# The use of Artificial Intelligence in the Dutch courtroom

# The use of Artificial Intelligence in the Dutch Courtroom

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

## MASTER OF SCIENCE

## in **Engineering and Policy Analysis**

Faculty of Technology, Policy and Management

By

Ibrahim Jabri MSc LLM

Student number: 4291034

To be defended in public on October 25, 2022

## Graduation committee

**Chairperson:** prof.mr.dr. J.A. de Bruijn, Organisation and Governance
**First Supervisor:** mr. J.M. Kooijman, Organisation and Governance
**Second Supervisor:** dr. A.M.G. Zuiderwijk- Van Eijk, Information and Communication Technology

a ro·bot ju·dge (*noun*): a machine who is in charge of a trial in a court and decides how a person who is guilty of a crime should be punished or who makes decision on legal matters.

# Preface

Before you lies the master thesis *"The use of Artificial Intelligence in the Dutch courtroom"*. It has been written to fulfil the requirements to obtain the Master in Engineering and Policy Analysis at the University of Technology in Delft.

It was summer 2019 when I asked myself the question: *what now?* At that time, I was finishing my masters in Tax Law, Data Science and Marketing Analytics, and Economics of Taxation at Erasmus University Rotterdam. Even though my journey in Rotterdam has always been fulfilling, it felt like it shouldn't just stop there.

A lot of thought processes further and I decided to enrol in the master's in Engineering and Policy Analysis. Hadn't it been for my background in data science, I wouldn't be here. And what a choice it was. The different courses in the master, the international students and even the campus at The Hague: it was all the fresh wind I needed.

Whilst I was working two jobs, as a lecturer at Erasmus University and a Process Controller at Rabobank, it did not at all feel like it was *too* much. While writing this thesis, I got the chance to pursue a new opportunity in the Middle East. The evenings and weekends were packed with writing the thesis, and even my working hours weren't spared.

It is therefore why I am so grateful everything in the end worked out just fine. I couldn't have achieved this without the support, feedback and flexibility of my committee – this product is also the results of your continuous support.

Friends, family and colleagues, whom not all know I am still studying: the videocalls and chats were just what I needed to have some time off and recharge.

To you, the reader, I hope you enjoy this thesis as much as I did writing it

Ibrahim Jabri
Utrecht, October 24, 2022

# Abstract

With the rise of data, many sectors have been implementing a form of automatising in recent times. This brings a lot of possibilities. As a matter of fact, an automated system is – unlike the human brain – not dependent on food nor sleep and less prone to errors. The efficiencies that could be gained by 'working' 24 hours, 7 days a week without a single moment of pause is enormous. An Artificial Intelligence system (hereinafter: AI system) could provide a relief for many sectors. In the literature, the application for an AI system within the judicial system has been advocated for. Nevertheless, the judicial system, especially regarding the rights of a suspect, needs many more considerations than the workload-relief argument alone. In this thesis, the possible implementation of AI – or a robot judge – within the Dutch judicial system will be explored. In this research, a special focus will be given to the rights of a suspect as set out in article 6 of the European Convention on Human Rights (hereinafter: ECHR). Article 6 ECHR provides the right of a fair trial to a suspect, where relevant, with the following provision: *"In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law"*. From clause 1 of article 6 ECHR, the following components can be extracted: a fair, independent and impartial tribunal. Implicitly, article 6 ECHR also provides the right to an explanation on how this tribunal reached to its verdict. In practice, this is also referred to as the right to transparent verdicts. With the use of the research question *"to what extent can Artificial Intelligence be used in the Dutch courtroom while still maintaining the rights for the parties involved, especially as how has been set out in article 6 ECHR?",* an answer is formulated on the question whether the Dutch judicial system is ready for – and capable of – the next judge: a robot judge. This research uses a systematic literature review as a starting point. With such review, information is gathered from both technical as well as legal point of views to answer each sub-question individually. This research also specifically looks at the way other countries have implemented or created a form of an AI tool *in* and *as* court. Next to this, four interviews have been held with potential stakeholders as well as persons of interest. The reason why these interviews are held is to confirm and check what has been found from the literature

assessment as well as to gain new (internal) insights on this topic. As an answer to this research question, an AI system can be implemented in the Dutch courtroom if it consists of two (main) components: a linguistic system and a machine learning system. Additional rules to use this robot judge are proposed as well, for example that continuous feedback-loops (of the probabilities for example) should be built in and that the system should be continuously monitored (and adjusted) for potential biases. The robot judge can both function *as* and *in* court. Additionally, a robot judge can sit at the table of the Dutch Council for the Judiciary and their ambitions to (further) litigate digitally and publish more, if not all, cases online.

*Keywords:* Artificial Intelligence, Dutch judicial system, article 6 ECHR, robot judge

# Table of Contents

# Chapter 1: Introduction

## 1.1. Introduction to the problem

Recently, there has been a rising discussion on the implementation of Artificial Intelligence (hereinafter: AI) in the courtroom (Sourdin, 2018). In a period of growing suspicion towards the capabilities of digital technologies and tech-companies, the judicial system aspires to take a step further in the automation of work processes surrounding legal procedures in criminal law.[1] A robot judge that automates the decision-making process in court could be reality in the near future. This robot judge can be used to gather evidence, examine patterns in large amounts of data, and provide (relevant) case studies (Katz, 2013). All this could eventually be used to provide verdicts: the robot judge becoming reality.

Potential positive impacts can be achieved when implementing AI. As an example, AI could decrease the subjectivity of a judge's decision. In a recent study, parole hearings in criminal cases by judges in Israel have been investigated (Bryant, 2011). This research found out that judges gave more lenient decisions at the start of the day and after a lunch break, possibly due to a so-called glucose depletion. After each meal, approval rates spiked reaching 65%, whereas on average only 35% of the convicted were granted parole. Since a robot judge is not exposed to a glucose depletion, a protentional positive impact can be achieved.

However, past the promising possibilities of the AI judge, concerns are raised regarding the fairness of (criminal) *robotic* trials (Deeks, 2019). Most fundamental rights European citizens have, are established in the European Convention on Human Rights (hereinafter: ECHR). Article 6 of the ECHR states that every citizen, who has been charged with a criminal offence, has the right to a fair trial. Nevertheless, it is this *fair* trial that seems to form a crucial clash with the use of AI

---

[1] See for example in this context the current discussion that is held on the use of data of big tech companies and how they have been using (and might even misuse) data of users, source: https://www.cnet.com/news/privacy/facebook-struggles-to-control-all-the-data-it-gathers-report-says/.

in the courtroom (Schermer & Oerlemans, 2020). As AI systems lack any actual knowledge of the juridical system and criminal law, AI judges cannot make juridical nor moral considerations (Prins & Roest, 2018). AI systems do not feel or care, so whenever they would make an inequitable or morally unjust decision, they would not be conscious of it. Besides, while at first glance it seems that AI systems could possibly increase objectivity by ignoring factors such as gender and race, AI systems may learn and mimic bias from their human inventors or the data they have been trained with. This will eventually only increment the bias that is already present. Ruha Benjamin, an associate professor at Princeton University, states that machine-learning algorithms are trained on *"data produced through histories of exclusion and discrimination"* (Benjamin, 2019).

Next to this, studies have shown that humans have the tendency to over-trust automated decision-making systems (Hao & Stray, 2020). The trust in AI systems appeared to be so significant, that people in the studies followed a robot's advice even though they knew the robot had made mistakes in the past. Lastly, many AI systems, in particular those based on machine learning, lack transparency (Prakken, 2018). Their algorithm based decision-making is too complex to provide explanations on why a certain decision has been taken. Usually, the most accurate AI models are not very explainable and the AI models that are better interpretable are less accurate (Gulum, 2021). And in the end, when sentenced to imprisonment, the charged criminal wants to know on which grounds. Therefore, research should be conducted on how AI could and should replace the human judge in the courtroom – if it should have a seat in the courtroom in the first place.

## 1.2. Problem statement

As has been stated previously, a potential implementation of AI faces positive effects as well as challenges. AI is – other than human judges – not limited to the amount of time or workload. In addition to this, AI uses (collected) data on which a sentence is given. This benefits the legal certainty – one of the corner stones of law (practices) – as a verdict will be similar to previous situation(s) (Casey & Niblett, 2020).

On the contrary, AI as a judge is still a new technology. Therefore, a plan of action or set of criteria should be given in order to provide a 'working' AI judge, which will fit within the judicial system and its objectives. Special attention should be given to the rights as provided in the ECHR – mostly article 6 ECHR. This thesis will try to find a way in which AI can be used as a judge while still maintaining the rights a citizen has as set out in the ECHR.[2] This research will be done using the following research question:

> *To what extent can Artificial Intelligence be used in the Dutch courtroom while still maintaining the rights for the parties involved, especially as how has been set out in article 6 ECHR?*

## 1.3. Research method

To answer this research question, a qualitative approach will be followed. This research tries to evaluate the (possible) implementation of AI in the Dutch courtroom. In addition, the objective is to show the effects of this policy on the Dutch judicial system and to provide recommendations on how to implement AI in the courtroom while maintaining the rights a suspect has as specified in article 6 ECHR. Since AI is not yet implemented, this leaves room for a framework or a guideline.

In order to conduct this research, the following sub-questions have been formulated. These sub-questions will organise this research in a structured way; every sub-question will be answered in a chapter.

1) From a technical perspective, what are the promising potentials of AI and a robot judge within the Dutch courtroom?
2) From a juridical perspective, how can AI and a robot judge contribute to an improvement within the Dutch courtroom?
3) What challenges will an AI system within the Dutch courtroom face from a technical perspective?

---

[2] Even though this thesis uses examples from other countries, the research focuses on providing a action plan for Dutch courtrooms and the Dutch Council for the Judiciary.

4) What challenges will an AI system face within the Dutch courtroom from a judicial perspective, also considering the rights of a suspect as set out in article 6 ECHR?

5) On the bases of which requirements should an AI system be developed for a successful implementation within the Dutch courtroom?

The starting point for this research is the current information as drawn from the literature. Even though this research is specifically focussed on the Dutch courtroom, information is taken from multiple (technical) journals across the worlds and not necessarily restricted to a Dutch perspective. Using a systematic literature review, information is gathered from both technical as well as legal resources to answer each sub-question individually. For this review, the literature is extracted by conducting manual searches through publicly available sources such as Google Scholar and textbooks within the field of artificial intelligence, automated decision-making and AI in the courtroom. Next to this, four interviews have been held with potential stakeholders as well as persons of interest. The reason why these interviews are held is to confirm and check what has been found from the literature assessment as well as to gain new (internal) insights on this topic. A lot of times, it is the practical insights that cannot be found within the literature as to why interviews were necessary for this research. The interviews are conducted with the following organisations (in their order of conducting): a Professor in Legal and Data at Erasmus School of Law, the Dutch Council for the Judiciary, the WRR and lastly, a company that built an AI model to predict court outcomes. The interviewees are assigned as such so a diverse range of stakeholders and persons of interest are considered. Whenever specific information within the chapters is taken from the interviews, this will be mentioned explicitly.

To structure this research, two forms of AI in the courtroom are investigated. The first form is AI *in* court. In this form, AI systems will assist in the courtroom (such as research, comparison between courthouses and drafting verdicts) but will not rule in the case itself. This role will be solely given to human judges. The second form is AI *as* court. In this form, AI will do all the potentials it can perform under form 1 but also rule in the underlying case. As such, this process does not require an active role for the human judge.

## 1.4. Prospective results and drawbacks

As a prospective result, the belief is that AI can be implemented in the Dutch courtroom and still be able to guarantee the rights of the suspect – if and only if these rights considered when building the system (Chen et al., 2021). As a matter of fact, the implementation of AI seems promising because it removes subjectivity of a judge that might currently still be in the system. However, on the downside, a robot judge might be based on any subjectivity that exists within current court verdicts. This is because a robot judge searches for patterns in previous verdicts (Campbell, 2020). Some form of checks and balances should be implemented to minimise or remove this subjectivity.

On the other hand, this research is only limited to the Dutch judicial system. A possible answer that could follow from this research is that under current circumstances, AI cannot be implemented because it cannot guarantee the rights of a suspect. Even in that case, this will be an answer solely based on the Dutch judicial system.

## 1.5. Relevance

### 1.5.1. Academic relevance

AI is a concept that has been widely accepted from a theoretical perspective and is even used in a lot of fields as of today to assist humans.[3] The possibilities are extensive, from predicting whether to invest within the financial markets to the predictions if a person has, or is at risk of getting, a certain disease (Dunis et al., 2016; Reddy et al., 2019). The use of AI within a judicial system is also argued for as it could potentially increase legal equality, decrease biases and provide better reasoning. However, most of this research is conducted within the US or China, countries which are generally leading in terms of the possibilities of AI. In the Netherlands, some research has been conducted on the implementation of AI in the

---

[3] For example, AI is used in education (Marr, 2018), the field of medicine (Kabir, 2019), COVID-19 detection (Chen, 2020) and even police enforcement (Greene, 2018).

courtroom. This research mainly focussed on the limitations of AI as to why it could never replace the human judge. In this context, see for example research from Prakken (2018) and Sourdin (2018). However, those studies do not approach the AI judge from a 'positive' perspective. In this research, another approach will be followed. From the belief that AI could play a role in the Dutch courtroom, research is conducted on what should be taken into account for a successful implementation and which role this AI judge could have. This research therefore tries to contribute to the current knowledge by providing a list of criteria that should be met to ensure a proper implementation, considering the perspectives from different stakeholders as extracted from the interviews conducted.

### 1.5.2. Societal relevance

The Dutch judicial system has always been open for ways to improve the quality of law within the Netherlands. Examples of these initiatives are the project Quality and Innovation or the current projects regarding filing cases digitally (Van Duijneveldt, 2017). As this research is focused on providing criteria for a successful implementation of AI, the Dutch judicial system would benefit as it can be included in their ambition to upskill the system from a digital perspective. However, the benefits can be much wider than previously mentioned. Any person could use the foundations of this research to build an AI judge and present this to the Dutch judicial system.

Additionally, the purpose of AI law is to provide better law than the current status quo. Therefore, any party involved (e.g. a suspect or the State) will benefit from the correct implementation as the quality of law will be improved. Likewise, this research also gives room for future research on bigger possibilities of digitalisation in general other than providing case law (e.g. comparing quality of law and researching effectivity of clauses of law). This will not only benefit the previously mentioned parties but (the Dutch) society in general.

### 1.5.3. Relevance within the EPA-program

This research focuses on the possible implementation of AI in the Dutch public domain. In this specific case, AI provides potential possibilities for actors within the

Dutch judicial system. These actors include the Dutch judicial system represented by the Council for the Judiciary (Dutch: *de Raad voor de Rechtspraak*). As such, this research tries to find a bridge between AI on the one hand and the Dutch judicial system on the other. Because this system has not been implemented (yet), the goal is to propose guidelines on how the implementation of such system in Dutch courthouses. As such, this research tries to find an answer to the question whether AI can be implemented in the courtroom and if so, to what extent. This research combines a technical character with a legal perspective as well as a social relevance. By trying to find a guideline that considers the judicial system, in this case from a Dutch and European perspective, as well as a technical component, this research complies within the multidisciplinary character of the EPA-program. Therefore, this topic fits within the master program.

Next to this, this research builds on a paper that was written by the author (together with another student) in the course Political Decision-Making *(EPA1424)*. In this paper, the Dutch Public Prosecution Service was advised on the use of AI in the courtroom. This topic was found so interesting that it is used as the foundation for a research topic. The expected deliverable fits the purpose of the program therefore as well. The guideline that is expected to follow out of this research could be proposed to the problem owner: the Dutch Council of the Judiciary *or* any courthouse for that matter.

# Chapter 2: potentials from a technical perspective

## 2.1. Introduction

Even though computers and electronic systems have been widely accepted over the last 30 years, AI is a concept that was introduced even before the first computers were used. At the end of the second world war, Vannevar Bush was the first person to introduce the concept of AI (Bush, 1945):

> *"It is strange that the inventors of universal languages have not seized upon the idea of producing one which better fitted the technique for transmitting and recording speech. [...] Consider a future device for individual use, which is a sort of machinezed private file and library. It needs a name, and, to coin one at random, 'memex' will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement of his memory."*

Even though Bush did not mention the word *Artificial Intelligence* explicitly, his ideas set the first stone. In 1970, even before the widespread adaptation of computers and technological systems, the legal system showed interest in AI (Bobrow & Raphael, 1974). In that specific research, tools were developed that could perform processes that require some form of human thinking and rationale. Later, the Natural Language Processing was mentioned as (one of the most) beneficial in these judicial systems (Xiao et. Al., 2021). This tool can be used to predict an outcome based on textual input such as court rulings (Luo et al., 2017).

> ### *Natural Language Processing: a practical tool*
> *Natural Language Processing tool (or more general: text mining) is a method used to analyse (large) chunks of text and to generate patterns. The main advantage of this tool is that it does not only look at the type of words used, but also their position within sentences and context (Nadkarni et al., 2011). To stay within legal examples, the word "object" would have two different meaning in a legal court setting as compared to an art magazine.*

*The tool can analyse context and placing of words and would be able to recognize the different meanings of "I object" versus "nice object". This tool uses two types of analyses: syntactic and semantic analysis. In a syntactic analysis, words and phrases are analysed based on word combinations, order and grammatic rules. In semantic analysis, words and phrases are analyzed based on context and relationships within words and sentences (Locke et al., 2021, Chary et al., 2019). This might give the (false) impression that this tool seems like a technical tool only used by data scientists. Believe it or not, it is something most of us use (unaware): personal digital assistants, such as Alexa or Siri, are a combination of speech recognition and Natural Language Processing (Locke et al., 2021).*

In the literature, a lot of arguments have been proposed in favour of a role of AI within a judiciary system. In his book, Ethan Katsh stresses the increasing role of AI within judicial decision-making (Katsh, 1995). Ethan defines three main advantages for the use of judicial AI over a human judge. These three arguments, known as efficiency, experience and objective, will be analysed in the next paragraphs. Using a systematic literature review on the previous mentioned arguments, the findings from this first chapter will be used to answer the first sub-question: *from a technical perspective, what are the promising potentials of AI and a robot judge within the Dutch courtroom?*

## 2.2. Efficiency

The research of *Law in a digital world* stressed out the challenges that are faced within judicial systems: more cases and (relatively) less people (Katsh, 1995). Since then, the workload has increased drastically. As a result of this, cases are not given the time they should be given (or 'deserve') in the first place or even put aside for cases that have a higher urgency. This creates a backlog which is a crucial risk for any judicial system. To illustrate this crucial risk by means of an example, the current workload of the Dutch Council for the Judiciary is illustrated below. This example proves the need for drastic changes.

