, which is a trivia game spanning general knowledge.[47] In 2017 the best Go player in the world was beat by Googles artificial intelligence AlphaGo. Just months later an updated version called AlphaGo Zero could learn, through unsupervised training, to beat its predecessors.[48] Also in 2017 an artificial intelligence in the medical field was able to reach dermatologist level of detecting and classifying skin cancer.[49] The applications of AI in the medical field can also be exemplified when researchers with the help of AI developed a new kind of antibiotics that has effect on multi-resistant bacteria.[50] Another prominent and topical example is the pharmaceutical company AstraZeneca that uses AI to accelerate the research and development of new medicines.[51] This is advanced even further by a group of pharmaceutical companies utilizing so called 'blockchain technology'[52] to collaborate and share data between each other, without breaking anti-trust regulations, that could be used to train AI.[53]

Within the legal field AI is used as a tool for analysis of contracts and documents in general, however the adoption of AI tools is still remarkably slow.[54] Within the judiciary in some states of the United States of America a system called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) helps judges to analyse the

---

[47] Markoff, the New York Times 2011.
[48] See DeepMind, *AlphaGo*.
[49] See Esteva et al, pp. 115-118.
[50] Stokes et al, *A Deep Learning Approach to Antibiotic Discovery.*
[51] See for example the press release from AstraZeneca 2019.
[52] Blockchain is a technology that is beyond the scope for this thesis. One can simply imagine it as an encrypted set of data that can be shared in a secure yet accessible and verifiable fashion.
[53] See Kuchler, Financial Times 2019.
[54] See Toews, Forbes 2019.

risk of recidivism.[55] The system has however been criticised for being biased, obscure and in reality no better than humans in general at determining recidivism.[56] Another AI tool known as Clearview focuses on facial recognition. The tool can take an image of any individual's face and scan the open internet for all occurrences of that face in anything from videos to images. The tool has been used by law enforcement agencies to both track criminals and unidentified victims. Critics say this is a serious violation of privacy since there is no possibility of opting out.[57] In spite of this, the emergence of AI tools within the judiciary, law enforcement and legal field at large can most likely be expected to continue. Again it is worth emphasizing that in present time it is not feasible to replace judges with AI equivalents. A study[58] from the CoE on algorithms and human rights in 2016 concluded that the correct prediction rate of predictive algorithms was at 79% which is far too low if the same would apply for decision making.[59] Another interesting application is the possible use of AI within arbitration and dispute resolution. Arbitration services could use AI to facilitate much more efficient proceedings while also keeping the costs down for the parties as the cost is a known drawback of using arbitration. The complexity of arbitration may however render AI fruitless according to some critics.[60]

In summary there are certainly viable applications of AI within the legal field but the line between discrimination and impartiality, privacy infringements and effectiveness of law enforcement, and last but not least the respect for due process and the rule of law must be kept in mind when implementing and using the tools. AI is particularly useful where a large amount of data needs to be analysed or where traditional methods of analysis would be much too slow and cumbersome.[61] It would for example be impossible to task a human with analysing millions of images within a reasonable time in order to find and identify an unidentified victim or criminal in a video or picture. The aforementioned examples are just a select few but show that there are advantages and possible uses of AI that society can potentially benefit from.

---

[55] Weller, pp. 28-29.
[56] Yong, The Atlantic 2018 and Larson et al, ProPublica 2016.
[57] Alba, the New York Times 2020.
[58] MSI-NET, *Algorithms and Human Rights,* pp. 11-12.
[59] Note that the referenced study concerned predictive algorithms and not decision-making algorithms.
[60] See Scherer, pp. 509-512, for her skepticism concerning the direct substitution of human arbitrators with AI arbitrators.
[61] Coppin, pp. 23-24.

2.5 Summarizing Key Points of Artificial Intelligence

To summarize this chapter before moving on to the theoretical model one can first keep in mind that there is no consensus of a single definition of AI. Artificial intelligence is today mostly built on neural networks that rely on different kinds of machine learning which enable the AI to find patterns in data. Machine learning is the overall technique behind how the AI learns whereas deep neural networks is a further development of translating the learning into a prediction, decision or result. Due to the nature of these technologies there are inherent issues regarding bias and transparency that must be taken into account when developing and using AI. Both issues need to be addressed in order for AI to gain public trust.

Additionally some problems with training neural networks comes from the data itself and quality of the data. When training a neural network it is important to keep in mind that the data needs to be accurate and representative in order to minimize errors. It is both problematic to include and exclude data when training the AI. Including too much data means lesser accuracy and excluding data opens up for bias or subpar pattern recognition of the AI.

Lastly as AI is a powerful tool when developed we must consider which guidelines to use that will keep the AI from undermining privacy and fundamental human rights. These guidelines must be adopted before implementing AI in ways that could be harmful for society.

2.6 A Theoretical Model of Adjudicative Artificial Intelligence

In order to discuss AI adjudicators effectively there needs to be a theoretical model as a foundation for discussion. Discussing AI as adjudicators in this thesis therefore assume the following conditions to be true.

The AI judge is…

i) …based on known technology and trained on case law;

ii) …able to interpret both oral and written information and follow procedural rules as human judges;

iii) …a "black box" where the processing that happens between input and output is unknown to outside viewers;

iv) …impossible to distinguish from a human in terms of constructing convincing legal reasoning in a written judgement.[62]

[62] This should by no means be understood as the AI arriving at the *correct* conclusion at all times. Instead it is meant to shift the focus to the adjudicatory function from the purely material outcome of adjudication.

3 A Fair Trial – A Fundamental Right

3.1 General

Moving on from the technical details of AI to the field of human rights. This chapter looks closer on the notion of a fair trial and the subsequent right to a fair trial according to the case law of the European Court of Human Rights. Initially it can be said that the delicate line between the sovereignty of the national courts and the jurisdiction of the ECtHR is rarely crossed by the Strasbourg Court as it follows the 'fourth instance' doctrine.[63] This means that the Strasbourg Court does not involve itself in determining errors of fact or national law unless it means that the rights and freedoms in ECHR is affected.[64] The Strasbourg Court is primarily concerned with how the implementation and application of national law correlates with the rights and freedoms that ECHR is set to protect.

Two general, but nonetheless particularly important, aspects of rights are that they need to be effective and practical and not illusory and theoretical, otherwise they merely serve a symbolic purpose and cannot be invoked in a meaningful way.[65] The right to a fair trial in Article 6 ECHR may at first glance seem like such a symbolic right due to its broad formulation. However, the ECtHR has via its extensive case law on Article 6 fleshed out a multitude of implicit rights and principles that are by no means only theoretical, such as the access to court doctrine. These implicit rights and principles form a safety-net that serves as a procedural and general protection for human rights. As the right to a fair trial is imperative in order to be able to secure any other right of the convention Article 6 cannot tolerate restrictions or impairment. The ECtHR has therefore stated that the interpretation of Article 6 should not be made restrictively since the right to a fair administration of justice holds such a prominent place in democratic society.[66] Article 6 is also said to be the single most invoked provision of the ECHR before the ECtHR which paves the way for a large volume of cases.[67]

---

[63] Harris, O'Boyle et al, p. 374.
[64] Garcia Ruiz v. Spain para. 28.
[65] E.g. Airey v. Ireland para. 24.
[66] Delcourt v. Belgium para. 25.
[67] Harris, O'Boyle et al, p. 374.

Article 6 guarantees the right to a fair trial in both civil and criminal matters. The fundamental requirement for the Article to be applicable is that the case determines a dispute concerning civil rights, civil obligations or a criminal charge.[68] The precise definition of what constitutes a civil right or obligation, as well as criminal charge have not been exhaustively listed by the ECtHR but have been the object of several disputes.[69] The definitions also have autonomous meaning that the ECtHR can expand or restrict, as such it is not up to the Contracting States to determine the meaning of what constitutes a civil right, obligation or criminal charge. In spite of the ECtHR's ability to decide the meaning of civil rights or obligations the Strasbourg Court cannot itself create rights that have no prior foundation in domestic law.[70] The Article itself is clearly divided into three parts. Article 6(1) concerns both civil and criminal matters while 6(2) and 6(3) concerns criminal matters. Nonetheless, the ECtHR has indicated that certain principles that stem from Articles 6(2) and 6(3) are also applicable by analogy to civil proceedings that fall under Article 6(1).[71] This thesis explores both the civil and criminal parts of Article 6 in a broad relation to artificial intelligence and as such no strict division of the Article is made.

The following chapters will cover important areas of Article 6 that have been developed via extensive case law from ECtHR such as: the right of access to court, the requirement of impartiality and independence of the court, what constitutes a fair hearing and what defines a court or tribunal. These areas are examined in order to determine what constitutes a fair trial in relation to the formal requirements of the court and if an AI judge can live up to the standards?

3.2 The Right of Access to Court

3.2.1 Access to Court

Since the right of access to court must be upheld an AI judge would have to comply with the rights. We must therefore start by looking at its prerequisites. The right of access to court is not explicitly stated in Article 6 ECHR but has grown through case law and is

[68] Schabas, p. 272.
[69] Ibid, p. 272.
[70] Ibid, p. 273.
[71] E.g. Albert and Le Compte v. Belgium, paras. 32-33.

viewed as an integral part of the article. The right of access to court means that one has the right to institute proceedings before a court of law.

The first case where the ECtHR established the right of access to court was *Golder v. the United Kingdom*. The ECtHR states that the right of access to court follows implicitly from Article 6(1) since it is not explicitly stated, as mentioned earlier. However, it is an inherent element of the Article without interpreting the article in an extensive way. The Strasbourg Court also explains that the right to a fair trial in Article 6 is built from many underlying principles or rights, which as a whole constitutes what is known as a 'fair trial'. [72] As such there may be implicit rights that are required for a fair trial to exist. The right of access to court is one of these rights. The Strasbourg Court has summarized that Article 6(1) secures a right for everyone to have a claim concerning civil rights and obligations brought before a court or tribunal. A further development of the access to court doctrine is made in the case *Deweer v. Belgium* which concludes that the right of access to court also applies both to civil and criminal cases.[73] No general distinction between them should therefore be made in principle.

The right of access to court is also said to be subject to certain limitations. In the case of *Ashingdane v. the United Kingdom* the Court has stated that the right of access to court is subject to limitations decided by the Contracting States. However, such limitations may not modify the core of the right and must always be proportional. In the aforementioned Ashingdane case the ECtHR states that "*any limitations must not restrict or reduce the access left to the individual in such a way or to such an extent that the very essence of the right is impaired*".[74] This principle has later been repeated in several cases.[75] As such it is up to the Contracting States to protect the right of access to court, while still maintaining certain practical aspects of the way courts operate such as time limits of appeals or court fees.

The right of access to court is also an integral part of the individual's access to justice. If there is no way of securing an individual's rights before a court then justice can hardly be said to be administered. However, the fact that one cannot institute a case before a

[72] See para. 28 and 36.
[73] See para. 49.
[74] See para. 57.
[75] See e.g. Tinnelly & Sons Ltd and Others and McElduff and Others v. the United Kingdom.

traditional court does not necessarily mean that the right of access to court, or the access to justice, is impaired which will be explored in the next chapter.

3.2.2 What Constitutes a 'Court' Or 'Tribunal'?

In order to further outline what the right of access to court really means the definition of the court itself must be determined. The concept of 'court' or 'tribunal' is autonomous within the meaning of the convention.[76] As such the national concepts of what constitutes a proper court may not necessarily be the same as ECHR and is of little importance in the eyes of the ECtHR. An authority or official body does not for example need to be strictly classified as a court in order for Article 6 to still be applicable, nor does the presiding individual need to be titled judge. Instead the ECtHR has stated that it is of greater importance to look at the judicial function of the body.[77] The ECtHR has also stated that if the authority determines matters within its competence on the basis of the rule of law and holds proceedings in a prescribed manner then the authority falls under the substantive definition of 'tribunal' or 'court' as it serves a judicial function.[78] As such there are many different authorities and bodies that can constitute a tribunal or court in the sense of Article 6. The right of access to court does therefore not necessarily mean that one is entitled to proceedings before a court in the traditional sense with a judge or jury. Other official bodies can fulfil the requirements such as disciplinary boards or committees.

Another important factor in determining the characteristic of a court or tribunal is that it has the power to determine disputes. It is not enough to be able issue non-binding statements or opinions.[79] One of the true characteristic of a court is therefore if it has the power to issue a binding judgement that determines the outcome of a civil dispute or a criminal matter. In other words the court must have the ability to issue enforceable judgements that cannot be quashed or retried other than by appellate courts.

In summary a court in the autonomous meaning of the convention must serve a judicial function, determine matters within its competence on the basis of the rule of law and hold proceedings in a prescribed manner for it to count as a court or tribunal.

---

[76] Sramek v. Austria para. 36.
[77] Belilos v. Switzerland para. 64.
[78] Ibid para. 64 and Sramek v. Austria para. 36.
[79] Benthem v. the Netherlands, para. 40

3.3 A Court 'Established by Law'

The statement that a court or tribunal must be 'established by law' in Article 6(1) ECHR is intended to indicate that the judiciary power cannot be arbitrarily decided by the executive power. Something that reflects the principle of the rule of law.[80] The courts must have a legal basis, both in terms of organization and function.[81] The ECtHR has also stated that the composition of the bench cannot be left to the arbitrary discretion of the executive power or left to the judiciary itself to decide but also has to have a legal basis.[82] Even the appointment and renewal of a judge's term of office cannot be left to the executive or left to be regulated by the internal practice of the judiciary.[83] The process must also follow the principles of independence and impartiality.

The definition of 'law' in this situation does not only include legislation concerning the establishment, competence and organization of the tribunal but also "any other provision of domestic law which, if breached, would render the participation of one or more judges in the examination of a case irregular"[84]. A decision to appoint, renew or dismiss a judge that is deemed irregular in relation to other decisions of similar nature would therefore risk rendering the court unlawful as in not established by law. If there are arbitrary decisions concerning the participation of one or more judges that breaches national law it can be seen as irregular and in breach of article 6(1) ECHR. A tribunal that is without legitimate jurisdiction or oversteps its limited jurisdiction may also risk not being a tribunal 'established by law'.[85] However, in general the ECtHR does not question the interpretation of national law unless there has been a flagrant breach of the legislation.[86] What is considered such a flagrant breach has to be decided on a case-by-case basis by the ECtHR.

As the court has to be regulated there needs to be legislative measures made in order to implement AI judges. It may be problematic to irregularly appoint AI judges without proper legislative action. As such there is no general hindrance to create for example

[80] Harris, O'Boyle et al, p. 458.
[81] Barkhuysen et al, p. 611.
[82] Sokurenko and Strygun v. Ukraine, para. 24.
[83] Oleksandr Volkov v. Ukraine, paras. 151-156.
[84] DMD Group, A.S., v. Slovakia, para. 59.
[85] Sokurenko and Strygun v. Ukraine, paras. 27-28.
[86] Pasquini v. San Marino, paras. 104-109.

special AI courts with an AI that serves as adjudicator as long as the court is properly regulated and established in national law. Naturally the AI would need to follow the other prerequisites of a fair trial in Article 6 but this would not affect the ability to *establish* the AI court itself.

3.4 Impartiality and Independence

3.4.1 General

A truly essential part of the rule of law is that the judges or laymen that take part in adjudication are impartial and independent, otherwise the decisions will be questioned and the public may lose faith in the judicial system as a whole. Article 6(1) ECHR states that an individual has the right to be heard by "*an independent and impartial tribunal established by law*". The ECtHR has established prerequisites when examining the impartiality and independence of a bench or individual judge. Aside from these prerequisites, the ECtHR has also stated that a tribunal comprised of lay men, lay judges, experts or members of interested bodies does not by itself constitute a case of bias, see for example the cases *Le Compte, Van Leuven and De Meyere v. Belgium* and *Pabla Ky v. Finland*.[87] Furthermore, the same principles that apply to professional judges also apply to any other lay judge or lay member of the bench, see *Langborger v. Sweden* and *Cooper v. the United Kingdom*.[88] As such, the criteria for impartiality and independence to be met are not affected by the background or profession of the members of the tribunal but are equally applied for each member. Instead other criteria has to be met in order for the court to be questioned in terms of its impartiality and independence.

Additionally, as impartiality and independence are more or less connected terms they may require a joint examination which is commonly done by the ECtHR.[89] See for example the case of *Ramos Nunes de Carvalho e Sá v. Portugal* in which the court argues that a breach of one may result in breach of the other.[90] In the chapters below the terms will be examined independently according to the principles laid out by the ECtHR in order to identify the requirements for each situation.

---

[87] See paras. 57-58 and para. 32 respectively.
[88] Paras. 34-35 and 123 respectively.
[89] Harris, O'Boyle et al, p. 446.
[90] See especially paras. 150-156.

3.4.2 The Impartiality of the Court

In order to assess whether a tribunal is deemed to be impartial or not the ECtHR has put forth two tests. There is however no definite line between the two tests and as such they may overlap.[91]

i) the subjective test, which has been phrased by the Court as *regard must be had to the personal conviction and behaviour of a particular judge in a given case*, and

ii) the objective test, phrased as *whether the tribunal itself and, among other aspects, its composition, offered sufficient guarantees to exclude any legitimate doubt in respect of its impartiality*.[92]

Looking at the subjective test the ECtHR has stated that there exists a presumption that the tribunal is impartial, without prejudice and bias, until there exists proof of the contrary.[93] The characteristics of this proof must show that the judge has for example exhibited hostility or ill will for personal reasons in earlier cases or the case at hand.[94] The question is if it can be shown that the judge, or another member of the court such as the juror, has acted with personal bias.[95] As such, the presumption that the judge is impartial holds a strong initial position in ECtHR case law. The Strasbourg Court has acknowledged that it is difficult to show and procure evidence that a member of the court has acted with personal bias according to the subjective test. In order to remedy this the objective test provides a further guarantee.[96]

The objective test put forth by the ECtHR in case law is explained well in the grand chamber case of *Micallef v. Malta*. The test is portrayed as whether or not there can be objective questions raised against the impartiality of the court. The tribunal must itself, among other aspects, and by its composition offer sufficient guarantees to exclude any legitimate doubt of its impartiality.[97] An important factor to look at is whether there exists links between the judge and other actors in the proceedings, for example between the

---

[91] Harris, O'Boyle et al, p. 451 and cited case Morice v. France para 75.
[92] See Micallef v. Malta para. 93 and De Cubber v. Belgium para. 25.
[93] See Kyprianou v. Cyprus para. 119.
[94] See for example, Micallef v. Malta para. 94.
[95] Harris, O'Boyle et al, p. 451.
[96] Micallef v. Malta para. 95.
[97] Ibid para. 93.

judge and one of the parties. The relationship in question must have such nature and be of such degree that it indicates a lack of impartiality.[98] In the case of *Pescador Valero v. Spain* the Strasbourg Court states that professional relations between a judge and one of the parties may lead to objective doubt of the impartiality.[99] Other cases also cite personal or financial relations as cause for doubt.[100] Even the *appearance* of a link that can cause objective doubt must be taken into account when assessing the question of impartiality.[101]

Another important, but not decisive, factor is the fear of bias from the standpoint of the concerned party.[102] This fear must be objectively justified. As mentioned even the appearance of a link can be enough to objectively cause doubt. The ECtHR also stated in the *Micallef v. Malta* case, citing the case *De Cubber v. Belgium*, that "*justice must not only be done, it must also be seen to be done"*[103] which highlights the importance of the parties trust in the court's impartiality as well as the public view of the tribunal. If the confidence in the court(s) by the public is at stake because there is a legitimate reason to fear a judge's impartiality then the judge must withdraw from the case.[104] The appearance of the judiciary to the public is of great importance which is also emphasized by the ECtHR in the Micallef case.[105] The ECtHR has also stressed that national legislation for ensuring impartiality, such as regulating the withdrawal or dismissal of judges, shows that the legislator wants to remove reasonable doubt that the national courts are biased which is something the ECtHR takes into account when assessing whether a tribunal was impartial and if there can be legitimate fears that are objectively justified.[106]

In relation to racial discrimination and bias in the courts the ECtHR has found that if a judge takes sufficient precautions as to "*dispel any objectively held fears or misgivings*"[107] then the court is not objectively partial. In the cited case there was a fear of bias due to a note being passed from the jury to the judge saying that the jury showed

---

[98] Ibid para. 97.
[99] See paras. 27-29.
[100] Micallef v. Malta para. 102 and Wettstein v. Switzerland para. 47.
[101] Micallef v. Malta para 98 and Pétur Thór Sigurðsson v. Iceland paras. 45-46.
[102] Micallef v. Malta para. 96.
[103] Ibid para. 98 and De Cubber v. Belgium para. 26.
[104] Micallef v. Malta para. 98, see also Fey v. Austria para. 30.
[105] Micallef v. Malta para. 99.
[106] Ibid para. 99.
[107] Gregory v. the United Kingdom paras. 46-48.

racial overtones. The judge promptly warned the jury to lay aside any personal beliefs and biases after consulting with the prosecution and defence. This was deemed enough action to dispel any objective fears. Another case further develops the objective test in relation to discrimination.[108] After it is brought to the judge's attention that one juror has made racial jokes or remarks the judge asks the jurors to think through their ability to judge the case overnight based only on the evidence. Each juror affirms the next morning that he or she has no racial biases in signed letters to the judge. One juror also apologizes in a written statement to the judge for any racial remarks made but assures the judge of having no racial bias. The Strasbourg Court held that this admission of racist comments distinguishes it from the previous case. A dismissal of the jury would have been more appropriate in the latter case than vague assurances from written statements. As such, if racial bias can be objectively pointed out the court must take action. Since both cases concern bias in juries the applicability on a judge may be limited but nonetheless shows the importance of impartial proceedings and the importance of taking the fear of bias seriously. Additionally as stated earlier, if a judge cannot objectively judge a case he or she should, according to the importance of appearance mentioned earlier, withdraw from the case.