***The Dutch Council for the Judiciary***

*Within the Netherlands, the Council for the Judiciary (hereinafter: the Council) is responsible for the Dutch courthouses and is dedicated to ensuring that these courthouses can perform their tasks properly. Statutory duties are, among others, management, quality and integrity.[4] On the 20th of September 2013, the Council headlined on a high production and workload among judges after concerns had been raised. As a result, a deliberation took place on the initiative of the Dutch House of Parliaments. In this meeting, the conclusion was drawn that delivering quality has always been a top priority within the Dutch judicial system, and that this priority was under pressure due to the workload and the way the system works.[5] Within the judicial system, an allowance is received for every verdict produced, and vice versa. This loop paved the way for a higher production and workload. Next to this, the average case intensified over the years as cases got more diversified and smaller cases are to a bigger extent discharged in other ways such as mediation. The financing system based on output is partially responsible for the (exponential) heavy workload. Next to this, the judicial system faced modernisation with the implementation of the project "Quality and Innovation" (Dutch: Kwaliteit en Innovatie).[6] This pressurised the judicial system to a bigger extent. In 2016, 6 years later, Dutch news headlined with "Overtime normalised for years".[7] Claimed is that 40% of overtime is needed to operate at full speed. 2 years later, in 2021, Dutch news headlined (again) with a critical note to the so-called 'walk-in' courthouses, that allowed judges to rule in cases that had been postponed.[8] More importantly, the judicial system lacks 200 judges to fully operate as normal. Nowadays, this issue persists. On the 10th of May 2022, The Dutch Association for the Judiciary*

---

[4] Source: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Paginas/mva.aspx.

[5] Source: https://www.eerstekamer.nl/behandeling/20131113/debat_over_de_werkdruk_bij_de.

[6] Source: https://www.rijksoverheid.nl/onderwerpen/rechtspraak-en-geschiloplossing/vernieuwing-in-de-rechtspraak.

[7] In 60% of the cases, one of the parties involved is the government. The function of the judge is not only to ensure there is a balance between the government and citizens, sometimes this requires them to also provide counterbalance. As the workload increases, judges are less prepared sometimes. This puts the quality of cases (statistically) at risk, source: https://www.metronieuws.nl/in-het-nieuws/2019/03/werkdruk-rechters-enorm-overwerk-al-jaren-normaal/.

[8] Source: https://nos.nl/artikel/2382375-rechters-kritisch-over-inloopkamers-zaken-worden-afgeraffeld.

*addressed two points of issue to the Council.[9] First, a substantial decrease of overtime – at that time 33.6% – is requested. Next to this, it is requested to fully eliminate the overtime within 2 years. The question is whether this is theoretically possible, let alone practically considering the backlog that is already there. All in all, this means the Dutch judicial system needs 200 extra judges within 2 years – ceteris paribus. At most recent, on Thursday the 16th of June 2022, the Dutch newspaper NOS headlined saying that 1500 (light) cases will not be presented in front of a court due to a lack of judges.[10] If not yet before, now is the time to find a solid and efficient solution.*

As shown in the previous example, The Dutch judicial system is facing a high production and workload. This workload puts stress on both the quality of the work as well as the judges performing that work. To achieve the elimination of workload, change needs to happen. To achieve this change, *any* implementation of AI can contribute.[11] This will lead to a faster and better (in terms of quality) processing of cases – at least theoretically. Next to this, human judges can be used for the more difficult and intensive cases. Therefore, the implementation of *any* artificial system will be able to depressurise the current system, with the intention to maximise judicial resources (Zheng, 2020). The application of algorithms is numerous. In a less severe situation, AI will support judges by for example predicting the outcome in new cases or finding relevant case law. To a more severe extent, AI will replace the human judge, by filing the case, defining the legal context and ruling in the specific case.

It is worth mentioning that the project *Quality and Innovation* of the Council never succeeded, with a price tag of € 205,000,000.00. This emphasises the urgency and importance to come up with an outline to have a correct and successful implementation of AI, as mentioned in paragraph 1.3.[12] Within the context of the

---

[9] The Dutch Association for the Judiciary (*Dutch: Nederlandse Vereniging voor Rechtspraak*) addressed its concerns in a letter, source: https://nvvr.org/uploads/documenten/20220510-brief-Raad-voor-de-rechtspraak-Werdruk-TBO.pdf.
[10] Source: https://nos.nl/artikel/2432906-te-weinig-rechters-in-gelderland-1500-rechtszaken-geschrapt.
[11] As mentioned in paragraph 1.3, AI can take a form *in* court as well *as* court.
[12] As limited resources are scarce, it is worth mentioning that the same funds that could be used for the implementation of AI within the Dutch courtroom could also be assigned to hiring and training

judiciary system as explained in the previously mentioned example, two main factors can be identified: there are more cases and as such relatively less judges. In these sub-paragraphs, both will be analysed in the context of how AI will be able to provide a solution two both factors.

### 2.2.1. More cases: AI *as* court

One of the main issues within the Dutch judicial system is the fact that there is an exponential growth of cases filed at the courthouses. In addition, with relatively less judges available, these cases keep piling up. Adding delays due to the pandemic has resulted in an exponential number of cases unruled.[13] One of the interviews conducted for this research was with the Council. In this interview, one of the topics discussed was the recent developments on litigating digitally (Dutch: *Digitaal procederen*). Since 2018, the Council has focused on making sure documentation can be done digitally. In fact, as of November 15 2021, all Dutch courthouses accept digital litigation (Wees, 2022). In my opinion however, nothing is done with the possibility of this digital platform, other than the purposes for which is built. The sole purpose is that documents are sent digitally. However, to name a possibility; text mining system can be used to analyse these documents and do predictions on the outcomes. This does not necessarily mean that a whole predictive system should be built, rather a pilot that demonstrates the promising possibilities of AI.

As such, an AI-prediction when filing a case would be a solution. This works as follows: when litigating digitally, an AI-system can predict the chances of that specific case and give a prediction on the claim. Any person, mostly a lawyer, will be able to receive its chances and decide whether it should proceed with filing the case. These predictions can indicate the chances of the specific case for the filer and as such perform as an implicit second taught. This would, presumably, lead to less people filing a case (e.g. when a low prediction is given) and/or more people searching for other options of dispute resultion such as mediation (Bex, 2020). Nonetheless, these predictions should not stand in the way of filing a case at any

---

more judges. However, the implementation of AI, as is expected, is easier and faster to implement and more durable – at least in the long run.

[13] Part of the backlog is also due to cases that were filed during the pandemic lockdown and as such have been postponed to post-lockdown, source: https://nos.nl/artikel/2432964-geen-zitting-in-1500-zaken-maar-stapels-bij-rechtbank-torenhoog.

chance.[14] It is worth mentioning that an outcome, technically, might not be based on a prediction. In the situation where a comparable (or even identical) case was filed earlier, the outcome will be like the previous outcome – or at least that is expected.[15] In such case, no real 'prediction' was conducted. To fully understand the possible chances of an AI-system, the possible predictions will be explained. An AI-system can predict based on three factors: prediction based on non-content related factors, textual description of the case and prediction based on (actual) legal relevant factors.

### 2.2.1.1. Predictions based on non-content related factors

Some machine learning systems can predict based on non-content related factors of a case. In one research, an algorithm was created to predict in American cases that were filed at the American Supreme Court based on factors such as type of case, the date of filing and at which (lower) courthouse the (initial) case was filed (Katz et al., 2017). Other relevant factors could be the name of the lawyer that is filing the case. One can assume that some lawyers have a higher 'winning' rate. This specific algorithm predicted correctly in 70% of the cases. This means that in more than 2 out of 3 cases, the prediction was correct.

Questions can however be asked on the applicability of these predictions. It is trained on factors that are not judicially related. To which extent will such a system ever be able to replace a human judge? On the other side, such predictive systems might be able to spot other trends, such as differences in ruling between courthouses, or even patterns based on the filers last name or origin to name a few (more on biases in paragraph 3.3).[16] Such mechanisms, that practically would work as checks and balances, would benefit the quality of cases. It could work both post hoc as well as a prediction that could run simultaneously to the court proceedings. As part of article 6 ECHR, any party in court, except for the State, has the right to appeal for a

---

[14] Anyone has a right to file a case at a courthouse and to ask a judge to rule in specific case, for example based on article 6 ECHR. Therefore, even if the predictive system would result in a small chance in ruling in favour of the petitioner, access to a courthouse would still have to be granted.
[15] For example, Dutch law requires that an appeal to the courts should take place within 6 weeks after the decision from a Dutch governmental authority was made and notified. In case these six weeks have surpassed, any appeal will be inadmissible.
[16] In the US, some companies provide these predictions (at the level of courthouses) based on the judges, the lawyers or the parties involved (e.g. a State versus an individual). Examples of these commercialised algorithms are Lex Machina and Lex Predict (www.lexmachina.com respectively www.lexpredict.com).

substitution of the court (or a particular judge). The parties involved have the right to an independent and impartial judge. In case a party has the impression that this is not the case, a request to substitute the court can be filed. However, these requests are not very successful: in 2020, out of the 675 requests filed at a Dutch courthouse, only 17 times the court (or judge) was substituted.[17] This shows that it is very hard to successfully substitute a court. However, a prediction based on non-content related factors can provide insights and even function as proof for such requests. If a court is specifically sensitive to some non-content related characteristics, that could provide insights into the chances of cases even before one legal argument has been given. Therefore, even though the use as a judge is questionable, there are other ways in which these predictions can be applied as benefitting the quality of case law. From a behavioural perspective, however, there should be some form of protection that these requests to substitute a court will not be solely based on these predictions. One can imagine that otherwise, courthouses will be occupied by these requests. The fact that these predictions are available shouldn't pave the way for an exponential growth on requests for substitution. Therefore, it would be advisable that these predictions would only be available to a selected group of people (e.g. courthouses, judges and researchers).

### 2.2.1.2. Predictions based on textual factors

In another example, algorithms were trained to create patterns within textual factors of cases (Medvedeva, 2020). In this specific research, predictions were made based on cases of the European Court of Human Rights and whether a specific right from the ECHR was violated. The algorithm was trained to predict based on information on the case, such as facts and history of the case. The predictions had an accuracy of 75%. Even though this may seem like accurate predictions (compared to the first example), no relevant predictions were done. The predictions are merely done based on other circumstances and factors rather than 'real' legal arguments. The word *violation* had a high predictive value in combination with the word *the first applicant* just to name a relation. For a full overview of the relevant predictors of a

---

[17] From 2009 to 2012, the number of requests filed increased exponentially to approximately 600 appeals a year. From 2012 up until 2021, the number of requests varied between 600 and 750. Most of the requests are filed within the division of private law, source: https://jaarverslagrechtspraak.nl/wp-content/uploads/sites/2/2021/04/Jaarverslag-Rechtspraak-2020.pdf#page=66.

violation in this specific research, please refer to Appendix B – Figure 8. From Medvedeva's research, a predictive system was built on the judges and their relationship with a violation of an article from the ECHR. The salient conclusion was that it *matters* which judge would rule in a specific case on the conclusion of a violation (see appendix B – Figure 9). Such differences between judges are, first and foremost, an even bigger argument in favour of a robot judge. Additionally, like in the previous example of predictions, this type of system can increase the quality of case law even without having to decide in specific cases.

### 2.2.1.3. Predictions based on legal arguments

The last form of predictions are predictions based on actual legal arguments, and from a perspective of producing case law most interesting. In an American research, a system was created to predict in cases of misuse of company secrets. In this system (Ashley et al., 2009), manual words had been entered to create a prediction. Words like 'non-disclosure agreement' or 'unique product' were relevant legal arguments on which predictions were based. These factors were compared to previous cases and based on that a prediction was made whether a verdict of misuse of company secrets was given. These predictions had an accuracy of 82% and 88%. It goes without saying that these accuracies are high, especially compared to the previous examples. Connecting this to the exponential increase in cases, the predictions could also be done before trial (for example when filing the case digitally). It will function as an extra moment of 'assessing chances' which could result in related parties seeking other ways of solving the underlying problem. This on its turn could lead to less cases. In other words, the use of this predictive system does not require an extreme AI tool to begin with. On the other hand, from a behavioural perspective, this could also lead to an adverse reaction. If the 'winning chance' was low, parties could continue filing anyhow with the reason that nothing is lost. In a lot of cases, the reason parties go to court is not to win, rather it is to be heard (Katyal, 1998).

### 2.2.2. Less people: AI *in* court

All courthouses within the Netherlands have offices who are responsible for collecting and storing files to support the judge's decision-making. Examples of jobs

conducted by these offices are collecting and sorting case law, with the relevant job of finding precedents. This can be, as one might expect, a time-consuming task.

In the previous paragraph, predictive systems have been explained that could function as AI *as* court. The possibilities of AI are nevertheless more than just predicting. AI can also function *in* court as to replace the work of humans. This would lead to a more effective use of scarce resources (i.e. people). As a matter of fact, Corien Prins shares the believe that in a lot of research, the focus is on how AI can act *as* court while that might still be a road down too far (Rechtstreeks, 2019).[18] In the interview with the Council, I have found a similar focus: the feeling is suggested that it is an all or nothing when it comes to the implementation of the robot judge. Rather, more focus should be given to the functions that are ancillary to the actual deciding in a case (Tai, 2018).

To understand the possibilities, below are different classifications in which an AI system can be of use (Bex, 2019):

- Researching (including benchmarking) documents fast and in a smart way;
- Analysis of the current status quo given the underlying legal problem; and
- Support and decide in cases.

To name a few concrete examples, AI can (Rechtstreeks, 2019, p. 68-69):

- Assist in helping case law to be published online. One of the (ongoing) pillars of the Dutch Judiciary is to publish more cases online. In the interview with the Council, the percentage 4 is mentioned: as (roughly) only 4% is published online, there is room for improvement. AI can anonymise texts and publish these afterwards or even assist in writing these case laws (similarly as how any voice assistant such as Siri responds to spoken words);

- Provide insights into a consistent use of law. Within criminal law, insights into the use of law can be relevant, also in terms of effectivity of different clauses. Just to name an example, until today, there is no clear insight into what the 'average' given punishment is of theft ex article 310 Dutch Criminal Law;

---

[18] prof.mr. Corien Prins is a Professor at Tilburg University and chair of the Scientific Council of Government Policy (Dutch: *WRR*).

- In addition to previous mentioned, what are the effects of new laws or changes in current law? It is one thing to conceptually explain new policies, it is the other to substantiate using (data-driven) examples;

- The reimbursement of court costs is currently a calculation done by humans. This process is time-consuming but additionally, and if you ask me a bigger problem, there is no insight into similar calculations across different courthouses. Automatised systems can take over this process, which will not only lead to more efficient processes, but also to an increase in quality across the different courthouses; and

- Planning and organisation within a courthouse are currently done by humans. It goes without saying that this is a time-consuming process. However, these functions can be done using an automated process that will lead to a more efficient use of time and improve transparency.

## 2.3. Experience

In the literature, the current belief is that a human judge has some advantages over an AI system, such as rationale thinking and providing reasoning as to why a certain decision was made (Ulenaers, 2020). However, it does not necessarily mean that there is no place for automatisation within the judicial system. Going back to the way judges gain experience, it will take a long period of education after which experience is gained in the work field. All in all, this process might require 10 years or more. Some of these limitations can be taken away by an artificial judge. Let's take the example in which some form of evidence is needed to prove certain facts have taken place. In case this evidence is not available, or not in a substantial way, a judge would have to argue that evidence is lacking in the specific case. An artificial judge would be able to quantify the probability at which the elements of the evidence could lead to the facts that took place (Xu, 2022). It would be able to quantify this relationship based on (legal) precedents. A judge would heavily rely on personal knowledge and experience and might even be at risk of providing a wrong decision.

Next to this, difficult and complex decisions rely on personal experience to provide correct decision-making. Furthermore, and that is one of the biggest arguments against the current system, a human judge is bounded by its overall rationale. For the same reason as previously mentioned, an AI system can generate new knowledge and provide relevant insights (Ojo et al., 2019; Gomes de Sousa et al, 2019). As an example, in a murder trial, an AI system could quantify the chances of recidivism. This could play a role in whether to sentence a suspect to prison with or without detention under a hospital order.

### The AI judge: from a global perspective

*The previously argued AI systems are not just a 'one day dream'. In fact, empirically, a lot of countries have already adopted a form of an AI judge. In Brazil, the AI tool with the name VICTOR is used to analyse a case even before a human judge has had a look at it (Becker and Ferrari, 2020). One of the prerequisites to appeal at the Brazilian Supreme Court is the presence of a general repercussion. Before VICTOR, this process would take around 40 minutes for each lawsuit. VICTOR takes less than 10 seconds. There is additionally one aspect of VICTOR that is interesting as well. According to the Council, AI in the Netherlands is still a step too far because court cases are unstructured and do not follow the same pattern.[19] However, VICTOR uses Optical Character Recognition to obtain a machine-readable text-file, even if the input is pdf, jpeg, scanned or handwritten. Moreover, and that makes it interesting as well, VICTOR can distinguish legal reasoning (from any other text). In Estonia, an AI judge can rule in small cases of less than EUR 7.000 (Nillier, 2019). The Singaporean judicial system introduced a system that uses speech translation systems to transcribe court hearings in real time which allows judges to review oral testimonials (Chiang, 2017). In Colombia, the Constitutional Court has a tool called Prometa that predicts the outcome of cases with an accuracy of 96%. In addition, it can sort cases based on the urgency in less than 2 minutes. With the human hand, it would have taken almost 100 days (Rivera, 2020). With these examples, it might*

---

[19] Leo Wees wrote an article on this topic, in a personal capacity, and calls the current situation of court rulings in the Netherlands 'garbage' (Wees, 2022).