When looking at impartiality and the AI judge it is a disturbing thought that bias could be introduced systematically into the courts in the different forms of AI bias discussed earlier. Perpetuating injustice by solidifying existing bias, or creating new bias, in the judge itself would be a catastrophic development. However, this may not necessarily be the case. This will be explored further in chapter 4.3.

3.4.3 The Independence of the Court

An independent court is another essential part of a democratic society and cannot be overlooked. When assessing independence in relation to Article 6 ECHR one does not only mean independence from the other powers of the state, such as the executive government or the parliament. It is also important to take into account independence from both of the parties in a case.

The ECtHR has stated that Article 6 does not imply that the Contracting States are required to organize or separate the power of the judiciary and political power in a certain

[108] Sander v. the United Kingdom paras. 32-35.

way, nor does Article 6 have any constitutional implications regarding the limits of the interaction between the powers of the state.[109] Instead the importance lies in the case-by-case consideration whether the requirements of the ECHR are met in relation to the national legislation and the de facto connections between the judiciary and other actors.

The ECtHR has summarized the test of independence in four criteria that needs to be tried to cast doubt on the independence of the tribunal. These criteria have been repeated in extensive case law.[110] The criteria to assess the court's independence are as follows:

i) the manner of appointment of its members and
ii) the duration of their term of office;
iii) the existence of guarantees against outside pressures;
iv) whether the body presents an appearance of independence.

The above criteria seem to fall into three categories, (a) independence from the executive power, the legislature and the parties, (b) guarantees that the tribunal can operate independently and (c) that even the resemblance of dependence must be avoided.[111]

When looking at the independence from the executive, the requirement is that the tribunal, once appointed, cannot be pressured by the executive power before or during a trial. The ECtHR has stated that independence is undermined if the executive power can step in and influence the outcome of a pending case, regardless of whether or not the court has taken the executive's view into account. The mere incidence that the executive decides to or has the ability to intervene in a pending case may raise fear that the court is not independent.[112] The fact that a judge is appointed by the executive is not by itself a cause to question the independence according to criteria i), but it is instead important that the judge can serve its adjudicatory role without influence or pressure.[113] It is however not necessary for the appointment to be for the lifetime of the judge to ensure independence as long as they cannot be removed at will or on improper grounds.[114]

Independence from the parliament is as important as independence from the executive power. The ECtHR has however stated in the case *Sacilor Lormines v. France* that the

---

[109] Kleyn And Others V. The Netherlands, para. 193.
[110] E.g. Langborger v. Sweden, para. 32 and Findlay v. the United Kingdom, para 73.
[111] Barkhuysen et al, p. 600.
[112] Sovtransavto Holding v. Ukraine, para. 80 and Barkhuysen et al p. 601.
[113] Campbell and Fell v. the United Kingdom, para. 79 as well as Flux v. Moldova (No. 2), para. 27.
[114] See Barkhuysen et al, p. 600.

fact that a judge is appointed by the parliament does not by itself mean that the judge is subordinate to the same. This is similar to the independence from the executive. Nor does it mean that the judge is not independent as long as the judge cannot be pressured or receive instructions on their adjudicative role from the parliament.[115] The requirements for the independence from the executive and legislative power is therefore of similar nature and focuses on whether pressure can be exerted.

Lastly concerning independence from the parties the ECtHR has stated in the case *Sramek v. Austria* that if a member of a tribunal holds a subordinate position in terms of the duties or organization of the member, there may exist a legitimate doubt from the other party that the member in question is completely independent.[116] Certain situations may therefore cast doubt on the independence of the court. If the judge is part of an organization that one of the parties holds a superior position in this may be a cause for doubt. This clearly overlaps the requirement of impartiality as well.[117] The implications for the AI judge will be further discussed in chapter 4.3.

The next chapter will focus om the nature of the hearing itself and what is to be expected of a fair trial in relation to the following judgement.

3.5 A Fair Hearing and a Reasoned Judgement

3.5.1 What Is a Fair Hearing?

The detailed characteristics of what constitutes a fair hearing has been avoided by the ECtHR in terms of formulating an exhaustive number of criteria.[118] In the case *Kraska v. Switzerland* the ECtHR stated that the purpose of a fair hearing in Article 6(1) ECHR is: "*to place the 'tribunal' under a duty to conduct a proper examination of the submissions, arguments and evidence adduced by the parties, without prejudice to its assessment of whether they are relevant to its decision*"[119]. The proceedings must therefore be assessed on a case by case basis in order to determine whether a hearing has been fair or not.[120] Impartiality and independence plays a role here as well under the prerequisite of

[115] See para. 67.
[116] See para. 42.
[117] Barkhuysen et al, p. 601.
[118] Ibid, p. 561.
[119] Kraska v. Switzerland, para. 30.
[120] Barkhuysen et al, p. 561.

'prejudice'. Additionally, the ECtHR has stated that it is the proceedings as a whole that when looking at the bigger picture has to represent fairness.[121] Certain aspects may however violate the notion of fairness regardless of how far the proceedings have come. Such as the collection of evidence in a preliminary hearing.[122] The application of Article 6 in this regard may also depend on the stage of the proceedings, for example the requirement of publicity may be less strict on cassation proceedings than other proceedings.[123]

The main differences between criminal and civil proceedings concerning a fair hearing is outlined in article 6(2) and 6(3) that when taken literally only applies to criminal proceedings. Although, in principle the paragraphs also apply to civil proceedings as well as administrative proceedings which has been mentioned earlier.[124] Since the requirements for civil proceedings are not as explicitly stated compared to criminal proceedings in 6(2) and 6(3), even though the requirements apply in principle, the characteristics of a fair hearing in civil cases are not necessarily identical to those of criminal cases. As such there exists a wider margin for what is regarded as fair when dealing with civil proceedings.[125] For example the fact that cross-examination of witnesses in Article 6(3) applies to both civil and criminal proceedings may not be as strictly interpreted in relation to civil proceedings as for criminal proceedings.[126]

The right to a fair hearing includes the *right to be present* at one's hearing for criminal cases which follows from the object and purpose of Article 6 as a whole even though it is not explicitly stated.[127] Civil litigation does not guarantee the same right to be present but instead the *ability to present one's case effectively* needs to be upheld in order to live up to an equality of arms between the parties.[128] If one party cannot present its case effectively the conditions are unequal. This means that civil disputes can be judged based only on the written statements of the parties. Certain civil cases may however be of such

---

[121] Kostovski v. the Netherlands, para. 39.
[122] Barkhuysen et al, p. 561.
[123] Ibid, p. 561.
[124] See supra chapter 3.1.
[125] Barkhuysen et al, p. 562.
[126] Harris, O'Boyle et al. p. 411.
[127] Schabas, p. 316.
[128] Harris, O'Boyle et al. pp. 411-412.

nature that the right to be present at the proceedings should still be upheld.[129] Since our AI model[130] is assumed to function no differently in this regard than a human judge we will not delve deeper into the presence of the parties during the proceedings. However, it should be noted that if an AI judges substitutes a human judge to preside the trial it may affect both the nature of the proceedings and how well one can present the case effectively.

Another facet of the right to a fair hearing is the right to *effectively participate* in the trial. This right overlaps the aforementioned rights to be present at the proceedings and the ability to effectively present ones case.[131] The case of *Stanford v. the United Kingdom* clarifies that there is both a right to 'follow the proceedings' and a right to 'participate effectively' during the proceedings. In the aforementioned case the applicant's ability to participate effectively was impaired by his hearing difficulties as well as the acoustics of the court room in question which made it hard to hear the proceedings taking place.[132] The ECtHR however found no violation of Article 6 in this regard. Another case concerning the effective participation, *V v. the United Kingdom*, concerned the excessive formalism and nature of the proceedings. The fact that the accused was 11 years old at the time combined with the nature of the proceedings as well as the stress and fear afflicted by the situation meant that the accused could not effectively participate in the proceedings.[133] This case is on one side of the extremes since it concerns a young child and in relation to an adjudicatory AI it can be noted that, again, the assumption made is that the formal procedure of the trial should not be affected as the AI would act according to procedural law if presiding a trial. However, it cannot be completely ruled out that substituting a human judge with an AI equivalent could have effects on the parties that adversely affects the parties ability to effectively participate.

---

[129] Ibid, p. 412 for a longer list of examples of such cases.
[130] See supra chapter 2.6.
[131] Harris, O'Boyle et al. p. 414.
[132] See para. 26.
[133] See Harris, O'Boyle et al. p. 415 and V. v. the United Kingdom, paras. 17-18.