*even feel as if a lot of countries already have adopted a form of AI within*
*their respective courtroom except for the Dutch system.*

## 2.4. Objectivity

While writing this thesis, there has been a trend to an increased criticising society. This trend is not only noticed in the Dutch judicial system, but in all levels of society such as the (Dutch) parliament and police officers. It is partially in the easiness of writing critics, through social media for example, that led to this trend. According to the public, sentencing decisions seem too unfair, inconsistent and too lenient (Zdenkowski, 2000). However, not only the public criticises: academia and scholars have been criticizing the unpredictability of court sentences as well (Stobbs et al., 2017). Most definitely in sentencing, judges are responsible for providing the correct sentence and should assess objective and subjective factors. It is this, however, that allows for subconscious and even an (intentional) bias.

> ### *Judge v. decision*
> *There are many factors that influence human psychology. Especially in*
> *decision-making, a consistent and correct decision is required. However, the*
> *question is whether this, practically, is the case. Two phenomena are*
> *important in this concept. The first term is decision fatigue, introduced by*
> *George W. Bush as 'the decider'. Research has shown that decision-making*
> *can wear you down, especially later in the day. In parolee hearings,*
> *prisoners showing up early in the morning would receive parole in 70% of*
> *the cases, while those who showed up at the end of the day would be paroled*
> *in 10% of the cases (Tierney, 2011). It is more 'shocking' to know that*
> *approvals would jump back to 65% after a snack. All in all, it seems like what*
> *matters is not the law, but the lunch.*

In one of the interviews, the following statement was made:

*"Humans cannot think rational, they can rationalise their thinking."*

I partially agree with that statement, and as such wouldn't be against an AI system as a solution to human judgment and biases.[20] With an AI system, assessments can be done more efficient, cost-effective but more important, more accurate (Hogan-Doran, 2017). Now, theoretically, it could be possible that similar cases could have different outcomes in Amsterdam's courthouse compared to Rotterdam's. However, an artificial judge will ensure equality and unity of the same laws in similar cases (Gao, 2019).

One of the downsides of a human judge is (bluntly saying): they are humans. Humans feel, get tired and sometimes do or say things that they regret later. Let's use the example of the after-lunch dip, an AI system does not get tired nor hungry. As mentioned in the literature (Crootof, 2019; Xu, 2002):

> *"... due to the personal preference, bias, burnout, corruption and other problems of judges, it is also a difficult luxury to expect human judges to keep a neutral objective and fair attitude. Human judges are notoriously inconsistent, both as a group and as individuals."*

A robot judge is not faced with constant (outside) pressure. One of the characteristics of a 'black box' system is that there is no constant pressure on the relations of input, features and output. The AI system's rationale is free of hunger or tiredness, and (external) influences, and gets rid of the arbitrariness of human judges (which is partially due to their given discretion). This will lead to a better unification of the law and reduces subjectivity when interpreting clauses. In Figure 1, a general overview of a black box system is illustrated. Any black box system, such as the Natural Language Processing in paragraph 2.1, uses 'hidden' relationships to define the output based on input data. As part of this research, the illustration below will be used to define an outline for AI in the Dutch courtroom later in this research.

---

[20] Some sidenotes should be made here, as I will explain in paragraph 4.3 and 4.4.
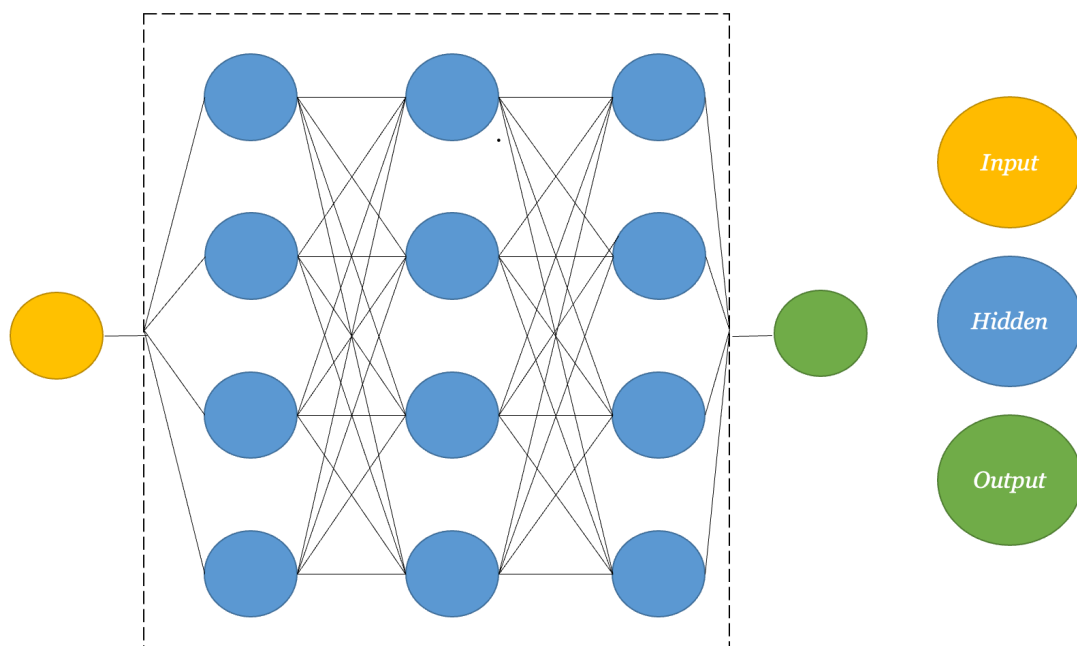
**Figure 1:** *a general overview of a black box system.*

## 2.5. Conclusion

In this chapter, the potentials of an AI judge have been addressed from a technical perspective to answer the following sub-question: *from a technical perspective, what are the promising potentials of AI and a robot judge within the Dutch courtroom?* It is crucial to note that, given the current circumstances in the Dutch judicial system, it feels like a *now or never* chance. AI systems are able, from a technological perspective to provide better case law due to their efficiency. These systems don't get tired, hurt or hungry. The systems can give an extra dimension to case law by making sure different courthouses will have similar outcomes and by providing the possibility to quantify chances and relationships. Moreover, AI systems are - in theory – not prone to constant pressure from outside such as corruption, biases and preferences. The fact that such a system will operate around a 'black box' system gives it preferential characteristics compared to a human judge. Even if AI will not function *as* court, the possibilities in terms of a system *in* court are numerous. It can help publishing cases online, provide insights in the consistent use of law or its effectivity, assist in writing verdicts or help in planning and organisation – to name a few examples. With an extensive proven track record across the world, the sub-

question of this chapter should be changed from 'what' are the potentials of AI to 'when' is AI implemented.

# Chapter 3: potentials from a juridical perspective

## 3.1. Introduction

In the previous chapter, the potentials of AI systems have been discussed from a technical perspective. The main takeaway from this analysis is that a robot judge can be a better judge compared to a human judge based on the three different factors being efficiency, experience, and objective. In this chapter, the potential of an AI system will be investigated from a juridical perspective. Such analysis will be done to answer the second sub-question: *from a juridical perspective, how can AI and a robot judge contribute to an improvement within the Dutch courtroom?*

In the interview conducted with the Council, one of the points of discussion was whether an AI system can rule while still maintaining the core values of the judicial system, and that of a judge in general: an independent court and unbiased and integer rulings. With the use of a systematic literature review, these three core values will be used as a starting point as to why an AI system will be better in providing and maintaining these values. It is important to draw this analysis because if an AI system wouldn't be able to guard these principles, the system will end up being an inadequate replacement in the first place.

## 3.2. Independency

One of the main key values of the judicial system is independency as being guarded by the Council. In addition, according to article 6, paragraph 1, of the ECHR, a trial needs to be conducted by an independent and impartial court which is established by law:[21]

> *"In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. Judgment shall be pronounced publicly but the press and public may be*

---

[21] A full extract of Article 6 ECHR can be found in Appendix C.

*excluded from all or part of the trial in the interest of morals, public order or national security in a democratic society, where the interests of juveniles or the protection of the private life of the parties so require, or to the extent strictly necessary in the opinion of the court in special circumstances where publicity would prejudice the interests of justice."*

Passing by the requirement that an AI system should be established by law, the system should provide independent rulings.[22] To fully grasp the potential of such system in this context, it is first important to understand the concept of independency.

### 3.2.1. The concept of independency

A judge should be able to decide in a court solely based on the evidence and arguments presented by the relevant parties. Only by such manner, a judge can provide fair justice. The concept of impartiality is in that sense also important. A judge should not act towards a specific case with a bias, both to the law as well as the parties involved.

Independency refers to the entire justice system as well as individual courthouses (Nowotko, 2021). This independency is not only in terms of providing independent case law, but also in terms of financing the judicial system. It refers to the position the judicial system has within society which is broader than judge X's verdict in case Y. Within court, it is important to mention that judges should appear (to the court) as independent and impartial. As such, AI can provide better and more independent and impartial rulings – at least compared to the current status quo. This conclusion can be drawn based on two factors: increased impartiality and removing of biases. Impartiality will be discussed in the next sub-paragraph, biases in the next paragraph.

Prior to the discussion on the improvement of the quality of law from the perspective of impartiality and biases, it is worth mentioning that *any* AI system must be

---

[22] To be able to provide for a robot judge, the system should be established by law. This practically (also) means that a robot judge should be accepted by the Dutch House of Representatives and Senate through a law before it could be used.

developed by the judiciary itself. As will be explained in paragraph 3.5, it is important to be independent and to ensure this independency. For obvious reasons, if an external party will build the model, this party will be albeit indirectly highly involved in producing case law. This would put the external party at position to know the system and to diminish (or even fully remove) the independency of a robot judge.

### 3.2.2. Increased impartiality

AI is, at least in theory, able to guarantee *complete* impartiality. In paragraph 2.2.2., three different forms of predictions have been explained. In all these systems, impartiality is increased. Impartiality refers to the situation in which the judge is free from *any* relationships with parties involved. Parties must be treated equally and the judge is guided by objectivity, also referred to in paragraph 2.4. Because an AI system cannot feel nor think, nor have any relationship with the parties involved, this impartial relation is guaranteed. As a matter of fact, a robot judge will consider the parties only as input data as reflected in Figure 1. It is of extreme importance to guarantee this characteristic because as mentioned by the European Commission (hereinafter: EC), partial judges will lead to a society at risk of being unpeaceful. In the literature, an impartial and fair judge would be interpreted as:[23]

> *"no-feelings, no-personal-agendas, no-political or religious ideals, no-needs, no-hates, no-passions, no-weaknesses, no-mistakes, round-the-clock-work, no-worries-about-poverty-at-retirement—just the cold reading of the applicable law, the jurisprudence, and its application through an objective and deep understanding of the facts presented."*

Theoretically and practically, this utopia will never be achieved using human judges because practically, any human will be influenced in some sort of way. It is a matter of dealing with these influences to *act* partial-free. The lack of transparency is a proposed argument as to why AI will not be able to provide a (more) impartial court. In one of the interviews, the argument was given that due to the lack of transparency, there is no possible control nor a check and balance to provide for an impartial court.

---

[23] Source: https://futurium.ec.europa.eu/en/european-ai-alliance/open-discussion/are-artificial-intelligence-courts-discrimination-risk.

However, that same argument can also be given in the discussion on human judges. To provide for an impartial (human) court would require a mechanism to control the minds of a human judge prior to their verdicts. On the contrary, machines can be implemented with checks and biases and do not have the ability to hide their real motives.

Another point that is worth mentioning is that most research on the implementation of AI is conducted within develop*ed* countries. However, the improvement that could be achieved by implementing AI in develop*ing* countries is even bigger. The system in developing countries might be prone to corruption and inefficiencies to a bigger extent than the more developed countries. Especially on the argument of impartial courthouses, bigger improvements might be achieved in developing regions (Ariel, 2021).

## 3.3. Biases

A machine learning system is fed by current rulings and learns from what relevant persons in court have said and ruled. In this context, AI is more of a relational judge to an independent judge. The system learns from previous cases and bases the outcome of the cases on previous decisions. From article 6 ECHR, a suspect has the right to a fair trial. This fair trial provides that it is free from any biases. To understand how biases can be omitted by AI *as* court, rather than AI *in* court, it is important to define the different biases that can occur in a decision-making process. The term decision-making process is broader than court rulings per se. The reason for this is that a verdict not only consists of the verdict itself, but also other processes that require decision-making or at least a rationale: should evidence X be included in the file? Is witness Y relevant to the case? Are the rights of the suspect sufficiently provided? All these questions require some sort of decision-making that is important to the case itself and can include forms of biases.

Das and Teng (1999) categorized four different types of biases that might occur in decision-making processes. These biases can be classified into four groups:

I.     **Prior hypotheses and focusing on limited targets**

The most common form of biases is those resulting from prior hypotheses and decision-makers who are focused on limited targets. In this form, decision-makers are driven by prior beliefs or orientations. Strictly speaking, a judge that has prior beliefs or orientations cannot be classified as a biased judge – as humans we all have these. The biases occur whenever the judge (or court in general) only focuses on selected interests and outcomes and ignores conflicting information because of their prior beliefs or orientations. The most famous form of this bias is the confirmation bias in which we seek information to confirm our beliefs, inadequately considering other beliefs. An example of a *possible* case of confirmation bias is the Dutch villa murder case.[24]

### *The villa murder in Arnhem: juridical error?*

*One of the most influential cases when it comes to potential biases within the Netherlands is the villa murder in Arnhem in which a 63-year-old woman was killed. On the 12th of December 2000, nine persons were sentenced for murder varying from five to twelve years imprisonment. More than two decades later, this case is being addressed to as probably 'the biggest juridical error' of all time in the Netherlands.[25] The Dutch Supreme Court ruled on the 20th of April 2021 that the sentences from 2000 will be respected and there is no need to reassess the verdict.[26] As such, strictly speaking, there does not exist a juridical error. However, the circumstances might lead to other conclusions: the verdict is a victim of biases. The suspects all happen to be involved in the drugs scene and were – at that time – drugs users. As such, they were linked to the crime scene even though factually they weren't seen at the crime scene. The belief that drug users were more prone to committing a crime led them to 'be' on the crime scene. According to research, the suspects were forced into confessing the murder even though they were lacking critical suspect information (e.g. information that only the real murderer could have*

---

[24] The word possible is intentionally written in italics. Even though the literature argues for a juridical error, such verdict has never been given by a judge or courthouse.

[25] In Dutch news, the case is referred to as 'probably the biggest juridical error ever taken place', source: https://eenvandaag.avrotros.nl/item/arnhemse-villamoord-waarschijnlijk-grootste-gerechtelijke-dwaling-ooit/.

[26] This verdict is published under number ECLI:NL:HR:2021:633.

*known). Suspects would be given this information and asked to repeat the information, suggestive questions had been asked and interrogators would use the fact that these suspects were not fluent in Dutch to their advantage. In the end, and after hours of interrogation, one suspect confessed. Based on this confession, all other eight were found guilty. One point worth mentioning is that all this was based on two factors: the fact that the suspects were drugs users and that they were non-residentials (e.g. did not have a fixed residency and a stable form of income). Statistically, it might – or might not – be the case that drugs users would commit more crimes. However, would that justify having a status-quo: all drug users and non-residentials will commit a crime? This would pave the way to biases based on prior hypotheses. The first request to reopen the case was turned down on the 21st of April 2021. The second request has been submitted and is pending. Another point worth mentioning is that biases (or even presumably biased) are hard to prove. While many questions can be asked on the legal justification of this case, so far, none of these led to another conclusion than the one dated from 2000.*

II. **Exposure to limited alternatives**

In decision-making processes, a set of alternatives should be considered to propose the correct solution. Due to for example time constraint or philosophical orientation, decision-makers reduce the problem to simpler forms and consider fewer alternatives. In considering the best solution, decision-makers often rely on intuition as compared to a rational analysis.

III. **Insensitivity to outcome probabilities**

In this form of bias, cases are often seen as independent and unique. Therefore, a case is being decided on using subjective judgements as opposed to rationale and probabilities, irrelevant to outcomes of previous cases.

IV. **Illusion of manageability**

These biases occur in cases decision-makers tend to overvalue their power to control and manage everything. The fact that a suspect was sentenced with a

longer sentence doesn't necessarily have to mean this suspect will become a 'better' person afterwards. Rationally, a decision-maker can only think of certain consequences since a person is not able to fully think through the implications of their decisions.

In general, these biases can be grouped in two forms of biases: social biases and cognitive biases. The first two biases, prior hypotheses and exposure to limited alternatives, are social biases and occur due to someone's impression and thoughts based on the social group they belong to. Research has shown how in an interviewing process, for example, similar people tend to be 'grouped' as safe, and managers tend to not hire people with a different (educational) background (Bohnet, 2016). Moreover, we might assume someone from a different ethical background to not tell the truth merely because this person communicates with an accent. Furthermore, anchoring bias can be classified as a cognitive bias. In this form of biases, humans tend to hold on to the first suggestion or information that was received, as such the reference to the word 'anchor'. The latter two can be classified as cognitive biases in which systematic tendencies might lead to errors.

Next, the question is how and to what extent AI *as* court can mitigate or even remove biases and which biases can be removed or reduced better by AI (as compared to human interventions). The difference between mitigating and removing biases is that in the first, the bias itself stays intact but a work-around is found. In the latter form, the biases are diminished directly.