3.5.2 The Requirement of a Reasoned Judgement

In both civil and criminal cases the court is obliged to arrive at a final decision.[134] Furthermore, the reasoning leading up to the decision is required to be sufficiently clear as to allow a litigant or accused to meaningfully use any right of appeal and to understand the court's decision.[135] If a decision cannot be sustained by the reasoning of the court the likelihood of forming a successful appeal on material grounds is limited. The interest of the public in the reasoning behind a judgement is also important, especially in high profile cases.[136] Regarding criminal proceedings the accused should have the right to know why they were convicted and on what grounds. Additionally, according to my opinion the principle of legal certainty cannot be upheld if judgements are not supported by an adequate reasoning. It is hard, if not impossible, to deduct what importance different circumstances of the case had on the judgement if no legal reasoning is presented.

The ECtHR has stated that precisely how articulate the reasoning behind a judgement has to be depends on the case at hand. The Strasbourg Court also stated that the requirement does not imply that the court necessarily has to respond to every argument or point made.[137] Despite this, should a point raised be decisive for the outcome of the case it may need further elaboration by the court.[138] Furthermore, should the given reasoning not be good in law or on the facts it is not of satisfactory nature.[139] Depending on if the proceedings take place in a lower court or have been appealed to an appellate court the requirements may also differ. A lower court has a greater responsibility to give a well-reasoned judgement whereas an appellate court may reference or incorporate the reasoning of the lower court.[140] One can imagine if the lower court fails to sufficiently justify its judgement a greater responsibility falls on the shoulders of the higher instances to rectify such inadequacies should the appeal be successful.

In relation to our AI model it should be noted that the assumption made is that the AI is able to form convincing legal judgements. How eloquently the AI judgements are

---

[134] See e.g. Lupeni Greek Catholic Parish and Others v. Romania, para. 86 and Marini v. Albania, paras. 118-123.
[135] Harris, O'Boyle et al, p. 431.
[136] Ibid, p. 431.
[137] Garcia Ruiz v. Spain para. 26
[138] Ruiz Torija v. Spain para. 19.
[139] Harris, O'Boyle et al, p. 431 at note 582 and 583.
[140] Ibid, p. 431.

formulated would be a factor dependent on the prior case law the AI is trained on. The issue of inadequate reasoning would therefore be assumed to be as common or uncommon as the existing case law makes it.

3.6 Summary on the Right to a Fair Trial

In summary the right to a fair trial of Article 6 ECHR embodies multiple principles and rights that make up the notion of what is considered fair. The Article is complex and the most invoked article of the convention. The right of access to court is an essential right that enables all other rights protected in the ECHR to be secured. When the access to court is impaired or denied the individual's access to justice is crippled. Subsequently what constitutes a court, or tribunal, in the autonomous meaning of Article 6 ECHR is important in order to determine if an AI judge can fulfil the requirement of a tribunal as well as the access to court doctrine. The prerequisite is also that the court is also established by law, which is a technical hurdle that needs to be cleared by legislative means.

Concerning the principles of impartiality and independence the notable thing to keep in mind is that the criteria are overlapping in certain ways, or at least they are closely related or hard to separate and may therefore be examined jointly. The fact that a court must live up to basic standards of impartiality and independence is a cornerstone of the rule of law. It also has a causal relationship with the public's view of the whole justice system. There are presumptions that the court is impartial and independent and therefore proof is required to doubt the impartiality or independence. Therefore an AI judge would initially be presumed impartial and independent until proven otherwise.

As for the character of a fair hearing the fairness itself is dependent on multiple factors. The proceedings as a whole has to be seen as fair which naturally incorporates the impartiality and independence criteria. An overall examination of the proceedings is therefore important and has to be done on a case by case basis. Additionally, the proceedings has to allow for the parties to effectively participate in the trial as well as be present if the nature of the case requires it. The court's judgement must be supported by a sufficiently articulated reasoning as to allow for the parties to be able to form an appeal in a meaningful way.

Some of the aforementioned principles and rights that make up the right of a fair trial may be affected if an AI judge is implemented, other principles or rights may not be

affected directly if the assumptions of the AI model hold true. Nevertheless, there is room for a reasonable concern that an AI judge may in fact have a negative impact on the judiciary in light of these principles and subsequent rights. In the next chapter these concerns will be more thoroughly discussed and challenged.

4 The Collision of AI and a Fair Trial

4.1 Introduction

The technology behind the current stage of artificial intelligence may prove incompatible or give way for issues that are in violation of the right to a fair trial. In this chapter the collision of artificial intelligence and different aspects of the right to a fair trial are explored. Certain problems that have been identified in chapter 2 and 3 may have technical solutions while others are dependent on soft values such as trust, how justice appears to the public and how fairness is perceived by the parties in a dispute which require something more than technology to solve.

In order to analyse the collision of artificial intelligence and the right to a fair trial a starting point is to look at the institutional requirements of the tribunal as well as the formal requirements of the proceedings and compare this to the possibilities of an AI judge. As we have seen in the previous chapter there are certain components of a trial that cannot be overlooked for the trial to still represent fairness. When the formal requirements have been analysed we can move on to discuss the aspects of impartiality, independence and a fair hearing. Finally a discussion about transparency, trust and perceived fairness is well-needed to answer the question whether or not an AI judge is a suitable substitution of a human judge.

4.2 The Formal and Institutional Requirements of the Tribunal

4.2.1 Organization of the Tribunal and Appointment and Dismissal of AI Judges

At the centre of the question whether an AI judge could replace a human judge are the formal requirements that have been explored in chapters 3.2.2 and 3.3. By looking at the prerequisite of Article 6 that a tribunal has to be *established by law* it is hard to refute that an AI could technically be given the appropriate legal basis for its judicial functioning. The AI judge would also have to be able to issue binding judgements within its competence and perform its duties in a prescribed manner in order to comply with the strictly formal requirements. As there is no constitutional obligation to structure the court, apart from the prerequisite of independence from the executive and legislative power, in a certain way an AI judge does not by definition seem to violate the ECtHR's definition of a court or tribunal. As long as the same or similar laws that apply to human judges in

terms of the tribunal's legal basis is established for AI judges there seems to be no formal issues. The fact that a human judge is substituted with an AI judge should therefore not by itself cause the court to not be considered *established by law* as long as the national law recognize an AI judge as a proper judge equal to a human judge. Important aspects of impartiality and independence would naturally have to be taken into account as well, which we will discuss in detail later in this chapter.

The process of appointing an AI judge may prove to be different than that of a human judge which demand attention. Making appropriate comparisons between the appointment of human judges and AI judges here is difficult in a meaningful way. An AI judge exists as a program that is run on a computer. As such it is arguably not applicable to talk about appointments (or dismissals) of AI judges in a traditional sense. The technical intricacies of the fact that we are dealing with an algorithm makes it even harder. Is it the *computer* that runs the AI algorithm that is appointed, or is it an *instance* of the AI algorithm that is appointed? This may come off as plain bureaucracy or nitpicking and most people may say that it is of course the algorithm and not the physical computer that would be appointed. Nevertheless the fact that several instances of the same AI could potentially be run on the same or different computers requires certain reflection as to exactly what is being appointed.

When discussing dismissals of judges it also becomes apparent that an AI judge could simply be dismissed by switching off the computer or suspending the process on the computer. Convenient, but of course also highly vulnerable. It could of course be argued that a human judge is also inherently vulnerable due to humans mortal state. And maybe even more so than an AI judge since our consciousness is not preserved if we are "shut off" and we cannot be backed up in a secure data facility if need be. The issue at hand is that the legislation regarding both appointments and dismissals of judges would have to be adjusted to account for these differences between a human and an AI. A dismissal of an AI judge may also not carry the same implications as dismissing a human judge that may carry, for example, political undertones. As such the ECtHR's view on arbitrary appointments, renewals and dismissals of judges may not be directly applicable on AI judges.

As there naturally is no relevant case law on the area yet it is hard to determine what the Strasbourg Court's position would be. As mentioned briefly in the introduction this

issue may have a technical solution for a developer as well as a legislative solution for a legislator.

4.2.2 The Requirements of the Trial

The overall requirement for the proceeding in relation to Article 6 ECHR is that the proceedings has to represent fairness when looking at the case in its entirety.[141] This means that if the proceedings as a whole cannot live up to a certain level of fairness it is deemed unfair. The *Kraska* case[142] further states that the tribunal has to conduct a proper examination of the material, arguments and evidence presented. Concerning our model AI the ability to conduct a proper examination should be no less than the ability of a human judge if we assume our AI model is good. If anything one could argue that an AI judge would be able to recall information in a much more efficient and precise way than a human judge since it would have instantaneous access to the case files digitally or in memory. This procedural requirement of Article 6 would therefore most likely not be an issue for an AI judge.

What may turn out to be problematic is the right to effectively participate in the trial. Assuming that the AI judge would be equal to a human judge in the way the judge conducts the trial and interact with the parties in the formal sense there may still be differences as to how an AI judge is perceived by a human. As seen in the case of *V v. the United Kingdom* even though the trial is not conducted in an irregular manner the parties' ability to effectively participate is dependent on the nature of the proceedings. In relation to an AI judge one could argue that especially in cases involving children it may be inappropriate to remove even more of the 'humane' aspects of an already strenuous situation. The aforementioned case is however of such extraordinary nature in relation to other ECtHR cases concerning Article 6 that drawing too general conclusions may be unwise.

The final point of discussion in relation to formal requirements is that of a reasoned judgement. As mentioned earlier in chapter 3.5.2 the ECtHR has stated that the national courts are obliged to provide a sufficiently reasoned judgement as well as obliged to provide a final decision in a dispute. Our model AI is assumed to be able to present

---

[141] See supra chapter 3.5.1 at note 104.
[142] Ibid at note 102.

convincing legal reasoning that is indistinguishable from the same work of a human. Thus it should be less of a problem to fulfil the requirements. An important remark to be made is that when discussing AI judges it may be of value to see the 'reasoning behind the reasoning'. It is one thing for the AI to present convincing legal reasoning but it is another to know exactly which facts have been taken into account in order to know if there are unwanted biases in the neural network. One could however argue that the same issue is applicable for human judges. We cannot for certain *know* which facts the human judge weighed or if the judge carries subliminal bias. The main difference between the human judge and the AI judge is that we have created the AI and are also able to finely tune the algorithm at our own discretion. With that in mind it seems wrong not to demand to know which variables have been taken into account by an artificial judge created by ourselves. It is also hard to demand a detailed account of which facts have been taken into account and how they have been weighed by a human judge since the process of judging in many ways happens intuitively after experience or at least not in a step-by-step way that can be visualised by a flow chart or similar way. If this possibility of explainable decision making is present in AI judges we should demand that it is used. A quick note is that even though it could be theoretically possible to provide a flow chart of how the AI has reasoned (i.e. a map of the nodes and their weights) the result may not be interpretable, as in not give any meaningful information, by humans.[143] This will be covered in more detail later in this chapter when we are looking at the transparency problem.

4.3 Artificial Intelligence, Impartiality and Independence

4.3.1 The Issue of Impartiality

In this chapter our AI model will be tried against the impartiality tests put forth by the ECtHR. Initially it can be said that the subjective test is difficult to apply on a theoretical level since it is only speculative whether the AI judge would show such personal bias. As subjective impartiality is also presumed there needs to be concrete evidence that the AI holds a personal bias. One could argue that if the AI judge has a history of judgements where it can be shown that the AI took into account variables that implies bias then the presumption could be rebutted. Another thinkable way to rebut the presumption would be

[143] See Samek & Müller, pp. 7-9.

if one could show that the neural network in fact is biased. As mentioned earlier it may be hard to procure evidence of personal bias and it is of less interest to speculate whether or not this evidence will be more or less complicated to procure for AI judges.

Moving on to the objective test which is more easily discussed hypothetically. Can there be *objective* concerns of the AI judge's impartiality? The criterion that there should be no links between the judge and either of the parties in a case may not in general be a cause for any doubt. As a matter of fact since the AI judge would have no personal connections to the outside world one may argue that it would even be better than a human judge as it has no real incentives of being biased to favour personal relations. However if one of the parties of a dispute were involved in developing the AI algorithm it may certainly cause initial concern regarding the AI's impartiality. How can one party know for certain that the algorithm does not favour the developer? If the developer is a well-known and established company this may cause great concern should the company actively engage in litigation. Another concern is the case of state-funded or state-developed AI judges. How can a person accused of a crime know that the AI judge is not partial towards the prosecution, i.e the state, or otherwise biased? A direct link between the state and the AI judge could undoubtedly cause doubt of impartiality. As mentioned in chapter 3.4.3 even the appearance of a link may be concerning.

Another concern when it comes to bias is the for us unknown bias that the AI judge could conclude from previous cases which will be perpetuated in future adjudication. A real example could be that for certain crimes, especially concerning the possession and use of narcotics[144], the law enforcement targets certain groups or individuals in society more often than others via so-called criminal profiling. The selected cases in the training data will seem to show a correlation between belonging to the often profiled social group and being guilty or at higher risk of committing the crime. However, correlation does not imply causation which may cause issues for an AI judge to properly determine and take into account. This is a logical fallacy that even humans fall victim to. How can we effectively know that the AI judge does not put a link between certain crimes and certain groups? Each case must be tried on its own merits and evidence, and such correlative

[144] See Fellner, pp. 257-291.

conclusions may damage the development of justice and only perpetuate a status quo in case law.

Another reason to remain sleepless over AI bias are the unforeseen biases that may show up after months or years of the AI being in use, as the case with the HR recruiting tool mentioned earlier showed. Such biases will ultimately affect lives of people in a negative manner. When minorities, underprivileged or marginalized groups are subject to an AI judge's decision there might be even more incentive to make sure that bias is not affecting the outcome of a case since these groups already have a weak position in society where structural bias may hurt even more.

In general as mentioned in chapter 2.3.3 bias is a legitimate fear when discussing artificial intelligence. What may seem like an unbiased AI may also with time turn out to have systematically discriminated against a certain gender or even on the mere basis of names. The thought of such discrimination from within the judiciary being perpetuated systematically via artificial intelligence, in essence automated decisions, is frightening and the consequences abhorrent. Artificial intelligence without proper safeguards against bias and impartiality cannot be accepted.

Looking exclusively at the machine learning process where the AI is assumed to be taught on previous cases, there is the underlying possibility that bias is carried over from the training data. If the training data in turn reflects our existing structural problems in society how can we possibly hope to overcome social injustice, inequality or discrimination? The big difference from a human judge and our model AI is that a human can reason and think outside the limitations of the current scope. The possibility that an AI would be able to determine how a decision may affect society in broader terms is in my opinion slim at best. The likely scenario seems to be that the AI would become too reliant on previous case law without the ability to apply the law in novel ways, for example with the help of analogous reasoning. If all human judges would be replaced by AI judges the development of legal discourse would risk stagnation and solidifying existing biases instead of dispelling them. On the other hand, these issues could have a programmatic solution. Can we not program the AI to not account for variables such as race in general (naturally not in cases where the dispute concerns race such as discrimination lawsuits)? This does however not help with the fact that the AI itself may develop causal links between correlating variables. These issues are also much harder to

detect due to the lack of transparency in deep neural networks which will be discussed further in chapter 4.4.

There can evidently be reasonable doubt of the impartiality of the AI judge if the developer of the AI is one of the parties in a pending case or bias carried over from the training data. If the state or government has funded or developed the AI judge the accused could have a legitimate doubt that the AI can uphold the presumption of innocence or not draw conclusions based solely on correlations in previous case law in a way that serves the state. Political dissidents may also have a legitimate doubt that the AI is in favour of the government if the AI was funded or developed by it. Silencing dissidents through AI (in)justice may seem unrealistically dystopian but nonetheless from the standpoint of a dissident one would certainly like to be assured that the judge lacks any such ties to the government or state.

To roughly evaluate the impartiality of our AI judge in relation to the impartiality tests one could say that there are legitimate fears that bias may exist or develop over time. Personal bias is still hard to prove to rebut the presumption of impartiality. Meanwhile the objective links between the AI, its developer or the state can be a cause for legitimate doubt. Nevertheless there still exists the possibility that the legislator could implement regulations that makes AI judges viable if they can be dismissed and replaced by a human judge. Such regulations and safeguards could ensure that cases be tried by humans instead of AI judges. In certain ways this may however defeat the purpose of implementing AI judges if the goal is to streamline the judiciary and cut the time of lengthy proceedings. One of the hardest obstacle to overcome will be the appearance of the court to the public. A general scepticism towards an implementation of AI judges is possible in light of the problems discussed. This in turn may have a negative effect on the perceived fairness of the courts even if the AI judge is compliant with the requirements of national law and case law of Article 6 ECHR.

4.3.2 The Issue of Independence

The case of judicial independence is important for the separation of powers within a state and to ensure that no single interest gets favoured by the courts. As such there exists three situations where a court must uphold its independence:

i) a dispute between the government and a private actor;

ii) a dispute between two branches of government; and

iii) a dispute between two private actors.

The focus in this chapter will be on situation i) and iii) as these affect individuals' rights protected by Article 6. Even though in principal the second situation is also important in terms of independence the situation does not typically concern the right to a fair trial where one party is an individual.

The first situation can be illustrated by a criminal case where the prosecutor, i.e the state, accuses an individual of a crime. In this situation it is of high importance that the state cannot pressure the court to rule in favour of the prosecutor. Naturally it is important that the individual cannot pressure the court as well, but due to the "monopoly on violence" that the state enjoys compared to an individual this situation can be disregarded for now. The independence of the court is assessed based on the four criteria in chapter 3.4.3. Starting with the first and second criteria, manner of appointment and the duration of office for judges, we can make the assumption that while certain legislative measures would have to be made to adjust the current regulations to fit AI judges there should be no essential difference from the way human judges operate. As such there should be a presumption that in this context the AI judge would be equally independent as a human judge.

Regarding the third test criterion, i.e. the guarantees from outside pressure, one can again question the development and funding of the AI judge. The traditional way of thinking about pressure is meaningless for an AI judge as it is pointless to threaten it with removal of office, dismissal or personal consequences. The AI judge would be indifferent to such pressure as we are not dealing with a strong AI that comprehends and values its own existence like a human. The way pressure could be exerted is if the AI judge is either biased in favour of the state or there is an inherent security flaw that allows for manipulation of the AI. The possibility of such a backdoor into the AI would certainly mean that the judge could never be seen as truly impartial. As such there would need to be technical solutions or transparency in place to ensure that no such manipulation is possible. One such solution could be open source code or audited code by independent auditors. Something to also keep in mind is that the AI would be running on computers which are susceptible to hacking or exploiting meaning that there would be an ever looming threat of outside pressure. Security would play a crucial role if the AI judge is to be guarded from outside pressure.

The fourth criterion which is the courts appearance of independence is vital for the appearance of the judicial system as a whole. If there are legitimate and justifiable reasons

to distrust the judiciary based on the lack of appearance of independence from the state then one can question the fairness of all trials where the state holds interest in the outcome. Would an AI judge automatically enjoy the same appearance of independence as a human judge? Depending on the factors discussed earlier, such as security, development and possibility of bias, my opinion is that AI judges would have to prove themselves much more than a human judge. The AI judge should not be blindly trusted without proper due diligence as we have historically seen that unforeseen consequences, such as discriminating behaviour, may occur. Would there therefore be legitimate reasons to fear that the court is not independent, and would this fear be objectively justifiable? Since the model AI is based on known technology and that an inherent issue with these technologies, i.e. the machine learning process as well as the "black-box"-element of deep neural networks, is the transparency and ability to rule out bias there is at least a legitimate reason to fear that the AI is not completely independent. This fear could be objectively justified as previous examples of AI implementations support the argument that AI could hold bias. One can imagine that this fear could be very harmful to the court and the trial itself could be perceived as unjust to a person accused of a crime. From the perspective of a wrongfully accused person it may seem "Kafkaesque"[145] to be stuck in a trial judged by an AI judge that potentially is not independent from the prosecutor or otherwise not impartial. However if the issue of AI transparency could be averted then the fear of a lack of independence (as well as impartiality) may be less justified. One could argue that to the public a transparent AI judge may even seem more independent and unbiased than a human judge who of course hold political views, personal prejudices and biases and is dependent on factors such as money, reputation and future career.

The second situation where there is no state actor active in the dispute the characteristics of independence from the parties is different in the sense that the parties are more equal, yet the aforementioned questions remain. The central questions are therefore still if there are guarantees from outside pressure as well as the overall appearance of independence. Should one of the parties be the developer of the AI the aforementioned issues arise. The looming issue of flawed security which could open up

[145] As in a bizarre or surreal situation that imposes great feelings of helplessness or injustice without any practical possibility to get out of the situation.

for outside pressure is still present. The key seems to partially lie with transparency. If the model AI is sufficiently transparent the parties can not in the same way hold objectively justifiable fears that the AI is not as independent as a human judge could be.