In fact, AI will be able to mitigate both forms of biases but will tend to be relatively better in removing social biases. In the US, a software tool is developed that takes out *possible* biases of current verdicts: *Correctional Offender Management Profiling for Alternative Sanctions ('COMPAS')*. COMPAS is found to be an effective way used in court to quantify the likelihood the defendant will be a recidivist, free from factors that it cannot control itself (e.g. race or gender). Judges (and juries for example) can rely on these likelihoods to define the sentences (Dieterich et al., 2016).[27]

---

[27] The results of COMPAS will be discussed in paragraph 4.4.1.

On the general question how AI will *eliminate* biases, the following argument is given in the literature (D'Amato, 1977):

> *"Law might seem more impartial to the man on the street if computers were to take over large areas now assigned to judges. There is certainly some degree of belief on the part of the public that judges cannot escape their own biases and prejudices and cannot free themselves from their relatively privileged class position in society. But computers, unless programmed to be biased, will have no bias. They will give the same result on the same facts irrespective of the race, colour, wealth, talents, or deference of the litigants."*

It is a fact that the system will give the same results on the same facts which will ensure a judicial system free from social biases. One additional point is made: *unless programmed to be biased*. From a practical perspective however, in *any* case, AI will be better in *reducing* biases. A machine learning system offers way more visibility in terms of the components and ingredients of a verdict and allows for the possibility to include probabilities. Thus, this paves the way to a far greater opportunity to mitigate biases (Kleinberg et al., 2018). Even if we use the software tool COMPAS, much research is done, and is being conducted, on whether such a tool actually does what it claims. It is way harder to investigate a human brain.

### *The Dutch childcare benefits scandal*

*As mentioned previously, biases can happen in any decision-making process. Therefore, this does not only apply to a judge or courthouse. Even if a process is automated, which this research investigates, it is important to keep 'an eye' on the process and make sure the results are as intended. Sadly, this is not always done. In The Netherlands, one of the biggest scandals on biases was the Dutch childcare benefits scandal (Dutch: Toeslagenaffaire). The additional point from D'Amato became a Dutch reality: unless programmed to be biased. In this example, an algorithm labelled more than 20.000 parents as fraudsters. Because the machine is self-learning, it mimics its output and learned from the qualifications. However, no physical control was ever done on this algorithm. Learning from this, it is important that there should be some process to ensure systems are correct: detect, mitigate*

*and repeat (Belchev, 2022). Dutch news headlined on the 5th of October 2022, while finishing this research, that the Dutch Tax Authorities used systems that automatically assigned immigrants and people with a low income as fraudsters. Algorithms without rules can destroy lives and we should never be facing a computer with empty hands.[28]*

## 3.4. Truth-finding

In a recent study on the differences between human judges and machine learning systems, a different approach is followed. In a 2022 report, an empirical study concluded that AI is a more promising technology in the detection of deception through facial expressions as compared to a human judge (Monaro et al., 2022). In fact, a machine learning technique performs better in finding liars compared to humans in all the tests conducted. The machine learning system was able to differentiate liars from suspects telling the truth with an average accuracy of 75%. Compare this to the accuracy of human judges, which was approximately 57%, and the conclusion can be drawn that an AI system is more promising in truth-finding. Even surpassing the fact that in 25%, the robot was not able to tell the truth, it is an increase of 18% in the current situation. In terms of improving the quality of case law, this would result in an improvement. From a behavioural perspective, it could even lead to bigger wins. Suspects can think it can 'fool' humans, but would they still (try to) lie if they are faced with a computer?

In another research, the use of 'expert robots' is argued for to test testimonies in court (Katz, 2014). Under the current system, in case a statement is done by one of the parties involved (including a witness), this statement needs to be verified. Using a current Dutch example, inhabitants around the area of the factory of Tata steel are preparing a claim for compensation because they are allegedly exposed to bad air quality (due to Tata Steel) and this led to more cases of respiratory diseases.[29] Let's just say, for argumentative sake, one of the claimants proposes the argument that

---

[28] Source: https://www.ad.nl/politiek/europarlement-walgt-van-toeslagenaffaire-schandvlek-in-de-geschiedenis-van-de-rechtsstaat~ab375069/.
[29] Source: https://pointer.kro-ncrv.nl/gezondheidsschade-tata-steel-leidt-tot-aanklacht-en-massaclaim.

due to bad air quality, chances of diseases increased with 30%. To analyse this percentage, lots of time is required and might lead to inefficient use of resources. However, an AI system can 'pick up' this claim and analyse its statistical truth. It will also take away one of the current challenges: that of the credibility of current experts. Let's refer to the case in which (a group of) citizens are in court against a big company. It might be the case that this big company is able to provide statistical proof as to why their claim is not true. I am not saying that these companies will ask experts to lie. However, it is not what you know but what you can prove in court. If the expert robot can analyse the proof itself, the court will be less dependent on what is provided for in court.

Additionally, and a less severe application, in case a party involved in a labour case claims to be working at a firm since a specific year, the AI system can pick up this statement and fact-check it with for example information available at (local) Tax Authorities.[30]

These previous examples are illustrated to show the possibilities of an AI system in the court. Not only will it lead to efficient use of scarce resources, but also the resources that are (made) available can work (efficiently) on other tasks.

## 3.5. Principles of law: legal certainty and equality

In one of the interviews with the Council, the argument was raised that an AI system would never be able to fully replace the human judge because it is against the provisions of article 6 ECHR. The reasoning to this statement was that a verdict from an AI system would be not transparent and is not able to explain 'its reasoning'.

> ### Predicting vs. explaining
> *In the interview with the Council, it was mentioned that there is an ethical reason not to implement AI. One of the characteristics of AI is that it is a 'black box' system, and the relationships between input and output variables*

---

[30] Of course, such interaction between the robot judge and other systems, such as data from the Tax Authorities, should be conducted in line with regulations on the use of data (e.g. GDPR) and requires a legal competence for both parties to obtain and share such information.

*cannot be explained, at least not to a complete extent. A suspect that receives a verdict should be able to know on the bases of which ground the verdict was based. As the 'black box' does not allow for rationale, it is hard to provide a suspect with that reasoning. Even if, theoretically speaking, a certain explanation would be provided, an average suspect would still not be able to interpret it. There allegedly is a big gap between predicting and explaining.*

*The same argument was proposed in the interview with the WRR.[31] As part of the argumentation, the fact that AI is a 'black box' is in general not a problem. The verdict is based on precedent case law and if it is correctly anticipated, the verdict would be correct too: that ís your explanation. On top of that, in general, human judges decide on a case within four walls.[32] A suspect would not be allowed to be part of that discussion whatsoever. In that sense, no suspect knows on what grounds the verdict is based on other than those mentioned in the verdict specifically. The latter is something an AI system could do as well. The argument that an AI system can't think rationally didn't last very long either. In fact, every human being can't think rationally; they are only capable of rationalising their reasoning rather than thinking rational. To that extent, an AI system would be an improvement to the judicial system.*

In paragraph 2.2, the use of AI *in* as well *as* court has been argued for. In this context, both the applications will lead to a substantial increase in the quality of law taken from one of the core principles: legal equality and legal certainty. As part of the capability of an AI system to be able to provide efficient searches based on (other) court's rulings, a similar explanation of open norms can be provided. If these open norms are explained equally across the different courts within a specific region (e.g. the Netherlands), this will lead to a higher legal equality within the country and across the different open norms. This legal equality might not be provided for in the current status quo based on the simple fact that it would be too time-consuming to

---

[31] To avoid confusion, the abbreviation of the Scientific Council of Government Policy is kept at its original abbreviation 'WRR'.
[32] Judges will deliberate in a room away from public, cameras and parties involved. The process that is going on behind closed doors will never, at least not one-to-one, be available to the public.

analyse interpretations of different open norms across courts. Additionally, the lack of qualified people to do this, as referred to in paragraph 2.2.2., will not make it possible to provide 'human' checks and balances within the judicial system.

This would one-on-one also lead to a higher legal certainty. Due to the fact that it is certain how different open norms are being explained, parties involved can decided (prior to filing a case) whether it would be beneficial to ask for a court ruling or to find other ways of dispute resolution (e.g. mediation). This will decrease the burden on courts and will lead to more time and (physical) capacity for cases that require human involvements.

> ### *The problem of knowing the system*
> *As part of the interviews, I have asked both the Council as well as the WRR who should be held responsible to build such an AI system. The Council explained that it is a big question who should build these systems. The current expertise within the Council is lacking as well as the capacity and funds to provide for such a system. Rather, a company such as Deloitte, IBM or Microsoft would be able to provide better systems, purely since building automatised systems is part of their core business. The same question was asked to the WRR and their answer was firmly that such system should be built by the Council. The Council is responsible for the Dutch courthouses and their rulings, and this implementation will be part of their responsibilities.*

Combining what is previously said on legal certainty and the arguments of both the Council as well as the WRR, it is inevitably that only two answers exist to the question who should build the system. Either a third party will build the system but will never be allowed to use the system in any court hearings (e.g. start an AI-procedure), or the Council is solely responsible for this system. For some obvious reasons, the organisation that builds the system is also aware of the relationships and probabilities within the system. As such, there is extensive information available that will be against the equality of parties in front of a robot judge in case this organisation ever gets involved in automated court procedure. In terms of legal certainty, it is clear how the court will rule in specific cases. This knowledge, however, should be protected with priority.

## 3.6. Conclusion

In this chapter, the potentials of an AI judge have been addressed from a juridical perspective to answer the following sub-question: *from a juridical perspective, how can AI and a robot judge contribute to an improvement within the Dutch courtroom?* Based on the analysis, the conclusion can be drawn that a robot judge is better in ensuring the core values of the Dutch judicial system: independent courts and unbiased and integer rulings. An AI system will be able to provide independent and impartial rulings because it cannot feel, sense or be influenced. Additionally, it can both remove as well as mitigate social biases and provide insights on the way different characteristics, such as race, gender or sex, have an influence on the outcome of a case. On the other side, because case law is part of 'law' of a jurisdiction in general, AI will be able to increase legal equality. The same case will have the same outcomes irrespective of the courthouse or the judge. This will also lead to more legal certainty as open norms will be interpreted similarly across courthouses. The applicability of a robot judge in terms of truth-finding proves how technical initiatives will be able to benefit a juridical system. As the sub-question specifically refers to a 'contribution', it is evident that a robot judge will bring a proper contribution to the table – or seat.

# Chapter 4: limitations from a technical perspective

## 4.1. Introduction

In chapter 2 and 3, the potentials for a robot judge have been discussed from both a technical as well as legal perspective. In chapter 3, the potentials from a technical perspective proved that an AI system will be beneficial to a judicial system in terms of efficiency, experience and objective. Based on this reasoning, the impression can be drawn that human judges can be replaced by AI with ease. However, some limitations exist in the implementation of AI, or should at least be considered. In this chapter, the limitations from a technical perspective are explained and the potentials are used as a starting point. By means of a systematic literature review, the following sub-question, *what challenges will an AI system within the Dutch courtroom face from a technical perspective,* is answered.

## 4.2. Efficiency and quality are not synonyms

As mentioned in paragraph 2.2., one of the solutions that AI brings is that of efficiency. Due to the numerous amounts of cases and the lack of judges, workload is increasing. AI can be both beneficial to the judicial system *in* court as well *as* court. However, a critical note should be placed as efficiency and quality are not synonyms.

### 4.2.1. Input in, input out

First, a robot judge is a system that can predict future outcomes based on data. This data consists of past verdicts and therefore, future verdicts are related to past decisions. As such, and also in terms of legal equality, it is positive that judges' outcomes are predictable. The quality of decision-making processes is related to the input: the cases. However, taking this to a practical level, questions can be asked on the correct relation that is drawn between these past and future outcomes (Xu, 2022). There is no body or organ that is responsible for the quality of the verdicts if the judge is left out. In case such a body or organ does exists, will that person not (practically) be the judge in this case? Additionally, taking the quality as given, questions can be asked on the capability of an AI system to *read* the data correctly.

For some types of cases, such as basic cases on property valuation (Dutch: *WOZ-waarde*), the input data can give a prediction with a fair accuracy. This has to do with the fact that corresponding law (and therefore the cases) are straightforward and structured. However, criminal law cases are subject to a lot of details that cannot be interpreted by an 'AI brain'.

### 4.2.2. Human brain in, AI brain out

Second, the AI brain is non-existent at all, or at least to a limited extent. The brains' capacity is limited by the capability of the design and input by humans (i.e. the data scientist responsible for the model). Fact is that humans in general cannot achieve an intelligence level of 100%, so will the model itself never reach this level. More severe or difficult court cases require multiple judges could be a reason to ensure maximum intelligence is achieved. Additionally, humans can find some sort of way to make up for the fact that a 100% intelligence level is not reached. In legal context, judges can seek advice of other judges, or third parties, or look up theory that might be relevant for the case. However, an AI system does not have such capabilities outside of its system. Even though many machine learning techniques allow for deep learning, it is still very limited by the possibilities of the program used. The fact that the robot judge works with and on justice is a component that shouldn't be forgotten. Even if an accuracy (e.g. intelligence) is reached of a certain level, caution should be taken into interpreting this level.

> ***Moonlit: an accuracy of 73%***
> *In the interview with Moonlit, one of the first questions was on the prediction performance. The average accuracy is about 70% with a most recent accuracy of 73%.[33] This sounds like a promising accuracy, however, some questions are in its place. Taking a step back to the way an AI system works, and the specific Natural Processing Language, it is all about syntaxes. The outcomes of syntaxes are defined in terms of zeros and ones. In other words, it is either "party one wins" or "party two wins". Using a confusion matrix, the accuracy is determined. In a confusion matrix, the two predictions are*

---

[33] As taken from https://moonlit.ai/ - "Rolling performance".

*set out against the actual outcomes. Using the example of Moonlit, their*
*confusion matrix will look like the following:*

| | | Actual values | |
|---|---|---|---|
| | | Taxpayer wins | Tax Authorities win |
| **Predicted values** | Taxpayer wins | *Correct prediction* | *False prediction* |
| | Tax Authorities win | *False prediction* | *Correct prediction* |

***Figure 2:*** *a confusion matrix for the predictive system of Moonlit.*

*Going back to the accuracy of around 70%, the most important reason why it*
*is not so promising as it looks is because a correction for guessing should be*
*considered. With two outcomes available, either Taxpayer wins or Tax*
*Authorities win, I will have a correct prediction of around 50% if I predict*
*one outcome for all the cases. Therefore, the relevant prediction (and thus the*
*system of Moonlit) only led to an additional increase of around 20%. The*
*second sidenote that should be put in place is that of the outcomes. In the case*
*of a taxpayer v. tax authorities, generally speaking, there are only two*
*possible outcomes. It is either the taxpayer was right (e.g. less or no taxes*
*should be paid) or the tax authorities were right (e.g. taxes (or more) should*
*be paid). These obvious outcomes are less present in other cases. In a murder*
*trial for example, outcomes could vary a lot. Additionally, a point worth*
*mentioning is the fact that between the period January 2020 – January*
*2022, the system did not perform better in its prediction. The accuracy in*
*2020 v. 2022 is comparable.[34] As part of an AI system that is suitable to*
*function as court, it is important that the system learns from these*
*predictions to perform better. This self-learning component seems to be*
*lacking or missing in the case of Moonlit. Lastly, it is a critical point that on*

[34] As taken from https://moonlit.ai/ - "Rolling performance".

*the other side, the system did not predict correctly in around 30% of the cases. As to correctly apply the rights of article 6 ECHR, any system would need to perform with a 100% accuracy – or close to that. Especially because a system should learn from its own verdicts, these 30% 'incorrect' verdicts can compound to a bigger number of inaccurate predictions (e.g. the system uses incorrect verdicts as input data and builds further on this data). This would lead to the classical example of 'garbage in, garbage out'. On the contrary to all what has been said, another side note should be made. The fact that the system predicts differently than the actual outcome might not be a system issue. What if not the prediction but the actual verdict is incorrect, for example because of biases as explained in paragraph 3.4? What if the real outcome should have actually been the predicted outcome? In a human v. robot judge difference, which of these is correct and who should decide? As easy as this sounds, this might be more of an ethical than an actual juridical problem. Additionally, how would a party involved react to the 70%? Unintendedly, the system has the effect that it is 'wrong' in 30% of the case. If I was a party involved, and my chance of not winning this case would be obvious, I could still go for a shot since the system is incorrect in 30% of the cases. Would that mean that abovementioned system increased my chances from 0% to 30%?*

Lastly, passing by the fact that AI is dependent on the quality of the input and the limited extent to which it can 'think', there is the factor that humans overly rely on artificial systems and thus requires a higher accuracy compared to human judges. Fact is that society is less prone to errors of humans as compared to any (technological) system since these technological systems should 'work' regardless (Perc, 2019).

## 4.3. Experience and data are not synonyms

As mentioned in paragraph 2.2, AI systems apply current cases to patterns drawn from input data and the correlations, and based on that, apply this to provide a decision-making process. As the number of cases would pass, the system would learn not only from input data but also from new cases: deep learning. This would lead to a

self-learning machine. Or at least, so promises the theory in favour. However, experience and data are not synonyms.

The first reason on why experience and data are not synonyms, and therefore not all knowledge can be determined in terms of data, is the fact that decision-making conducted by an AI system is free of human involvement: it does not come with the ability to listen to the real meaning of words, to the emotions in a certain case, trying to mediate between parties involved and any other factors that would require social skills. For a fact, these social skills are not available in this system and a judge would have only required them throughout an extensive career and by training-on-the-job. There is no such code or deep-learning method that could replace this. Some cases would not (necessarily) require this form of decision-making, because their core problem is a matter of technical explanation of the law or case. In these cases, a binary decision, in terms of a 'yes' or a 'no' would suffice. However, there are cases in which an exception or interpretation should be done, which can only be conducted by a human.