To summarize the issue of impartiality and independence, since these terms overlap as mentioned earlier, there are legitimate fears that an AI judge may not in the same way as a human judge embody these core values of a fair trial. The solution seems to be a high level of transparency and security of the AI that could rebut such fears. The next chapter will discuss the transparency in more detail.

4.4 Transparency and Trust

Transparency is not a literal prerequisite of Article 6 ECHR however it can certainly be derived in some way from the underlying *right to a fair hearing* which requires certain transparency, *a reasoned judgement* which requires transparency of how the judgement is justified as well as the *publicity of the trial* to the public which ensures transparency of the proceedings. Aside from this the transparency seems to be connected to the public's trust in the judiciary on a general level. One can therefore argue considering that the model AI is a "black-box" a certain degree of trust may be lost. If there is a lack of trust then there could be a disinclination to embrace and accept the judgements of the courts which in turn weakens the rule of law as well as the constitutional position of the courts. The level of trust in the judiciary could be assumed to be a measurement of the overall health of the judicial system.

The transparency of the courts, or the judiciary at large, has a direct effect on the way justice is perceived as being properly dispensed. In broader terms a lack of transparency in society means that corruption, where bias is but one facet, more easily can gain foothold. A transparent system makes it easier for the public to examine and question the powers and trust that the rule of law is followed. Transparency within the courts should therefore be a priority and even if it is not an explicit prerequisite of Article 6 ECHR I would argue that transparency effectively enables a right to a fair trial and an overall healthy judicial system.

As mentioned earlier the issue of transparency as well as bias is known within the field of artificial intelligence and machine learning. Therefore an AI that lacks a basic level of transparency will most likely not realistically be at risk of being implemented. It has also been theorized that exactly how we determine the level of transparency is up for debate

and may differ from situation to situation depending on whom the transparency concerns.[146] For example there is a difference in the transparency a developer requires, i.e. full access to the source code, and the transparency a member of the public may request. To the latter a piece of the source code is most likely useless or otherwise not interpretable in a meaningful way. Transparency may therefore mean different things to different individuals. A party in a dispute may want transparent reasoning that justifies the outcome of a case. An individual accused of a crime may want to be ensured that no racial bias is present. These two aspects of transparency is maybe the most valued when discussing the use of AI judges. The issue of transparency and trust must be taken into account and given proper thought when assessing the overall appropriateness of AI judges.

4.5 Concluding Thoughts on the Effect of AI on a Fair Trial

In relation to the right to a fair trial in Article 6 ECHR the conclusion is that while there may not be any distinct formalities that would prevent an AI judge to be implemented, as long as the judge can live up to the same requirements put on a human judge in terms of the procedural requirements which is assumed, there would most likely be severe issues with impartiality and independence. These requirements are technically also part of the formal or institutional requirements put on courts but in this context they have been discussed as separate criteria.

Another issue is that of transparency and the trust in courts. This could be said to not be strictly included in the literal context of Article 6, however since the ECtHR has put emphasis on the appearance of the courts it is not unimaginable that transparency is a facet of the appearance which in turn affects the trust of the judiciary.

If the issues of impartiality, independence and the appearance (transparency) of the AI judge is not solved one can easily imagine that it would risk undermining the right to a fair trial.

[146] Weller, p. 24-25.

5 The Artificial Judge and the Character of Justice

5.1 Introduction

The previous chapter discussed the collision of a fair trial and AI judges. This chapter will focus on discussions on other difficulties of AI adjudication as well as characteristics of justice that an AI judge would have to solve or otherwise reflect to become on par with a human judge.

5.2 Difficulties of Training an Artificial Judge

An issue to discuss is the difficulty of training the model AI, or any AI judge for that matter, while still having a judge that is up to date on the state of the established law. If the AI judge, presumably presiding in a lower court, is only trained on case law from the lower court(s) the judge would not have the full picture of the law since any judgements from appellate court(s) or the supreme court would be left out. As such the fear of stagnating the development of the lower courts may come true. There is also the possibility of a lower accuracy due to fewer cases to train the AI judge on.

However, should we instead train the AI judge only on supreme court cases to give it a "proper idea" of what the established law is we may face the issue of not having enough data to reach satisfactory accuracy of the AI. As discussed earlier the risk of bias is greater if the training data is not sufficient. This kind of training does not seem better than the previous.

The last idea would be to train AI judges on all of the case law from lower courts, appellate courts and the supreme court. This would most likely be enough to provide sufficient training data. The result should be an AI judge that is well equipped with "knowledge" of both the lower courts and the precedents set in the supreme court. One could ask if this is done, could the resulting AI judge not replace our supreme court judges? Should we not trust its judgement as being correct? We will come back to this point in chapter 5.4.

A problem that has not been mentioned earlier is the difference of common law and civil law systems. While the common law countries may have it easier to create and train an AI judge that is based on case law, a civil law country may have trouble as case law is not a source of law in the same manner. Civil law relies on codification of the law which

may create problems when an AI judge cannot rely on previous judgements but has to make its own judgement only based on a framework of rules. On the other hand programming an AI might also be easier if codification of the law can be translated to code that the AI could interpret programmatically. In general another difficulty which affects both legal systems is what happens when new legislation is adopted. No previous cases can be found and the AI judge would therefore have to either analogously adapt its previous knowledge or be unable to apply the new legislation as it would be out of scope of its narrow field. An AI judge that were to apply codified law in a civil law system may have an advantage compared to a common law AI judge. This would surely have a major impact on the aspect of fair trials and justice depending on which legal system is adopted nationally.

5.3 The Importance of an Appearance of Fairness

As discussed in relation to impartiality and independence in chapters 3.4 and 4.3 there is an emphasis by the ECtHR that "justice must not only be done, it must also be seen to be done".[147] In light of the statement by ECtHR in the *Micallef v. Malta* case that there is an importance that the courts inspire confidence, i.e. trust, in the public there could be said to be a more general importance of the *appearance of fairness* of the courts. If the process or trial itself does not inspire confidence then it does not matter if the result is right, the confidence in the courts will be damaged. If the public does not trust the courts then the judicial system will be undermined as disputes may be settled outside the courts or judgments from the courts may not be respected. An AI judge would therefore have a substantial burden not to seem to perpetuate injustice or give the appearance of biased proceedings even if the judgements are correct.

Without delving deeper into exactly what *fairness* is one can still conclude that it is not only the result, i.e. judgement, that has to be fair and justifiable. The system as a whole must represent fairness otherwise it has lost something of great symbolical value and risks losing the trust of the public.

[147] See supra note 87.

5.4 The Human Aspect of Adjudication

Is there an inherently human aspect of adjudication that an AI judge could never fulfil which cannot be disregarded? Looking back at the model AI we can deduce that the AI is assumed to be fully capable of producing convincing legal reasoning. For the sake of discussion we are not taking bias into account here. If the AI reasoning is fully on par with a human judge – why should we not accept its judgements? It would make no difference in terms of result or effect if the outcome is the same and the justification for the outcome is the same. Is there therefore any *de facto* difference in being judged by an AI judge and a human judge?

One considerable difference between an AI and a human is the accountability. A human is both legally and morally accountable for his or her decisions. An ideal human judge would abide by moral standards as well as the legal requirements. If the human decides to deviate from what is morally justified then they could be held morally accountable for said deviation, at least in theory. A human judge is also assumed to have a conscience and a concept of what is fair which could be argued to act as a "soft" safeguard against morally wrong decisions even if it is by no means guaranteed. If a human on the other hand deviates from what is legally required then there may be repercussions such as dismissals or even charges of professional misconduct. These aspects of accountability would arguably be less effective in keeping an AI judge to abide by what is *right*. An AI judge cannot be assumed to have human morals other than what may be included in the training data or pre-programmed in the algorithm by the programmer. Any repercussions that an AI faces is essentially to be shut off. Is this a satisfactory repercussion if one is subject to AI misconduct? To some people it could surely be sufficient but to others it may not have nearly enough effect.

Aside from accountability there is also the aspect of being tried by one's peers. While a human jury may still be in place with an AI judge presiding a considerable difference would take place in disputes that do not have juries. The parties would therefore presumably be met by an AI judge through the entire proceedings. It cannot be ruled out that this "lack of humanity" could have effect on the characteristics of the proceedings before a court. Having a human serve the adjudicatory and presiding function in a trial could arguable create a feeling in the parties of being sufficiently heard as well as acceptance of the judgement. In light of Article 6 ECHR and the right to effective

participation one could therefore reasonably argue that the effective participation may be altered if the parties are met by an AI and no human is present.

The importance of being judged by one's peers, the peer being a human judge or a human jury, may not be so easily dismissed. However, one could challenge this view by acknowledging that human judges and human juries are evidently no safeguard against for example wrongful convictions. Humans are not perfect and therefore an AI trained on human reasoning and laws created by humans will not be perfect. Can we reasonably put the expectation that an AI judge is to be perfect before we implement it? My opinion is that if we are to change something that works in general, the change has to ensure an improvement from the status quo.

One way to possibly achieve commonly sought after goals such as lower procedural costs and more efficient proceedings could be to establish special courts with limited jurisdictions where an AI judge could act as an adjudicator. In order to comply with the access to court doctrine the "AI court's" decisions could be appealed to the appellate courts where humans could still preside, thus not restricting the core of the right. Certain types of cases may be well suited for such special courts while others with too complex matters or cases that are too sensitive might be more appropriately decided by human judges. By doing so we may find a middle way between procedural efficiency and respect for the rule of law. However it could be questioned if the judgements from the AI court would not systematically be appealed and therefore only shift a heavy case load from the lower courts to the appellate courts. One could however argue that this is already the case with lower courts today. Although disputes could initially be solved in the lower AI court in a short amount of time and as such give people quicker access to justice with the ability to appeal. The Council of Europe Commissioner for Human Rights have opposed artificial intelligence being able to infringe on human rights without the affected person having access to court.[148] In my opinion if AI judgements could be appealed to an appellate court of humans this may be less of a problem but nonetheless an issue that should not be too easily dismissed.

Lastly there is a deeply existential and constitutional issue of giving up the power of judging from humans to an artificial intelligence if we cannot appeal the judgement to a

[148] CoE Commissioner for Human Rights, *Unboxing Artificial Intelligence,* pp. 13-14.

court of humans. By doing so we would be subordinate to the artificial intelligence and a fundamental part of human sovereignty would be given up, especially if AI judges would be implemented in the appellate courts or even the supreme court. How else can we ensure that the AI judge is just without in turn subordinating the AI to our judgement and hold it accountable? *Quis custodiet ipsos custodes* – who watches the watchers?

6 Concluding thoughts

6.1 Artificial Intelligence and the Right to a Fair Trial

This thesis is based on assumptions made of how an AI judges could function which has been put into a model that the discussion is based on. If this model was to be implemented in the courts right now we can conclude that there are certain associated risks that may follow. The main issues are the impartiality and independence of the AI judge. It is up for debate if it is an intrinsic property of the which results in us never being able to know for certain whether an AI judge is wholly unbiased or completely independent. This could however be challenged by the fact that human judges are by no means perfect and also carry bias. The main difference in my opinion is that an AI judge is created by us and should therefore be made to lead to *better* results than what we can achieve ourselves and not focus too heavily on efficiency or cost reduction of the proceedings.

Would an AI judge therefore undermine the right to a fair trial? In my opinion there is a substantial risk that an AI judge, if adopted too early when the technology is still too novel, could damage the trust in the courts and not live up to the requirements of Article 6 ECHR. We have historically seen bias entering AI's in unforeseen ways and it would be naive to rule out completely in the future. The possibility of perpetuating or sustaining for example racial bias or gender bias that exist within the judiciary should be a serious concern and be taken into thorough consideration before implementation. The morally perverse outcomes of a judiciary that cannot rid itself of such biases because humans have fully or partially denounced their adjudicatory function in favour of more efficient proceedings, or however one decides to argue, is hard to justify. Although this grey dystopian view may not necessarily be the outcome we should not give up rights that were hard fought for without an assurance that the results will be better than the status quo.

The present capabilities of the technology behind AI does not live up to the requirements needed to create an AI adjudicator that could substitute human judges. In the not so distant future this might change and it would therefore be wise to weigh the possibilities against the liabilities. While an AI judge may live up to the purely institutional requirements put on a trial it is still up for debate whether impartiality and independence can be guaranteed and fairness achieved.

6.2 Artificial Justice

As discussed in chapter 5.4 there may be certain human aspects of adjudication that cannot be overlooked. Is it really appropriate to implement AI judges if we can determine that there lies a value in being judged by other humans? This is also up for debate, and according to my opinion this could also be challenged with the view that for example countries plagued by a corrupt law enforcement or judiciary may in fact benefit from implementing AI judges by removing the human element from the adjudication. Additionally it could be argued that if AI judges rule the lower instances by efficiently working through heavy caseloads the finer aspects of adjudication and interpretation of the law could be left to the higher instances where humans still preside. This would most likely also be the most compliant with the right of access to court. The argument however relies on the fact that one has the possibility to appeal as well as the funds to finance proceedings in higher instances. The effect on the nature of litigation this kind of process would have is hard to foresee.

As for the public's view of the judiciary, i.e. the perceived fairness, it will be important to ensure that AI judges can be sufficiently transparent. Transparency builds trust which is of utmost importance to the courts. As mentioned the precise level of transparency may have to be determined on a case by case basis to actually be meaningful. The important point being that it may not be acceptable to have a "black-box" judge with little or no insight in its reasoning when it is something that we have created that cannot be held to the same accountability standards as our human judges. In the end it is almost certain that the public's view of the judiciary will be affected by an implementation of AI judges. If this view turns out to be positive or negative depends highly on the implementation. However one would be wise to not blindly trust the implementation of an AI judge that lacks transparency as we have little reassurance that impartiality or independence is upheld.

It may also be difficult to justify the concept of adjudication if we decide to remove the human element from the judgement. A judgement could be thought of as a statement of what the humans in a society think is fair or right. If an AI produces a judgement it could be argued to no longer be what the humans think is fair but instead be what the AI has conceptualised as society's view on fairness and righteousness. In my opinion these

two concepts may not necessarily be the same even if the result may be equal. How we justify adjudication has a certain importance.

An AI judge could undermine the right to a fair trial if the implementations lacks the aforementioned qualities that we desire in a judge. In spite of AI adjudicators being able to live up to purely institutional requirements it may still be inappropriate to substitute human judges without a possibility to appeal to a human judge in a higher instance.

6.3 Steps to Make AI Judges a Reality

In order for AI judges to be trusted they would need to be sufficiently transparent. The transparency is the backbone of the trust in the judiciary as well as the ability to examine if the AI holds any bias or is susceptible to outside pressure.

Furthermore there would need to be assurances made that the AI is sufficiently secure. This could be achieved via third party audits of the source code or by making the code of the AI judge open source, i.e. accessible by anyone to audit. The latter run the risk of openly broadcasting exploits in the code for a malicious use. At the same time it is a much higher reward as the code could be peer reviewed and kept up to date with the public as a watchdog. The former alternative may not leave the AI open to exploitation in the same way, however it does not leave the public with a total assurance that the judge is in fact sufficiently transparent, impartial and independent. It may also be up for debate if the ECtHR would accept such a solution unless there are sufficient guarantees in place.

Additionally the AI judge should have to abide by some ethical standards. The Council of Europe has through its judicial body CEPEJ adopted an ethical charter that is meant to be used as a guiding document.[149] This could ensure that developers, legislators and the government take necessary precautions to allow AI judges to operate ethically and in compliance with the ECHR.

Finally there would need to be legislative measures made to ensure that AI judges can be properly integrated into a system that relies on humans. Adopting regulation for the appointment and terms of office that better reflect how an AI judge would be appointed.

[149] See CEPEJ, *Ethical Charter*, pp. 5-7.

6.4 Conclusion

The purpose of this thesis has been to analyse the use of artificial intelligence as adjudicators based on the provisions of Article 6 ECHR in order to evaluate its suitability in courts. To fulfil this purpose the questions posed in chapter 1.2 must be answered.

To answer the initial question, an AI judge could undermine the right to a fair trial in different ways. The two most prominent ways would be the impartiality and independence of the court. Due to the technology behind AI there may be inherent properties that allow bias to enter the decisions which would violate the provisions of Article 6 ECHR. This may have detrimental effects and could hit already vulnerable groups hard as well as damage the reputation and confidence in the courts. The independence will also always be able to be questioned if the AI is developed by a single entity such as the state or by a large company. Both of these problems may have technical solutions but can never be overlooked.

The second question, if AI judges can live up to the requirements of Article 6 ECHR, is therefore dependent on the previous answer. An AI judge could live up to the institutional requirements with proper legislative implementation as long as the issues of impartiality and independence is solved. Article 6 is a complex provision of ECHR that in many ways overlaps itself as the prerequisites of a fair trial are almost never completely separate. The fact that it is an overall picture that decides if a trial has been fair also points in this direction. The issue at hand for AI researchers and developers to solve seems to be transparency. If a sufficiently transparent AI is developed then issues of impartiality or independence could be rebutted.

The two final questions regarding the public's view of justice as well as if AI judges are suitable from a perspective of an appearance of fairness can be answered both in favour and against an AI judge. An impartial and independent AI judge, for example presiding in an AI court with the possibility to appeal to a "human court", i.e. a regular appellate court, could be seen as fair and may receive positive praise from the public if it shortens the time of the proceedings As for criminal proceedings it is up for debate if it is appropriate for another intelligence than humankind to administer criminal justice and for us give up such sovereignty. If the AI in retrospect proves to be biased then the confidence in the courts would most likely crumble if innocent people have been convicted. The same would be true if an AI judge lacks necessary transparency that results in a feeling of

helplessness or frustration. As an AI cannot be held accountable by the same standards as human judges it may be so that certain cases are less suitable for AI adjudication. Thus it seems to be some human element to adjudication that is not easily dismissed or replaced by artificial adjudication. The reception of an AI adjudicator would be highly dependent on its implementation and whether or not it has overcome the issues of transparency, impartiality and independence. In my opinion it seems difficult to overlook the positive effects that AI judges could bring, such as cost effectiveness, possibly shorter procedures and maybe even true impartiality if trained properly. The possible downsides are however grim enough that being completely optimistic may seem naive. If we cannot guarantee that the implementation of AI judges will lead to *better* results with less bias and risk of outside pressure then we should not too hastily abandon the current system. Therefore an alternative may be AI special courts that are subordinate to our appellate courts to evaluate AI judges without risking the trust of the judiciary as a whole.

Whether we support it or not, as history shows the technology will improve with time and with time and sufficient development there will most likely be further achievements that could eliminate the present concerns. If the technology then complies with our view on fair trials it will be up to us to decide if there really is an inherent human element of adjudication that makes it inappropriate for humans to be judged by others than their own peers whether it be in our present judicial system or in established special courts.

Bibliography

Council of Europe Publications

Council of Europe, The Committee of Experts on Internet Intermediaries (MSI-NET), *Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*, March 2018 Study DGI(2017)12, available at https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5 (accessed 2020-08-02) [Cit: MSI-NET, *Algorithms and Human Rights*]

Council of Europe, European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018), February 2019, available at https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (accessed 2020-08-02) [Cit: CEPEJ, *Ethical Charter*]

Council of Europe, Council of Europe Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, 19 May 2019, available at https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 (accessed 2020-08-02) [Cit: CoE Commissioner for Human Rights, *Unboxing Artificial Intelligence*]

Literature

Barkhuysen, T., van Emmerik, M., Jansen, O. & Fedorova, M. (4th ed. revised by van Dijk, P. & Viering, M.), *Right to a Fair Trial (Article 6)*, In: van Dijk, P., van Hoof, F., van Rijn, A., Zwaak, L. (Eds.), *Theory and Practice of the European Convention on Human Rights*, pp. 497-654, 5th edition, Intersentia 2018 [Cit: Barkhuysen et al]

Bolukbasi, T., Chang, K-W., Zou, J., Saligrama, V., Kalai, A., *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings,* Conference on Neural Information Processing Systems 2016 (NIPS 2016), available at

https://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf (accessed 2020-06-28) [Cit: Bolukbasi et al]

Collins, H., *Artifictional Intelligence : against humanity's surrender to computers*, Polity Press 2018

Coppin, B., *Artificial Intelligence Illuminated*, Jones and Bartlett Publishers 2004