### *Law v. code*
*Researchers have asked questions on the practical possibilities of using an AI system to rule (Kauffman & Soares, 2020). The law can be very static and contains a lot of open norms. To provide an example, the Dutch law contains the following article on self-defence:*

> *"He who commits an act, provided by the necessary defence of one's own or someone else's body, honour or property against immediate, unlawful assault is not punishable." (Article 41 Dutch Criminal Law)*

*What is necessary? What is immediate, unlawful assault? And how to define someone's honour? The law itself contains a lot of open norms, and is most of the times a very factual question. Even if all the words (and relevant) probabilities can be transformed into code, the question would still be how to apply this system in each circumstance. It is the same rigidity that provides benefits to such a system, that will also provide its drawbacks. Legal language, in terms of the law, is nuanced, requires interpretation and*

*contextual understanding. IT professionals do not have such experience or knowledge, nor have they been involved in or aware of the policy around certain laws. Nevertheless, it is these professionals that need to ensure that legislation and case law are correctly translated. Even if they would 'team-up' with lawyers and judges, it would still require information that cannot be transferred to code. Additionally, these codes would still require to be updated frequently based on new case law or amendments to the law itself. This means that an AI system will still be dependent on human involvement. The natural question here is if frequently updating will still result in efficiency improvements (as argued for in paragraph 2.2)*

As such, the application of AI can, at least for now, only be limited to situations in which there is a clear, single answer (e.g. right or wrong) and the underlying case is structured.

## 4.4. Objectivity is not the same as law-making

As mentioned in paragraph 2.3, the application of judicial AI is argued for as being objective. However, some critical notes can be placed on the objectivity of AI.

Judicial is all about providing justice to the parties involved. Justice can be compared to the flag on top of a mast. It goes wherever the winds pushes it to (Radbruch, 1997). In other words, justice is a dynamic concept rather than a static given. Practically however, AI treats justice as a static concept. This leads to a positive aspect of AI: its neutrality. This nevertheless might not always be the best solution. Neutrality (and to a bigger context objectivity) is not a synonym to law-making. To illustrate this, two concepts will be explained: bias in, bias out, and the functions of the court.

### 4.4.1. Bias in, bias out

As mentioned in paragraph 3.3, AI can be a tool to limit, or reduce biases. AI bases its verdicts on the provided input. As such, any biases that already exists within current verdicts, and as such be used as input, will live to infinity. In paragraph 3.3, the software tool COMPAS has been explained. This software tool is promising

because it can predict the likelihood of recidivism, free from biases that arise due to skin colour or race for example. The potentials are promising. However, in that paragraph, the specific results have not yet been discussed. Behind the promising possibilities, a downfall lies. Based on recent study, people with a black skin colour were found to score 45% higher compared to Caucasians in being qualified at a risk of recidivism. Of course, statistically speaking, chances might be that based on one characteristic (e.g. a skin colour or gender) a specific pattern can be denoted. In fact, a black male was found to be at most risk (Moore, 2011). However, doesn't that lead to labelling (e.g. a black male is at higher risk)?

The promising possibilities of *any* tool (re)lies on and in its data. It is becoming problematic if the AI systems starts to learn (i.e. to judge) from that specific characteristic. An AI system cannot start from scratch and is reliant on its input: we feed AI systems everything, including any biases (Metz, 2019). The system acts similarly to a child learning bad behaviour from its parents. The bad patterns can be controlled by society (e.g. taxes on cigarettes or by controlling through education). However, such controlling mechanism is not (yet) available for an AI system. It would require someone to check these probabilities and relationship by hand, one by one. Ignoring the fact that this might be a very time-consuming task to do (think of the number of probabilities that can exist), what is the position of this person (or group of people)? They are after all responsible for creating probabilities and inherently involved in the production of case law. Simply making this a responsibility of IT engineers (that if you ask me should create the AI system in the first place) would be against all principles of law. It should however be the responsibility of the Dutch judicial system to control for these probabilities.

If none of these control mechanisms is put in place, the use of any AI system would merely be the transmission of human biases into the predictive system. Even if all the biases in the input data would be cleaned, which is a pure theoretical argument, still other biases could occur. To name two biases, there could be biases in the training set, which is used to train the predictive model. Biases that occur could be the result of the selected training dataset not being representative of the actual cases. Therefore, the system would not be properly trained. Additionally, biases could occur on the level of labelling data: algorithmic biases (Malek, 2022). An AI system can

therefore reduce human biases, but create new ones too (Zabrsnik, 2020). But even to a bigger extent, biases occur because of the characteristics of an AI system: it focuses on probabilities, not certainties. Additionally, it measures correlations and not causations (Malek, 2022). This leads to the question on how the outcomes should be interpreted. The outcomes of a predictive system should therefore be interpreted the way they are: probabilities and correlations. However, they are (a lot of times) interpreted as certainties and causations. Moreover, in case an AI system is used as supportive to a human judge's decision, another bias could occur. The tendency of a judge to only act in accordance with their own perceptions, or broadly saying outcomes that confirm their prior beliefs, could lead the way to confirmation biases (Malek, 2022). Therefore, the argument that biases are removed in an AI system is partially true as it leads the way to other biases that should be accounted for.

### 4.4.2. The functions of the court

Some scholars criticise the (general) use of any AI system within the judicial system (Matsuo, 2016). Making 'law' is not just a matter of combining factors and predictions, and to provide a verdict based on that. In general, a lot of people go to court because they want to be heard, not because they want to be ruled. As an example, a Japanese scholar pointed out that divorce cases for example involve a lot of emotions that are just not suitable for an AI system: these cases involve a lot of emotional factors and require some form of coordination of the interests of parties involved (Matsuo, 2016). Considering that (almost) every case is unique in some sort of way, a judgment based on previous cases might be hard to justify – especially when parties want to be 'heard'. Therefore, as Matsuo argues, the application of AI systems in some fields is not appropriate (i.e. divorce or property inheritance). Additionally, the responsibility of the court is doing justice, guarantying liberty and enhancing order. Other than resolving disputes, the court is also responsible for maintaining the rule of law, provide for equal protection and to ensure due process of law. While we have seen an AI system will be able to strengthen the functions of due processing of law and maintaining rule of law in a better way than the human judge, questions can be asked whether the same conclusion can be drawn for doing justice and enhancing social order. In fact, the question is whether that ever will be possible, even over time.

## 4.5. Conclusion

In this chapter, the challenges of an AI judge have been addressed from a technical perspective to answer the following sub-question: *what challenges will an AI system within the Dutch courtroom face from a technical perspective?* Any new system will come with a set of challenges that need to be addressed and considered. The way the AI system works, as it is based on human input, provides challenges on the quality of law as well as ensuring no biases are mimicked. Additionally, concerns can be raised on whether an AI system is suitable in all fields of law especially considering the functions of a courthouse. As partial conclusion, the challenges do not stand in the way for a general implementation of AI to begin with. It is however a matter of making sure these challenges are dealt with and so-called 'work-arounds' are created. For example, the system should generally have 'checks-and-balances' to make sure biases are not being part of case law. Even if this might be the current case (e.g. case law nowadays might contain biases), the implementation of AI will provide the possibility to have a 'fresh start'.

# Chapter 5: limitations from a juridical perspective

## 5.1. Introduction

In chapter 4, the limitations of a robot judge have been discussed from a technical perspective. From a technical perspective, limitations exist in terms of the quality of predictions as law, the lack of 'real' experience and questions on whether objectivity and case law are interchangeable words. Additionally, some limitations might exist from a juridical perspective. Frankly speaking, it would be naïve to think that a robot judge would be able to replace a human judge from one day to the next. In chapter 3, the potentials from a juridical perspective have been addressed. These potentials are used as a starting point and by means of a systematic literature review, the limitations are explained from a juridical perspective to answer the following sub-question: *what challenges will an AI system face within the Dutch courtroom from a judicial perspective, also considering the rights of a suspect?*

## 5.2. The lack of data

As has been covered in this research extensively, the AI system draws patterns based on inputs. This will allow the system to improve legal equality. However, this is based on the foundation that the input used is complete. Even if the system can draw probabilities, it is impossible that this is done based on all the current court rulings. This leads to a possible drawback for the system to be used in the first place.

First of all, not all court rulings are published or in that sense available to be used. Only a selected 'few' verdicts are published. Nevertheless, hypothetically speaking, even if all court rulings would be published, not all relationships can be quantified. The court rulings are just too different for that. One of the drawbacks given in the literature is the fact that texture and wordings differ too much between different courthouses (Wees, 2022). The same case can be written down completely different between the courts in Amsterdam and Utrecht for example. It is the lack of a clear structure, or 'template' that contributes to this issue. Even if all the courthouses within a certain region (e.g. in the Netherlands) would follow the same structure, it is

the difference in wording that would provide extra challenges to translate these words into code. There are different ways of saying that any person is innocent, or free of any prosecutions. Would that automatically lead to a list of different probabilities? Or should there be one template or list of wording that will be processed as input data? If so, who will take this responsibility? It would be naïve to think that IT engineers can convert this input data without any legal background or understanding the meaning behind words. The other question that comes up is whether legal or law should be based on a calculative approach – as that is what an AI judge effectively brings to the table.

## 5.3. The lack of laws suitable for coding

### 5.3.1. Complex laws with open norms

In the interview with Moonlit, the hypothesis was given that a lack of coherent verdicts is one of the biggest challenges for any AI system to work. Moonlit focusses on specific verdicts, such as those from the Court of Justice in the European Union ('CJEU'), said that using only one courthouse as part of the pilot has been a big plus for the system. The CJEU has a specific template to provide these court rulings and the outcomes of these cases are binding for the respective countries or parties involved. However, the big issue according to Moonlit is not the lack of suitable verdicts, it is the lack of suitable laws. The law itself allows for complicated and incoherent verdicts. Part of this problem is created by the numerous amounts of open norms within the law, that are explicitly built in the law to provide room for interpretation. An example are the clauses on reasonable and fairness.[35] These clauses are open to interpretation and this interpretation depends on person to person. In this context, it is not a lack of data but more a lack of *integer* data. Take the example of a murder trial, there is an infinite number of rulings and as such infinite number of possible outcomes: 12 years in prison, 4 years in prison and preventive detention, no punishment because of self-defence, 12 years but additional 4 years because of the severity.

---

[35] As an example, the clauses in article 6:2 Dutch Civil Law is used.

### 5.3.2. The tendency to legalise everything

As been previously mentioned, the law itself contains a lot of open norms that allow room for interpretation. However, even going passed the 'overkill' of these open norms, there is the tendency to legalise everything which makes the law overcomplicated. In recent research on Dutch Public Law, some arguments have been proposed to step away from that current tendency of legalising everything (De Moor-van Vugt, 2014). The first reason is that times have changed, and the legislator should not try (and want) to determine everything in writing. One of the better examples to explain this is the fast pace of a digitalising society. It is the same digitalising that introduces the possibility of an AI system, that requires any law to be logic and not overly complex. A good example to illustrate this is the way different businesses operate these days. To determine where profits are taxed with a 'regular' business (e.g. a book shop or restaurant), it is a mere example of allocating where sales have taken place and where profits are generated. However, looking at digital businesses (e.g. Facebook or Twitter), the question where profits are generated and thus taxed is way harder to answer. The law itself shouldn't want (and need) to provide the answers to these questions. The pace of a changing society will ensure the law is always one step behind. The same reasoning can be given as to why the law shouldn't be 'updated' to ensure cryptocurrencies are taxed as part of a person's savings and investments. It is the fast-changing way of life that will ensure the law itself isn't able to cover all these changes in a *timely* manner.

Rather, the law should be simple, not overly complex and structured. Any queries that cannot be answered based on the law one-on-one will be provided for by court (based on the writing of the law and reasoning). This will ensure the capability to provide to the fast changing (complex) society, by still maintaining a form of legal bases. There is no prerequisite for a legislator to have any possible outcome in writing. The law should provide the core, any branches can be added by means of case law. It is a fallacy that laws should be predictable. As an example, one of the reason the *Code Civil* is extensive is because back in the days, society wasn't too complex and transactions involved were 'limited'. In recent days however, the rise of digitalisation, globalisation and different societies (also referred to as the branching of society) makes it (almost) impossible to have every scenario written out.

### 5.3.2. The need for more structured law

One of the reasons provided to allow for an extensive form of law is the argument of legal certainty. As part of the reasoning, a law that covers all (possible) forms and thinking will ensure parties involved know what to expect. However, two remarks are in its place. The first one is the fact that humans are not robots: the way they behave and act is a lot of times not based on the law (e.g. not completely rationale) Therefore, providing this legal certainty would be an overkill. Additionally, a lot of laws allow for deviations (Dutch: *hardheidsclausule*). These deviations are based on individual clauses and as such push legal certainty to the background.

All in all, the law itself should be less complicated, rigid and more structured to allow for robot verdicts. To achieve this, the clauses of the law itself should function more as a decision tree. Some laws in the Netherlands, for example the law regarding the valuation of real estate (Dutch: *Wet waardering onroerende zaken)* are already fairly structured and not too complicated. For some obvious reasons, there is an infinite number of possible effects on the value of real estate. However, the law itself does not cover all these aspects, rather that responsibility is given to case law. This type of dynamics between law and case law allows for an AI system to fully operate in a correct way.

## 5.4. A robot judge v. human judge

In paragraph 3.6, the argument was provided that an AI system will be beneficial because it increases legal certainty and equality. Because the system is built on the foundations of different court rulings, it will be able to provide court rulings that are equal among the different cases, irrespective of their region for example. However, there are two sidenotes that should be placed. The first one is the fact that the place of such an AI system is uncertain. What if a human court ruling is different than an AI system's, whose should be favoured? There are two possible answers to that question: will it be the human judge because this judge can, other than think, feel and innovate the law? Or is it the electronical judge because it makes sure all the different court rulings are equal and interpret the law in the same way, and as such has done more 'extensive' research. The strict answer is that neither of these two

should be favoured above the other one, they just interpret the law differently. In terms of favouring the human judge over the AI judge, no feelings (or even egos for that matter) are hurt. Such changes can be done to the system to ensure the AI judge rule similar to the human judge. However, the other way around is way more difficult as it is not as simple as setting the brains of the human judge to rule like the AI system. Some judges might be welcoming to electronic changes and believe it can benefit the judicial organisation. For these judges, amending their reasoning to be comparable to the AI's wouldn't be that hard. However, in case you have studied an extensive number of years and have a long track record in solved cases with over 20 years of work experience, it might be a bit harder to accept the fact that you were overruled by a computer. This feeling will be even more strengthened if the system itself is built by an IT consultant rather than judges with over 20 or 30 years of experience. To embrace technological change requires both the mindset of accepting it *and* being involved in it.

### 5.4.1. A robot judge: a philosophical interpretation

In previous paragraphs, the challenges for judicial intelligence have been described. These challenges have to do with the lack of data, lack of law suitable for this intelligence and the current tension between the human judge and the AI judge. For argumentative sake, let's assume these challenges are all regulated for: data on past verdicts is fully available, the law is 'technology proof' and human judges (and the legislator) have completely embraced the use of AI systems. Another issue that comes up is the quality of whatever outcome the judicial system rules. From a philosophical perspective, the implementation of AI is bounded by the (ongoing) discussion on syntax versus semantics. Any computer system is based on syntaxes; it interprets data and provides an outcome in a syntax form (i.e. zeros or ones). These outcomes of any case are abstract and do not allow for any deviation for the respective case and outcome. Although we have concluded that interpretation should be taken with caution, as it leaves room for biases to occur, the other side (e.g. no interpretation at all) is also not part of the solution. It is John Searle's main argument against any use of AI (from the use in courtrooms as their application in medicine) (Searle, 2002). As part of article 6 ECHR, any suspect has the right to a *fair* trial. The word fair in this context also refers to a court ruling that considers the relevant factors from both the

law itself as well as the arguments as brought up by the parties involved. Even if we would say that the system would be able to fully process input, it lacks feeling and human interpretation. To illustrate this, let's take the following example: the words *"he is coming to dinner"* or *"he is coming to dinner tomorrow"* are of a similar extent. However, the first sentence adds a form of urgency to the context which is less relevant in the second sentence. This urgency, or more general the meaning behind words, is something an AI system might lack. It will take time before such understanding of information, like that of the human mind, is replicated (Searle, 2002). This will draw a restriction on the use of AI *as* a Supreme Court. A Supreme Court, generally speaking, is responsible for interpretation of laws. This is, using the current way AI is built up, not possible to be done by a robot. It is worth mentioning, in this context, that judges do way more than just providing verdicts. It is the public and educative function of the human judge that cannot be fully replicated.

## 5.4.2. Lack of transparency

Let's move to a world in which AI *as* court has been widely accepted and all previous problems have been addressed. At a certain point, a suspect is faced with a verdict that for argumentative sake sentences him with five years in prison. As part of the suspect's right, he will appeal against the verdict. The next question comes up: on the basis of what? What are the arguments brought forward against the verdict? Additionally, the suspect should know based on what ground exactly. However, as mentioned in paragraph 3.5, one of the key factors of a machine learning system is its functioning as a black box and it will primarily lack in explanation. Shouldn't however these explanations be guaranteed? After all, we are talking about a person's freedom. The fact that a black box won't be able to provide this will undercut the suspect's sense of fairness and trust (Deeks, 2019). Even if we (in theory) say that probabilities and explanations are synonyms, would it be correct to assume the average suspect would be able to interpret these figures? In order to use AI *as* court, the first step is to provide explainable AI (in the literature also referred to as xAI) (Deeks, 2019). This will enhance trust between humans and the system, expose biases and unfairness and create insights into how a specific law (and thus 'the world') works. Just think about it, can anyone explain why Alexa or Siri respond the way they do? In fact, this paradox of 'not knowing how' is very common in practice:

do we actually know how a washing machine gets our clothes clean? The more important question if you ask me is, do we actually want to know? Practically, if clothes are not cleaned properly, we wash them again – perhaps even by hand. Why can't we apply this approach to an AI judge? If we are not happy with the outcome, we file an appeal – perhaps even at a human judge.

## 5.5. Conclusion

In this chapter, the challenges of an AI judge have been addressed from a juridical perspective to answer the following sub-question: *what challenges will an AI system face within the Dutch courtroom from a judicial perspective, also considering the rights of a suspect?* In general, none of the challenges addressed can be classified as immediate blockades to the implementation of AI judge. From a juridical perspective, challenges arise due to the lack of both data as well as laws suitable for a robot judge. Additionally, a comparison is done between a robot judge and human judge. This leads to a question on the philosophical position of a judge as well as the lack of transparency to its predictions. Taking this a step further, it does not necessarily lead to the conclusion that the challenges outweigh the possible benefits, to my opinion. It requires a different approach and even a rebranding of the position of the judge (and a courthouse) in general. Rather than the judge being intertwined with the law, it should function as a separate 'entity' that irreverent 'gets the job done'.

# Chapter 6: working towards a successful implementation

## 6.1. Introduction

In previous chapters, the implementation of the robot judge has been investigated from both a technical as well as a juridical perspective. This has led to both potentials as well as challenges that should be considered. In this chapter, all what is said previously will be combined to form a list of criteria and objectives that consider the challenges while not diminishing the potentials. This chapter combines information from both the literature as well as the interviews. The conclusions from this chapter will answer the last sub-question of this research: *on the bases of which requirements should an AI system be developed for a successful implementation within the Dutch courtroom?*

## 6.2. Design requirements

The analysis in previous chapters have all provided insights into how a robot judge would look like. In this paragraph, this information is combined into defining the design requirements for such a system. The previous chapters have led to a list of design requirements that a robot judge should meet to reap the benefits while still overcoming the challenges. The final step is to combine this into the model that considers all what has been mentioned previously. Figure 1 in paragraph 2.4 shows the general idea of a black box system such as the Natural Language Processing. This figure is used as a starting point and is extended in Figure 3. In Figure 3, the *input* has been replaced by *case*, and the *output* has been replaced by *verdict*. The factors within the 'black box' remain 'hidden'.
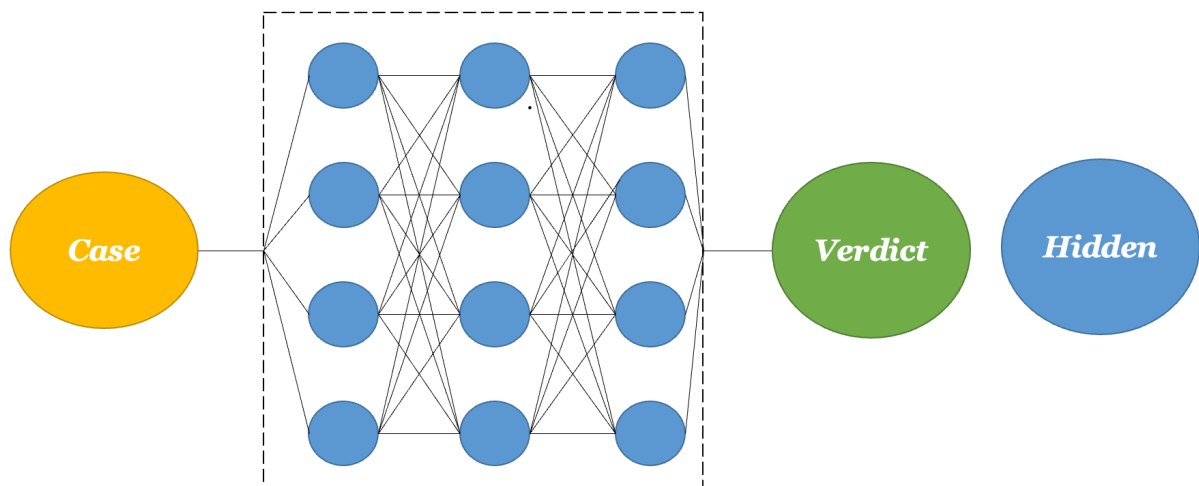
***Figure 3:*** *a high-level overview of a robot judge*

From the analysis in previous chapters, two important requirements can be defined. The first requirement is a system that can decompose cases into words without losing the meaning behind these words: a linguistic system. The robot judge would replace the capability of a human judge to read the case and define the legal problem.

The second requirement is a system that can rule in the underlying case and provide a verdict that is aligned with past verdicts as well as the law. This system is referred to as a machine learning system. In next sub-paragraphs, these two systems will be explained.

### 6.2.1. Linguistic system

If the claimant decides to file the case, the case firstly needs to be transferred to a linguistic system. The linguistic system is responsible for the 'decomposition' of the case and the meaning behind the words. To do so, it is dependent on syntactic rules, semantic rules and lexicons. The syntactic rules look at the way words structure a sentence. It is important to first understand the structure of sentence in order to grasp the meaning behind it. Below are two examples:

> Hans is managing partner.
> Joyce is managing partners.

The first sentence refers to the job title of Hans, while the second sentence puts more emphasis on the word 'is' and as such refers to the responsibilities of the job of Joyce.

Semantic rules are used to structure sentences. The two sentences

>    Anneke paid Joyce for the coffee
>    Anneke paid five euros for the coffee

have the same structure from a syntactic point of view. The first sentence however says something different about Joyce than the second sentence saying about five euros.

Lexicon contains information about the syntactic and semantic roles of different words. As an example, the lexicon would include the facts that "Hans", "Joyce" and "Anneke" are names, "coffee" is a noun and "pay" is a verb. Additionally, it would contain rules such as that "pay" should include a relationship between a buyer, seller or the amount of money. In case one of these are missing, the noun following the verb can be the referent of that information. This will lead to a decomposed case which basically depicts the case linguistically. In paragraph 2.1, the problem of different meanings behind the words "nice object" and "I object" was explained. The semantic rules would cover this, as a subject followed by the word "object" would have a different meaning compared to an adverb followed by the word "object". The linguistic system contains of continuous feedback-loops to ensure that meanings of word are continuous up to date. The word 'exposed' has a totally different meaning in 2022 compared to 20 years ago to give an example.

### 6.2.2. Machine learning

The decomposed case continues to the machine learning system. This system consists of two components: the database model and the verdict generator. The database model is the actual 'black box' system such as the Natural Language Processing. Input from the database model is taken from laws, regulations and decisions. Additionally, input is taken from past cases. The responsibility to 'provide'

laws suitable as input for the system is the sole responsibility of the legislator. Only the legislator can guarantee a correct interpretation of the law because they know the meaning and intention of the different clauses. The database model is a joint involvement of IT engineers and judges, however, only within the judicial system. As has been mentioned in paragraph 3.2, it is crucial to ensure independency of the AI system. The model itself also requires continuous feedback-loops to ensure it is free from bias and the relationships are accurate. Part of the machine learning system is the verdict generator transcribes the outcomes of the database model. Parties involved will know on the bases of what arguments the verdict was generated.

For both the linguistic system and the machine learning system to operate, two additionally 'supportive' functions should be considered as well. These are a legal dictionary and a legal databank which will be explained in the next sub-paragraphs.

### 6.2.3. Legal dictionary

To decompose the case correctly, the linguistic system must also extract knowledge from a legal dictionary. This collection contains information about specific relations, meaning or interpretation. The word 'pay' refers to a seller and payer and that must be a juridical person, either a natural person or an entity for example. Additionally, the five euros paid by Anneke is a legal form of payment in line with article 6:34 Dutch Civil Code. To discharge the obligation to pay, there must be an agreement for example. Is there one in that case? These are questions that can be asked based on the legal dictionary. If questions cannot be answered from the decomposed case, the filer could receive a notification that question cannot be answered (e.g. information might be missing). As such, this will ensure that the decomposed file is complete.

### 6.2.4. Legal databank

The legal databank contains the laws, regulations and decisions that are used in the machine learning system. It is important that this data is outside the machine learning system as the sole responsibility to translate law into code would be for the legislator himself. After all, the legislator knows the (true) meaning behind the words.

Following the design requirements as previously explained, the *robot* judge is born. A graphical illustration can be found in Figure 4. Only after having walked through the system in Figure 4, a case can become a verdict.
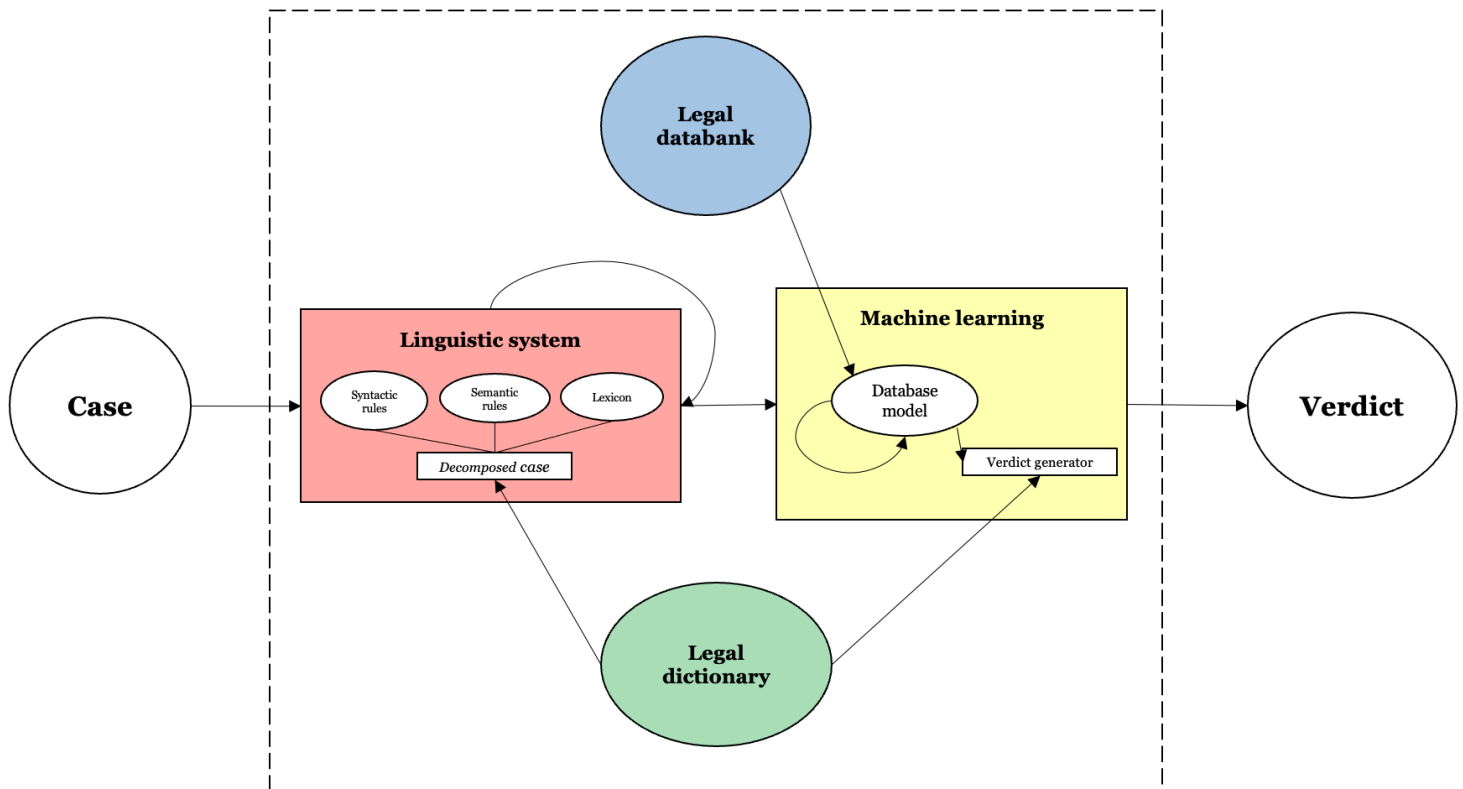


*Figure 4: a graphical illustration of the robot judge. The respective colours depict the responsibility to build that specific component (red = linguistics, blue = legislator, green = judges and yellow = joint involvement of IT engineers and judges). For graphical purposes, the case and verdict do not have a colour as in Figure 1 and 3.*

## 6.3. The seat of a robot judge in the Dutch courtroom

All what has previously been said leads us to the (implicit) question: even if all the requirements as in previous paragraph are considered and an AI system is built successful, what will then be the current 'seat' within the Dutch courtroom.

Both Floris Bex and Corien Prins agree with the use of AI technology as a closed and defined system that should be used in predictable situations (Rechtstreeks, 2019, p.

71-72).[36] In these cases, AI will be able to meet the objectives previously mentioned, especially those of feedback-loops and transparency. The Dutch judicial system consists of three layers of instances. As for the successful implementation of AI *as* the Dutch courtroom, it would require to be implemented only in lower courthouses (Dutch: *rechtbanken*) and not in all types of law. The function of the lower instance is to strictly apply and comply with current (case) laws, as if it mimics current behaviour. If this research has argued for at least one potential of a robot judge, it would be its capability to – perfectly – mimic behaviour. Additionally, these types of cases do not require far-reaching interpretation as for example compared to the Dutch Supreme Court.

AI can function in structured cases, with well-defined problems, that leads to relatively simple answers, also referred to as 'routine' cases. The question that comes up is how to define whether a case is 'structured'. An approach similar to the Estonian courthouse can be followed, in which a monetary threshold is set.[37] Additionally, type of punishment can be set as a condition (e.g. a fine rather than multiple years imprisonment). This is also the reason why AI *as* the Supreme Court would never work – at least not with the current technology. The role of the Supreme Court is to interpret law and provide rationale in usually very unstructured cases.

Additionally, AI could function *as* the judge in cases that involve a preliminary relief (Dutch: *voorzieningenrechter*). These types of cases have in common that they are (usually) well defined problems with more or less straightforward answers. Moreover, the verdicts will lead to relatively simple answers which will function as input data for future verdicts. The reason why a person files a preliminary relief is, most of the time, to have a fast answer rather than an extensive explanation of the law in the underlying case. Also, in terms of transparency, relationships or reasoning are easier to provide because of the way these cases will contain less difficult matters. Graphically, the use of AI in the courthouse can be summarised in Figure 5.

---

[36] prof. dr. Floris Bex is a Professor IT and Law at the University of Utrecht.
[37] As a comparison, the Estonian courthouse is using the robot judge in a pilot for claims up to € 7.000, source: https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/.

|                          | **Unstructured case** | **Structured case** |
|--------------------------|-----------------------|---------------------|
| **Robot judge *in* court** | *A seat*            | *A seat*            |
| **Robot judge *as* court** | *No seat*           | *A seat*            |

***Figure 5:*** *a matrix summarising the use of AI compared to the type of case.*

The system in Figure 4 is showing AI *as* court. However, the system can also be used to define AI *in* court. As a matter of fact, the different components can be conducted by AI *in* court. Laws, regulations and decision can be transformed into a database which will allow easy access for human judges to use in court by for example only looking up specific keywords.[38] The verdict generator together with the legal dictionary can be used to transcribe verdicts in the same way Siri can write messages using voice as input. Additionally, similar to the software COMPAS, a tool could quantify the likelihood of recidivism or relationships relevant in the specific case.

## 6.4. Rules for a correct use of the robot judge

It is worth mentioning that the design as depicted in Figure 4 is not limited to a specific field of law or a specific courthouse. The requirements are set up in such a way that the 'seat' of a robot judge can be at any table - if correctly implemented. To obtain correct verdicts, some rules should be considered as well. These rules are explained in the next sub-paragraphs.

### 6.4.1. Ethical criteria European Commission

In chapter 3, the challenges of AI have been discussed from a technological perspective. One of the challenges has to do with the use of data and the general challenges that come up when using an AI system. In 2019, the EC released a rapport with ethical considerations to use AI on a trustworthy basis. As a starting point, these ethical principles should be followed to ensure a correct implementation of AI in the

---

[38] Moonlit will, starting September 2022, focus less on providing predictions of EU cases. Rather, it will build a legal database of all cases within the European Union in the field of tax and their respective laws in the fields of VAT. It will use this database to analyse the way the different countries interpret clauses within VAT, as this law is regulated across the European countries. Such initiatives show the promising possibilities of an AI system.

courtroom. Therefore, these criteria should be used as starting point for the playing field of a robot judge (European Commission, 2019).

I. **Human agency and oversight:** AI systems should allow human beings to empower, by providing the opportunity to make informed decisions and consider their fundamental rights. This also means that there should be a human-in-the-loop mechanism that allows for human intervention.

II. **Technical robustness and safety:** AI systems should be resilient, robust and secure. There should be a back-up plan in case the system doesn't work, which would minimises unintentional harm.

III. **Privacy and data governance:** it goes without saying that complete privacy must be ensured for parties involved. Additionally, however, adequate governance mechanisms must be considered to ensure quality of data and legitimized access of it.

IV. **Transparency:** data (including the mechanisms) should be transparent. Systems that are built to function for this traceability will lead to a bigger transparency. More important, this transparency also means that humans know the capabilities and limitations of the AI system used.

V. **Diversity, non-discrimination and fairness:** unfair biases that exists within an AI system must be avoided and it should ensure the system to be accessible to everyone, regardless of any's ability.

VI. **Societal and environmental well-being:** AI systems should benefit everyone, including future generations. Therefore, the system must be sustainable and environmentally friendly. This also means that societal impact must be considered.

VII. **Accountability:** AI systems should include systems (or mechanisms) that ensure accountability of outcomes. Moreover, this also means that design of

the system can be controlled for as well as amended if needed. Hence, any outcome should be able to be traced back to the design.

The Dutch courtroom should ensure it is aware of the ethical guidelines and anticipates on them. Additionally, the organisation should be familiar with it (on all the levels). The previously mentioned requirements are criteria for the implementation of any AI system. Nevertheless, additional rules are required for the correct use of AI in the courtroom.[39] The rules in the following sub-paragraphs are the result of the research as explained in previous chapters, including information as taken from the interviews conducted. Additionally, these rules are linked to the rights a suspect has as set out in article 6 ECHR to provide a (partial) answer to the research question.