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., Thrun, S., *Dermatologist-level classification of skin cancer with deep neural networks*, Nature 2017, pp. 115-118, vol. 542 [Cit: Esteva et al]

Fellner, J., *Race, Drugs, and Law Enforcement in the United States*, Stanford Law & Policy Review 2009, pp. 257-291, vol. 20, issue 2

Goodfellow, I., Bengio, Y., Courville, A., *Deep Learning*, MIT Press 2016 [Cit: Goodfellow et al]

Harris, D. J., O´Boyle, M., Bates, E. P. & Buckley, C. M., *Harris, O´Boyle and Warbrick: Law of the European Convention on Human Rights*, 4th edition, Oxford University Press 2018 [Cit: Harris, O'Boyle et al]

Kleineman, J., *'Rättsdogmatisk metod*' in: Nääv, M., Zamboni, M., (Eds.), *Juridisk Metodlära*, pp. 21-46 , edition 2:2, Studentlitteratur 2018 [Cit: Kleineman]

Kurzweil, R., *The singularity is near: When Humans Transcend Biology*, Viking 2005

Krizhevsky, A., Sutskever, I., Hinton, G.E., *ImageNet Classification with Deep Convolutional Neural Networks*, Conference on Neural Information Processing Systems 2012 (NIPS 2012), available at https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf (accessed 2020-06-28) [Cit: Krizhevsky et al, *ImageNet Classification with Deep Convolutional Neural Networks*]

McCarty, L.T., *Finding the right balance in artificial intelligence and law*, In: Barfield, W., Pagallo, U. (Eds.), *Research Handbook on the Law of Artificial Intelligence*, pp. 55-87, Edward Elgar Publishing 2018 [Cit: McCarty]

Montavon, G., Binder, A., Lapuschkin, S., Samek, W., Müller, K-R., *Layer-Wise Relevancy Propagation: An Overview*, In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, pp. 5-22, Springer 2019 [Cit: Montavon et al]

Russell, S., Dewey, D., Tegmark, M., *Research Priorities for Robust and Beneficial Artificial Intelligence*, AI Magazine, pp. 105-114, vol. 36, no. 4 [Cit: Russel, Dewey & Tegmark, *Research Priorities for Robust and Beneficial Artificial Intelligence*]

Russell, S.J., Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edition, Pearson 2010 [Cit: Russell & Norvig]

Samek, W., Müller, K-R., *Towards Explainable Artificial Intelligence*, In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, pp. 5-22, Springer 2019 [Cit: Samek & Müller]

Schabas, W. A., *The European Convention on Human Rights: A Commentary*, Oxford University Press 2015

Scherer, M., *International Arbitration 3.0 -How Artificial Intelligence Will Change Dispute Resolution,* In: Klausegger, C., Klein, P., Kremslehner, F., Petsche, A., Pitkowitz, N., Weiser, I., Zeiler, G. (Eds.), *Austrian Yearbook on International Arbitration 2019,* pp. 503-514, Verlag C.H. Beck 2019

Skansi, S, *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*, Springer 2018

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., Collins, J. J., *A Deep Learning Approach to Antibiotic Discovery*, Cell 2020, pp. 688–702, vol. 180, issue 4 [Cit: Stokes et al, *A Deep Learning Approach to Antibiotic Discovery*]

Weller, A., *Transparency: Motivations and Challenges*, In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, pp. 23-40, Springer 2019 [Cit: Weller]

Table of Cases From the European Court of Human Rights

Airey v. Ireland, 9 October 1979, appl. no. 6289/73

Albert and Le Compte v. Belgium (Article 50), 24 October 1983, appl. no. 7299/75; 7496/76

Ashingdane v. the United Kingdom, 28 May 1985, appl. no 8225/78

Belilos v. Switzerland, 29 April 1988, appl. no. 10328/83

Benthem v. the Netherlands, 23 October 1985, appl. no. 8848/80

Campbell and Fell v. the United Kingdom, 28 June 1984, appl. no. 7819/77; 7878/77

Cooper v. the United Kingdom, 16 December 2003, appl. no. 48843/99

De Cubber v. Belgium, 26 October 1984, appl. no. 9186/80

Delcourt v. Belgium, 17 January 1970, appl. no. 2689/65

Deweer v. Belgium, 27 February 1980, appl. no. 6903/75

DMD Group, A.S., v. Slovakia, 5 October 2010, appl. no. 19334/03

Fey v. Austria, 24 February 1993, appl. no. 14396/88

Findlay v. the United Kingdom, 25 February 1997, appl. no. 22107/93

Flux v. Moldova (No. 2), 3 July 2007, appl. no. 31001/03

Garcia Ruiz v. Spain, 21 January 1999, appl. no. 30544/96

Golder v. the United Kingdom, 21 February 1975, appl. no. 4451/70

Gregory v. the United Kingdom, 25 February 1997, appl. no. 111/1995

Kleyn And Others V. The Netherlands, 6 May 2003, appl. no. 39343/98; 39651/98; 43147/98; 46664/99

Kostovski v. the Netherlands, 20 November 1989, appl. no. 11454/85

Kraska v. Switzerland, 19 April 1993, appl. no. 13942/88

Kyprianou v. Cyprus, 15 December 2005, appl. no. 73797/01

Langborger v. Sweden, 22 June 1989, appl. no. 11179/84

Le Compte, Van Leuven and De Meyere v. Belgium, 23 June 1981, appl. no. 6878/75; 7238/75

Lupeni Greek Catholic Parish and Others v. Romania, 19 May 2015, appl. no. 76943/11

Marini v. Albania, 18 December 2007, appl. no. 3738/02

Micallef v. Malta, 15 October 2009, appl. no. 17056/06

Morice v. France, 23 April 2015, appl. no. 29369/10

Oleksandr Volkov v. Ukraine, 9 January 2013, appl. no. 21722/11

Pabla Ky v. Finland, 22 June 2006, appl. no. 47221/99

Pasquini v. San Marino, 2 May 2019, appl. no. 50956/16

Pescador Valero v. Spain, 17 June 2003, appl. no. 62435/00

Pétur Thór Sigurðsson v. Iceland, 10 April 2013, appl. no. 39731/98

Ramos Nunes de Carvalho e Sá v. Portugal, 6 November 2018, appl. no. 55391/13, 57728/13 and 74041/13

Ruiz Torija v. Spain, 9 December 1994, appl. no. 18390/91

Sander v. the United Kingdom, 9 May 2000, appl. no. 34129/96

Sokurenko and Strygun v. Ukraine, 20 July 2006, appl. no. 29458/04; 29465/04

Sovtransavto Holding v. Ukraine, 25 July 2002, appl. no. 48553/99

Sramek v. Austria, 22 October 1984, appl. no. 8790/79

Tinnelly & Sons Ltd and Others and McElduff and Others v. the United Kingdom, 10 July 1998, appl. no. 20390/92; 21322/92

V. v. the United Kingdom, 16 December 1996, appl. no. 24888/94

Wettstein v. Switzerland, 21 December 2000, appl. no. 33958/96

Internet sources

Alba, D., *A.C.L.U. Accuses Clearview AI of Privacy 'Nightmare Scenario'*, the New York Times 2020-05-28 (accessed 2020-07-27), https://www.nytimes.com/2020/05/28/technology/clearview-ai-privacy-lawsuit.html

AstraZeneca, *Press Release: AstraZeneca starts artificial intelligence collaboration to accelerate drug discovery,* AstraZeneca, 2019-04-30 (accessed 2020-06-20), https://www.astrazeneca.com/media-centre/press-releases/2019/astrazeneca-starts-artificial-intelligence-collaboration-to-accelerate-drug-discovery-30042019.html

Cortes, C., Jackel, L. D., Chiang, W-P., *Limits on Learning Machine Accuracy Imposed by Data Quality*, AAAI Press 1995 (accessed 2020-06-28), https://aaaipress.org/Papers/KDD/1995/KDD95-007.pdf

Dastin, J., *Amazon scraps secret AI recruiting tool that showed bias against women*, Reuters 2018-10-10 (accessed 2020-06-18), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

DeepMind, *AlphaGo*, DeepMind Technologies (accessed 2020-07-13), https://deepmind.com/research/case-studies/alphago-the-story-so-far [Cit: DeepMind, *AlphaGo*]

Hao, K., *Facebook's ad-serving algorithm discriminates by gender and race*, MIT Technology Review 2019-04-05 (accessed 2020-06-16), https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/

Kayser-Bril, N., *Google apologizes after its Vision AI produced racist results*, AlgorithmWatch 2020-04-17 (accessed 2020-05-19), https://algorithmwatch.org/en/story/google-vision-racism/

Kuchler, H., *Pharma groups combine to promote drug discovery with AI*, Financial Times 2019-06-04 (accessed 2020-07-19), https://www.ft.com/content/ef7be832-86d0-11e9-a028-86cea8523dc2

Larson, J., Mattu, S., Kirchner, L., Angwin, J., *How We Analyzed the COMPAS Recidivism Algorithm*, ProPublica 2016-05-23 (accessed 2020-07-18), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [Cit: Larson et al]

Markoff, J., *Computer Wins on 'Jeopardy!': Trivial, It's Not*, the New York Times 2011-02-16 (accessed 2020-07-13), https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html

Racism is Poisoning Online Ad Delivery, Says Harvard Professor, MIT Technology Review 2013-02-04 (accessed 2020-07-12), https://www.technologyreview.com/2013/02/04/253879/racism-is-poisoning-online-ad-delivery-says-harvard-professor/

Russell, S., Dietterich, T., Horvitz, E., Selman, B., Rossi, F. *(and over 8,000 signatories)*, *Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter*, Future of Life Institute 2015-01-15 (accessed 2020-06-10), https://futureoflife.org/ai-open-letter/ [Cit: Russell, Dietterich et al, *Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter*]

Smith, C., *Facebook Users Are Uploading 350 Million New Photos Each Day*, Business Insider 2013-09-18 (accessed 2020-05-12), https://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9?r=US&IR=T

Toews, R., *AI Will Transform The Field Of Law*, Forbes 2019-12-19 (accessed 2020-06-20), https://www.forbes.com/sites/robtoews/2019/12/19/ai-will-transform-the-field-of-law/#481bfe837f01

Yong, E., *A Popular Algorithm Is No Better at Predicting Crimes Than Random People*, The Atlantic 2018-01-17 (accessed 2020-06-24), https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/