**6.4.2. Impartiality:** be aware of biases (and adjust)

Impartial courts require judges to have a neutral, open and unbiased point of view when deciding a case. Machines and algorithms don't have biases – in general. An AI system considers input data, rather than parties involved, and provides output in the form of data. However, data becomes a source of information in case a person provides it with a meaning. The part when this meaning is given depends on whether AI functions *in* or *as* court. *In* court, AI will have a supportive role and therefore, judges provide a meaning to the output. This human interpretation can be biased. Output can be interpreted wrong or in a way to confirm ones prior believes: cognitive and confirmation biases. *As* court, AI will decide on the case. Algorithms itself don't have biases, however, the system can be trained on data that has biases. If the training data set is skewed towards a specific factor, such as skin colour, race or gender, the algorithm will mimic this pattern. The biggest challenge for such system is therefore to be free from bias, because it would require that datasets are free from biases: an immense task. The first step is to admit these biases (in general) exists. The next step is to actively monitor for these biases, both on the level of input as well as the output. The last step is to correct where necessary. This last step should be conducted by random, reoccurring, checks conducted by judges in collaboration with

---

[39] Strictly speaking, these criteria are not additional to the requirements as defined by the European Commission. However, they are a more specific interpretation of the criteria of the European Commission for a usage in the courtroom.

IT professionals. When using AI *in* court, not only checks on courthouse A, or judge B should be performed. Similar, verdicts from an AI system should be checked randomly. The system won't fully work unless both cooperate and learn each other's language. In the literature, the implementation of AI is usually conducted around the drawing board, in which people focus on how to conceptualise this system. However, the conversation should be held in practice, in cases in which it acts with the professionals for which the system was built initially (Gillespie, 2016).

### 6.4.3. Transparency: verdicts should be explainable and transparent

It is of extreme importance that verdicts can be explained, whether they are done by a human or robot judge. Whether AI functions *in* or *as* court, the way algorithms have provided the outcome should always be transparent and explainable. The so-called 'thinking-process' should be clear and feasible. The relevant arguments (and data) should be clear so that it can be tested if needed as a check and balance.[40] This will also allow to control for any bias. The algorithm can be a deep-learning mechanism, however, the human (brain) stays in control. To my opinion, this does not mean that all probabilities should be put on the table (which would also allow for law-'planning'). However, criteria and objectives should be provided on the basis of which a verdict is given. It will suffice if these criteria are extracted from legislation or case law. In my opinion, what matters is that the verdict provides an explanation to the verdict, not to the system. Similarly, a verdict from a human judge provides an explanation to the verdict, not to his or her brains.

### 6.4.4. Fair trial: Data should (really) be anonymised

As mentioned in paragraph 3.4, one of the opportunities for an AI system is the fact that it can interact with different systems and can provide opportunities in terms of truth-finding. A downside to this opportunity is, however, the fact that information can be linked to other sources of information and provide insights. Currently, names and personal information of parties involved are anonymised in verdicts. However, this way of anonymising is not enough – at least not anymore. Using open data, a verdict (and its data) can be linked to other sources of data. If you think about it, the

---

[40] The same checks and balances also provide the right to the suspect to appeal to the verdict. Effectively, the suspect should be able to receive on which arguments this appeal should be based.

amount of data that is available in a verdict is huge, including details only related parties know. If the Council proceeds, which in my opinion it should, with the ambition to increase published case law from 4% to a relative high number, how much *possible* data would be on the internet? Therefore, there should be an adequate form of protection by ensuring data is really anonymised.

### 6.4.5. Fair trial: create continuous feedback-loops

AI system should be equipped with the continuous ability to learn from feedback-loops. These feedback-loops will allow the system to continuously improve and implement new laws. Without the ability to adjust 'the settings', humans (and judges) will never be able to fully trust the system and let it unleash its potentials. It was this rules that was lacking or missing in the Dutch childcare benefits scandal as explained in paragraph 3.3.

This is where the role of the judge, after implementation of the robot judge, is important. They will be able to adjust these settings. This can only be done if the AI system is able to continuously provide (visual) evidence of the data and processes so users can reflect and improve the system.[41] It is important that these feedback-loops are used for action, not just insight. Feedback-loops build trust in the AI system and can improve accuracy of the predictions (Snow, 2019). To my opinion, it is of extreme importance to understand (and accept) that the AI system will never reach an accuracy or intelligence of 100%. Neither will a human judge. There is also no rush to implement a system tomorrow. The judicial system can wait until a certain level of accuracy is reached.

### 6.4.6. Fair trial: Learn each other's language

Lastly, and if you ask me most crucial, it is of extreme importance that judges (both robot and human) lean each other's languages. It is one thing to give an AI system to provide an outcome, both in their roles *in* court as well *as* court. It is the other thing to have a judge (or any related party) interpret the outcomes. Especially when it comes to the use of AI *in* court, it requires an open attitude of judges. The system

---

[41] Similarly, it is like cutting a slice of a cake to assess the content and quality and to adjust it based on new laws or improved quality.

aims to help the judges, not to replace them. Additionally, human judges should know how to use the system efficiently. Education and training should be more focused on the use of technological systems. AI is not some sort of 'monster' that is trying to take over the role of the human judge. Rather, it is flower that requires continues plowing and watering (Chorev, 2019).[42] Therefore, a human judge should be involved all the stages of the system and collaboratively work together. The stages of the system can be defined as: creation, implementation, monitoring and education (Hond, 2022). In his research, Hond defined guidelines and criteria for the use of AI tools in healthcare. Similar to judges, doctors face questions on how to use AI tools in their day-to-day jobs and how these tools can be used efficiently. His guidelines can, where relevant, also be used as guidelines for human judges. An overview of the different stages and respective involvements can be found in Figure 6.

| | The involvement of a human judge |
|---|---|
| **Creation** | Stating the conditions, roles and responsibilities of the AI tool. |
| **Implementation** | Involve in configuration, training users and ensuring tool meets the set roles (including conditions). |
| **Monitoring** | Ensure the quality of the tool is meeting the set standards and failures are addressed (and solved). |
| **Education** | Keep knowledge across users to a certain level, including new users. |

*Figure 6:* the involvement of a human judge per stage of the AI tool

---

[42] Compare in this context the use of AI in the medical world. An AI system was developed that could predict the effectiveness of medicine for breast cancer based on genetic material, clinical features and medical history of the patient. The prediction would be presented to the doctor, however, he was sole responsible to make the final decision as to which medicine would be prescribed (similar to the function of AI *in* court). The predictions would deviate from the professional norms and routines the doctors would follow, which led to confusions as to how to interpret the results. Inevitably, because trust in the predictions was low, doctors would leave the predictions aside and would go with their 'own' knowledge. Doctors would have (too) little knowledge on the way the algorithm worked, the used data and how to interpret the result. This led to the algorithm not providing the anticipated policy change(s) (Chorev, 2019).

## 6.5. The robot judge within the current Council

In this research, a special focus was given to the Council. After all, it is their responsibility to make sure courthouses do their jobs and correct verdicts are provided. The Council itself is currently working on two projects to further improve the judicial system. These two projects are the possibilities of litigating digitally and publishing more cases online. A seat of the robot judge would fit within the abovementioned ambitions of the Council. To illustrate this, the following sub-paragraphs will deal with the seat of a robot judge at the Council's table.

### 6.5.1. Litigate digitally

As part of the ongoing project from the Dutch judicial system to litigate digitally, this platform can be used to file a case. In this platform, information on the case such as the relevant parties and the claim itself will be filed all as part of a certain questionnaire. The project to litigate digitally has many potentials. For example, an indication could be given on the chances of 'wining' the respective claim based on previous cases. Additionally, there should be an option to apply for other forms of dispute resolution. Therefore, the position of this would be prior to the 'black box' system. The part of litigate digitally can function as a portal to ensure all information is available just as Gmail gives a notification in case an attachment is referred to in an e-mail but not attached. Also, it can give an indication on the chances of 'wining' the claim or promote other forms of dispute resolution (e.g. as a nudge).

### 6.5.2. External database

The second pilar of the Council is to ensure all cases are published online, by means of a database. This database consists of all the case law that has been published. Before becoming a 'final' verdict, it is important to ensure the verdict is set according to the standards for it to be published and included in the database of case law (e.g. template and language). Additionally, this database can also be used as input for the legal databank. After all, case law is part of official legislation within the Netherlands. Additionally, a database can give the opportunity to ensure the quality of cases is in line with the current status quo. There could be an intentional delay between the verdict generator and the verdict itself to allow for quality assurance. Therefore, this

database, referred to as the external database, would be after the 'black box' system. The word external refers to the fact that the database is outside the machine learning system. The robot judge could function as a tool to make cases available in the external database and to also 'transform' these cases into input for the database on laws and regulations.

Figure 7 puts all the pieces of this chapter together.
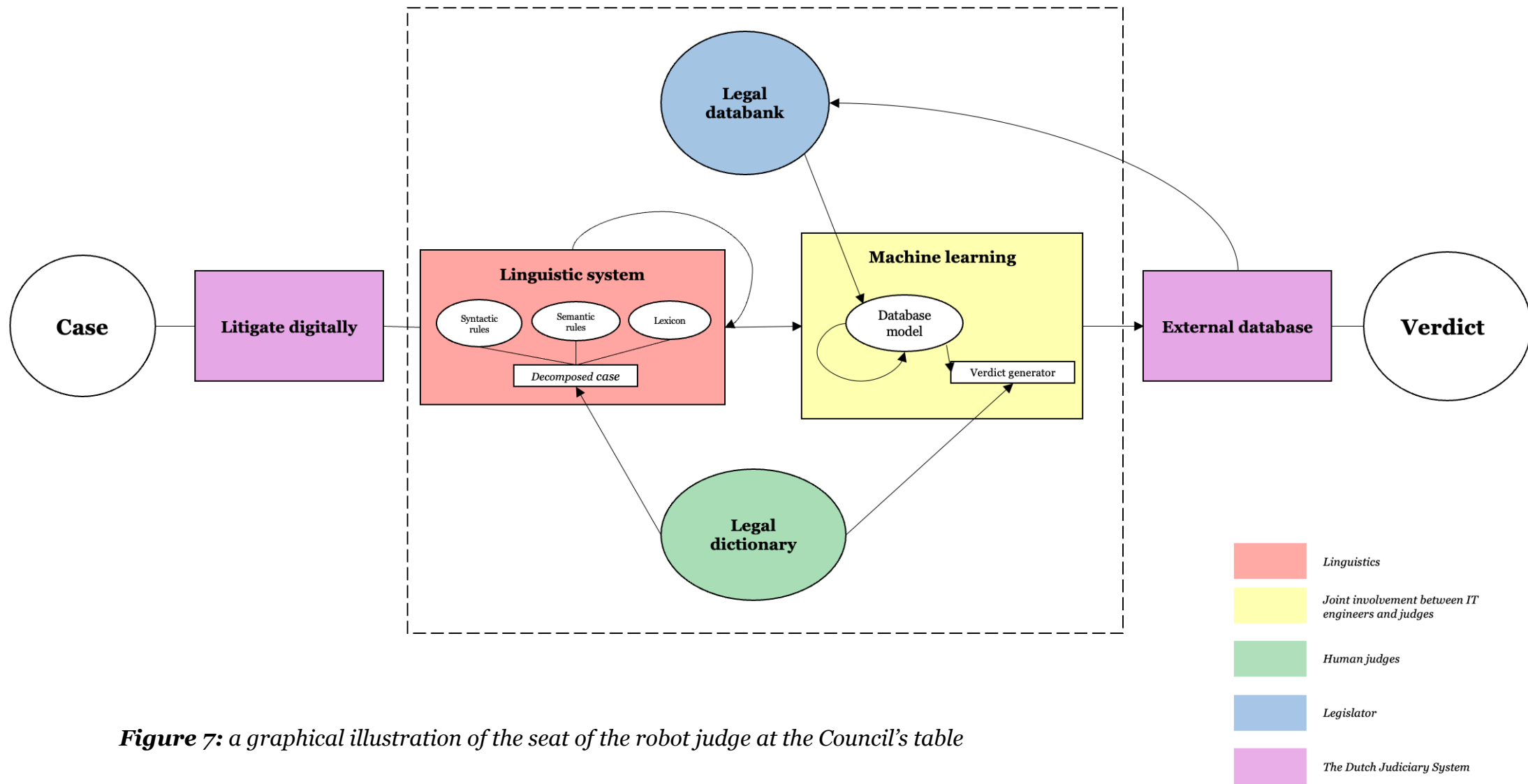
**Figure 7:** *a graphical illustration of the seat of the robot judge at the Council's table*

Case

Litigate digitally

**Linguistic system**

Syntactic rules

Semantic rules

Lexicon

*Decomposed case*

**Legal databank**

**Machine learning**

Database model

Verdict generator

**Legal dictionary**

External database

Verdict

Linguistics

*Joint involvement between IT engineers and judges*

*Human judges*

*Legislator*

*The Dutch Judiciary System*

## 6.6. Conclusion

As part of this research, the final point is to combine all what has been said previously and to propose a set of conditions for which a successful implementation of a robot judge can be guaranteed considering both the potentials as well as challenges of such an implementation. This includes considering the ethical rules as set by the EC. Additionally, a set of five rules are proposed that should ensure maximum potential of AI is achieved whilst not ignoring the challenges that should be considered. This has led to a proposed system architecture which consists of six main components to define the AI *as* court. This doesn't mean the system should be fully operational tomorrow. Each component is defined in such a way that it can be built and used on an individual bases as AI *in* court. The research question deals with whether the robot judge can be implemented in the courtroom. All what has been previously said will be used to form an opinion on this research the question – the final verdict.

# Chapter 7: conclusion

## 7.1. The robot judge: a final verdict

In the literature, the implementation of AI within the courtroom has been a hot topic recently. Many countries embrace to a more or less severe extent the technological potentials and have implemented *a* form of AI. In the Netherlands, the literature has both people in favour as well as against the implementation of AI. However, this has not led to *any* implementation or use of AI within the judicial system. In this research, the implementation of AI has been investigated with the use of the following research question:

> *To what extent can Artificial Intelligence be used in the Dutch courtroom while still maintaining the rights for the parties involved, especially as how has been set out in article 6 ECHR?*

To answer this question, both the literature as well as interviews have been conducted. Both technical and juridical potentials as well as challenges have been addressed to reach to a final verdict: will AI ever have a seat in the Dutch courtroom? Additionally, the robot judge is put into perspective of article 6 ECHR:

> *"In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. Judgment shall be pronounced publicly but the press and public may be excluded from all or part of the trial in the interest of morals, public order or national security in a democratic society, where the interests of juveniles or the protection of the private life of the parties so require, or to the extent strictly necessary in the opinion of the court in special circumstances where publicity would prejudice the interests of justice." – Article 6 paragraph 1 ECHR*

### 7.1.1. A verdict to each sub-question

To answer the research questions, five sub-questions have been used. Each sub-question will be answered individually.

*From a technical perspective, what are the promising potentials of AI and a robot judge within the Dutch courtroom?*

AI systems are, from a technological perspective, able to provide better case law based on efficiency. These systems don't get tired, hurt or hungry, nor are they prone to constant pressure from outside such as corruption, bias and preference. The systems can give an extra dimension to case law by ensuring different courthouses will have similar outcomes and by providing the possibility to quantify chances and relationships. The fact that such a system will operate around a 'black box' system gives it preferential characteristics compared to a human judge: it treats each case as input irrespective of any non-related factors. The numerous amounts of cases and less judges to cover the same workload led to the need of more efficient courthouses. The Council has already set the first stones by providing the possibility to litigate digitally and to publish more cases online. However, more is needed to cope with the exponential growth. Even if AI will not function *as* court, the possibilities in terms of a system *in* court are numerous. It can help publishing cases online, provide insights in the consistent use of law or its effectivity, assist in writing verdicts or help in planning and organization – to name a few examples. With an extensive proven track record across the world, the sub-question should be changed from 'what' are the potentials of AI to 'when' is AI implemented.

*From a juridical perspective, how can AI and a robot judge contribute to an improvement within the Dutch courtroom?*

The conclusion can be drawn that a robot judge is better at ensuring the core values of the Dutch judicial system: independent courts and unbiased and integer rulings. An AI system will be able to provide independent and impartial rulings because it cannot feel, sense or be influenced. It can both remove as well as mitigate social biases and provide insights on the way different characteristics, such as race, gender or sex, have an influence on the outcome of a case. On the other side, because case

law is part of the law of a jurisdiction in general, AI will be able to increase legal equality. The same case will have the same outcomes irrespective of the courthouse or the judge. This will also lead to more legal certainty as open norms will be interpreted similarly across courthouses. The applicability of a robot judge in terms of truth-finding proves how technical initiatives can benefit a juridical system.

*What challenges will an AI system within the Dutch courtroom face from a technical perspective?*

Any new system, in general, will come with a set of challenges that need to be addressed and considered. The way the AI system works, as it is based on human input, provides challenges on the quality of law as well as ensuring no biases are mimicked. Additionally, concerns can be raised on whether an AI system is suitable in all fields of laws especially considering the functions of a courthouse. As a conclusion, the challenges do not stand in the way for a general implementation of AI to begin with. It is however a matter of making sure these challenges are dealt with and so-called 'work-arounds' are created. To take an example, the system should generally have checks and balances to make sure biases within current case law, that is used as input, are not mimicked in the verdicts given. Even if this might be the current case (e.g. case law nowadays might contain biases), the implementation of AI will provide the possibility to have a 'fresh start'. These checks and balances would require random human judges to randomly assess verdicts from a robot judge. Both should be random (e.g. human judges and the verdicts) to ensure that the quality control is integer.

*What challenges will an AI system face within the Dutch courtroom from a judicial perspective, also considering the rights of a suspect?*

Humans will never reach a 100% intelligence level; the question can be asked if an AI system can ever provide that same level of intelligence. Challenges are additionally faced on the position of this AI judge compared to a human judge. Will it be superior, equal or inferior? To give the implementation a fair chance, both judges should learn each other's language. In general, none of the challenges addressed can be classified as immediate blockades to the implementation of AI judge. Challenges are faced

regarding the lack of both data as well as laws suitable for a robot judge. Additionally, a comparison is done between a robot judge and human judge. This leads to the question on the philosophical position of a robot judge as well as the lack of transparency to its predictions. Taking this a step further, it does not necessarily lead to the conclusion that the challenges outweigh the possible benefits, to my opinion. Rather than the judge being intertwined with the law, it should function as a separate 'entity' that irreverent gets the job done. Taking into account the previously mentioned challenges, a robot judge would only reach its potential in structured cases with more or less straightforward answers. This also requires the underlying law(s) to the problem to be structured.

*On the bases of which requirements should an AI system be developed for a successful implementation within the Dutch courtroom?*

Combining all what has been said, a preliminary verdict can be given: AI can have a place in the Dutch courtroom, both in terms *as* court as well as *in* court. To build such a robot judge, it is important it consists of two systems: the linguistic system and the machine learning system. Only if both systems are considered, a successful robot judge can be built. The linguistic will be responsible for decomposing cases filed; the machine learning system is responsible for the *actual* verdict. Additionally, these two systems should be accompanied with a legal databank and legal dictionary as to ensure that the robot judge can mimic or even replace a human judge.

### 7.1.2. The rules for working with a robot judge

It is not just a matter of providing the design requirements. Additionally, rules should be considered when using this robot judge. To do so, the robot judge should be used considering the ethical rules regarding the use of AI as set out by the EC.

Part of this research is also to find out under which rules such a robot judge would operate. To operate the previously mentioned robot judge, some additional rules should be set, also to guarantee the rights as set out in article 6 ECHR.

To ensure the robot judge is impartial, the robot judge should be free from biases. To achieve this, it is important one is aware of the biases and adjust where and if needed. To ensure the robot judge is transparent, its verdicts should be explainable. The tool functions as a 'black box'. Therefore, relationships within the system are not known. Nevertheless, that shouldn't result into a verdict that only shows correlations or causations. It should provide reasoning to its verdict and putting a party involved in a position in which they can verify the arguments, both in law as well as case law. It is important that the robot judge is transparent to its arguments, not the system – at least not to external parties. To ensure the robot judge provides for a fair trial, data should be really anonymised, continuous feedback-loops should be built within the system and both the human and robot judge learn each other's language. It is important that data should be anonymised as the possibilities to link data to other open sources *possibly* compromises the privacy that should be guaranteed for parties involved. Creating continuous feedback-loops will function as a check and balance for the system and ensures that the provided verdicts are correct. Additionally, it would allow the judicial system to improve the quality and accuracy of the AI tool. It is important to stress out that these feedback-loops are not necessarily the responsibility of a human judge. Such responsibility can also be given to IT specialists. Lastly, learning each other language will ensure that the potentials of the tool can be achieved and that (specifically) human judges can benefit from the AI tool *as* court but more specifically *in* court – in which there is relatively more collaboration between the human and robot judge.

### 7.1.3. Examples of AI *in* and *as* court

The possibilities considering previous criteria of a robot judge are wide. *As* court, a robot judge could provide case law at a first instance, with relatively straightforward cases that are capped at a certain threshold for example. This will still guarantee the rights as set out in article 6 ECHR while exploiting the potentials of an AI judge. *In* court, the possibilities are numerous, also in line with the current ambitions of the Council. To name a few examples, it could help filing cases digitally, transcribe verdicts or provide insights into the quality of case law (e.g. compare interpretations at different courthouses).

### 7.1.4. The seat of a robot judge at the Council

The Council is continuously looking for ways to improve the Dutch courthouses, based on the interview with them. Improvements can be achieved in terms of quality of case law, efficient use of resources (e.g. judges) or introducing new tools and initiatives. As part of this intention, two projects are currently (further) being implemented: litigate digitally and publish more, if not all, cases online. The requirements of the robot judge are designed in such a way that it would (practically) fit within these two projects. This would ensure that the robot judge can have a seat at the Council's table

### 7.1.5. What a robot judge can and cannot do

This leaves me to the final remark of this research. The potentials of AI *as* court and *in* court are not a pick and leave: both can be implemented at the same time. What matters is that every judge and the judicial system is ready for the change.

Even though this research is positive towards a robot judge, it does not mean it can replace all functions of the courthouse or all types of judges. The Supreme Court cannot be replaced by a robot judge as it interprets laws and therefore needs human rationale. The process of interpreting the meaning behind clauses in the law and the intention of the legislator is something that requires human intelligence rather than artificial intelligence.

Additionally, it requires the cases to be structured with straightforward, 'simple', answers. That would also require the fields of law to be simple, without too much overkill in terms of open norm and factual interpretation. At last, a robot judge would not – at least not *as* court – be able to replace the human judge in case it replaces a function of the court that requires human involvement. Most straightforward examples are those cases regarding divorces or inheritances.

In terms of the rights to a suspect, such as fairness and transparency, a robot judge would be able to provide fair and transparent court rulings if previously mentioned requirements and rules would be ensured. As a matter a fact, a robot judge would be able to provide more transparent and fair verdicts because interpretation would be

equal amongst the different courthouses and it would not be pressurised by biases or other (external) factors.

## 7.2. Research limitations

This research is limited in the fact that the starting point was a lack of previous research on this specific topic within the Netherlands. A lot of research is done on the use of AI in other fields, but limited sources were available on the use of legal AI. Additionally, there was limited access to data. As part of this research, interviews have been conducted. However, the *invitation list* was more extensive. To name a few, a request was sent out to the Dutch Supreme Court and researchers on the topic of biases and legal. Lastly, this research focusses on the requirements for a design. It would have been a plus too actually receive (a sample of) data from the Council that could be used to build and test the proposed system.

## 7.3. Relevance of this research

This leads us to the academic and social relevance of this research. The relevance from an academic perspective is that this research tries to follow a different approach from the current literature. In the literature, authors mainly focussed on the limitations of AI as to why it could never replace the human judge, for example Prakken (2018) and Sourdin (2018). From the belief that AI could play a role in the Dutch courtroom, research is conducted on what should be considered for a successful implementation and which role this AI judge could have. This research therefore tries to contribute to the current knowledge by providing a list of criteria that should be met to ensure a proper implementation, considering the perspectives from different stakeholders. Additionally, from a multi-actor approach, interviews with different stakeholders have been conducted. The research of this type has not been conducted yet. From a societal perspective, the benefits can be much wider than previously mentioned. Any person could use the foundations of this research to build an AI judge and present this to the Dutch judicial system. Additionally, the purpose of a robot judge is to provide better case law than the current status quo. Therefore, any party involved (e.g. a suspect or the State) will benefit from the correct

implementation as the quality of law will be improved. This will not only benefit the previously mentioned parties but (the Dutch) society in general.

## 7.4. Directions of future research

In terms of directions for future research, the limitation of this research could be used as a starting point. From a theoretical perspective, extensive research (in terms of interviews with different stakeholders) could be held. As an example, no perspectives from a judge on an aggregate level have been used. What would the attitude of human judges be towards a robot judge? Additionally, specific projects and software tool can be analysed in terms of how they would function within the Dutch judicial system, or even the European system for that matter. From a more practical perspective, an actual model could be built in collaboration with the Council. From a personal belief, it is important that such research does not focus on what the robot judge *cannot* do, rather it should focus on what it *can* do. On a personal note, I would have loved to get insights from the Dutch Supreme Court on the use of AI *in* and *as* the Dutch courtroom.

As a final verdict, the robot judge is here to stay (or sit) – if you ask me.

# Literature

Akselrod-Ballin, A., Chorev, M., Shoshan, Y., Spiro, A., Hazan, A., Melamed, R., & Guindy, M. (2019). Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*, *292*(2), 331-342.

Ashley, K. D., & Brüninghaus, S. (2009). Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, *17*(2), 125-165

Becker, D., & Ferrari, I., 'VICTOR, the Brazilian Supreme Court's Artificial Intelligence: a beauty or a beast?'

Belchev, B. (2022). *Recommendations on bias: detect, mitigate, repeat*, University of Twente.

Bobrow, D. G., & Raphael, B. (1974). New programming languages for artificial intelligence research. *ACM Computing Surveys (CSUR)*, *6*(3), 153-174.

Bohnet, I. (2016). How to take the bias out of interviews. *Harvard Business Review*, *16*, 1-5.

Buchholtz, G. (2020). Artificial intelligence and legal tech: challenges to the rule of law. In *Regulating artificial intelligence* (pp. 175-198). Springer, Cham.

Bush, V. (1945). As we may think. *The atlantic monthly*, *176*(1), 101-108.

Chary, M., Parikh, S., Manini, A. F., Boyer, E. W., & Radeos, M. (2019). A review of natural language processing in medical education. *Western Journal of Emergency Medicine*, *20*(1), 78.

Chen, B., Marvin, S., & While, A. (2020). Containing COVID-19 in China: AI and the robotic restructuring of future cities. *Dialogues in Human Geography*, *10*(2), 238-241.

Chiang, M. (2017). 'State Courts and A*STAR's Institute for Infocomm Research (I²R) Collaborate to Develop Real-time Speech Transcription System for Use in Courts' (State Courts, Singapore, 14 December 2017)

Crootof, R. (2019). "Cyborg Justice and the risk of Technological-Legal lock-in". *Columbia Law Review*, *119*(7), 233-251.

D'Amato, A. (1977). Can/should computers replace judges?. *Georgia Law Review*, *11*, 11-36.

Deeks, A. (2019). The judicial demand for explainable artificial intelligence. Columbia Law Review, 119(7), 1829-1850.

Dieterich, W., Mendoza, C., & Brennan, T. (2016). COMPAS risk scales: Demonstrating accuracy equity and predictive parity. *Northpointe Inc*, *7*(4).

van Duijneveldt, I., Wijga, P., & van Reisen, K. (2017). Werkdruk en organisatieontwikkeling in de rechtspraak.

European Commission (2019). Ethics Guidelines for Trustworthy AI, accessed on 30th of August 2022 on https://ec.europa.eu/digital-single-market/en/news/ethicsguidelines-trustworthy-ai.

Gillespie, T. (2016). # trendingistrending: When algorithms become culture. In *Algorithmic cultures* (pp. 64-87). Routledge.

Greene, D., & Patterson, G. (2018). Can we trust computer with body-cam vidio? Police departments are being led to believe AI will help, but they should be wary. *IEEE Spectrum*, *55*(12), 36-48.

Gulum, M. A., Trombley, C. M., & Kantardzic, M. (2021). A review of explainable deep learning cancer detection models in medical imaging. *Applied Sciences*, *11*(10), 4573.

Hogan-Doran, D. (2017, September). Computer says" no": Automation, algorithms and artificial intelligence in Government decision-making. In *Judicial Review: Selected Conference Papers: Journal of the Judicial Commission of New South Wales, The* (Vol. 13, No. 3, pp. 345-382).

de Hond, A. A., Leeuwenberg, A. M., Hooft, L., Kant, I. M., Nijman, S. W., van Os, H. J., ... & Moons, K. G. (2022). Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *npj Digital Medicine*, *5*(1), 1-13.

Kabir, M. (2019). Does artificial intelligence (AI) constitute an opportunity or a threat to the future of medicine as we know it?. *Future healthcare journal*, *6*(3), 190.

Katsh, M. E. (1995). *Law in a digital world*. Oxford University Press.

Katyal, N. K. (1998). Judges as advicegivers. *Stanford Law Review*, 1709-1824.

Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PloS one*, *12*(4), e0174698.

Katz, P. S. (2014). Expert Robot: Using Artificial Intelligence to Assist Judges in Admitting Scientific Expert Testimony. *Albany Law Journal of Science & Technology*, *24*(1), 19207.

Kauffman, M. E., & Soares, M. N. (2020). AI in legal services: new trends in AI-enabled legal services. *Service Oriented Computing and Applications*, *14*(4), 223-226.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, *10*, 113-174.

Locke, S., Bashall, A., Al-Adely, S., Moore, J., Wilson, A., & Kitchen, G. B. (2021). Natural language processing in medicine: a review. *Trends in Anaesthesia and Critical Care*, *38*, 4-9.

Luo, B., Feng, Y., Xu, J., Zhang, X., & Zhao, D. (2017). Learning to predict charges for criminal cases with legal basis. *arXiv preprint arXiv:1707.09168*.

Malek, M. (2022). Criminal courts' artificial intelligence: the way it reinforces bias and discrimination. *AI and Ethics*, *2*(1), 233-245.

Marr, B. (2018). How is AI used in education--Real world examples of today and a peek into the future. *Forbes Magazine*, *25*.

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, *28*(2), 237-266.

Metz, C. (2019). We teach AI systems everything, including our biases. The New York Times, 11.

Monaro, M., Maldera, S., Scarpazza, C., Sartori, G., & Navarin, N. (2022). Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models. *Computers in Human Behavior*, *127*, 107063.

De Moor-van Vugt, A. J. C. (2014). Uitgespeeld? Over de rol van de Raad van State in de Nederlandse bestuursrechtspraak. *De hervorming van de Raad van State*, 297-318.

Moore, L. D., & Padavic, I. (2011). Risk assessment tools and racial/ethnic disparities in the juvenile justice system. *Sociology Compass*, *5*(10), 850-858.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544-551.

Nillier, E., 'Can AI Be a Fair Judge in Court? Estonia Thinks So' (Wired, 25 March 2019)

Ojo, A., Mellouli, S., & Ahmadi Zeleti, F. (2019, June). A realist perspective on AI-era public management. In *Proceedings of the 20th Annual International Conference on Digital Government Research* (pp. 159-170).

Perc, M., Ozer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications*, *5*(1), 1-7.

Prakken, H. (2018). Komt de robotrechter er aan?. *Nederlands juristenblad*, *4*, 269-274.

Prins, C., & van der Roest, J. (2018). AI en de rechtspraak. *Nederlands Juristenblad*, *93*(4), 260-268.

Rivera, I. I. (2020). 'The implementation of new technologies under Colombian law and incorporation of artificial intelligence in judicial proceedings' (International Bar Association, 5 November 2020)

Snow, T. (2019). Three key principles to help public sector organisations make the most of AI tools.

Sourdin, T. (2018). Judge v Robot?: Artificial intelligence and judicial decision-making. *University of New South Wales Law Journal, The*, *41*(4), 1114-1133.

de Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, *36*(4), 101392.

Tai, E. T. T. (2018). Liability for (semi) autonomous systems: robots and algorithms. In *Research handbook in data science and law*. Edward Elgar Publishing.

Tierney, J. (2011). Do you suffer from decision fatigue. *The New York Times, 17*.

Ulenaers, J. (2020). The impact of artificial intelligence on the right to a fair trial: Towards a robot judge?. *Asian Journal of Law and Economics, 11*(2).

Wees, L. (2022). Computers zullen nooit rechtspreken! *Wolters Kluwer 2022, 1.*

Xiao, C., Hu, X., Liu, Z., Tu, C., & Sun, M. (2021). Lawformer: A pre-trained language model for chinese legal long documents. *AI Open, 2*, 79-84.

Xu, Z. (2022). Human judges in the era of artificial intelligence: challenges and opportunities. *Applied Artificial Intelligence, 36*(1), 2013652.

Zheng., X. 2020. Application and regulation of artificial intelligence technology in judicial judgment. *Chinese and Foreign Law* 32 (3):675–96.

# Appendices

## Appendix A

**Interview with Erasmus School of Law**

As part of this research, the first interview I had was with a professor at Erasmus School of Law. The specialisation of this professor was a combination of AI and Law. Additionally, this professor has a role at the Court of Justice in the Netherlands. Among other, recent publications consists of research on AI in healthcare, the perceptions of justice by algorithms and the trade-offs of any implementation at a courthouse. I was referred to this professor through my employer.

On the 21st of September 2021 (11:30AM – 1:00PM), I had a physical meeting at the Tax Law department of Erasmus University. The motive of this meeting was to gather insights on the use of AI as a replacement of the judge, both from an academic perspective as well as from his current role as judge. Specific focus in this interview was whether an AI system was applicable in all law fields, or just specific fields such as VAT or real estate valuation. Additionally, questions were raised on whether the binary choices of an AI system (i.e. zeros and ones) were equal to actual case law. The use of an AI system is favoured, though it should be regulated and well considered. The question was raised, without being answered, whether the role of AI system was to immediately function as a judge or whether certain functions (in between) can be replaced. I was given a specific advice to investigate whether an AI judge will improve legal equality.

The input is (mainly) used in chapters two and three and this interview is not recorded. I had received verbal consent to use the input (anonymously) throughout this research.

**Interview with the Council**

On the 18th of April 2022, I sent an email to the Council to request for an interview. I received a reply on the 20th of April 2022 and planned a (digital) interview on the 4th of May 2022 at 11:30. This interview was held with a senior board advisor Legal and IT at the Council and an intern and took place between 11:30AM – 1:00PM.

Starting point for this interview was a recent (at that time unpublished) article on the question whether AI would provide case law as of 2040. The reference to 2040 was made because that year was mentioned by Jaap van den Herik in his inaugural speech.

The Council is reluctant to whether an AI judge will be able to provide for case law (at least not as a general function). One of the reasons was the Dutch childcare benefits scandal. The argument was provided that an AI judge will be against justice as should be provided by a courthouse. The Council specifically mentions the differences between AI *as* court and AI *in* court, where the argument was raised that AI *in* court is something that could be implemented sooner than later. In this argumentation, a reference was made to the article by Corien Prins (Prins, 2018). To the question whether AI will function as a checks and balance within the judicial system, the Dutch courthouse is not ready for such implementation and we should start small and specific (e.g. a subject in tax). Additionally, we have discussed the different technologies that the Council is implementing (such as providing more cases online). The reason to this is on the one hand to provide more information to society but also to become 'smarter' and provide better case law in the future (e.g. to learn from cases). We have discussed whether the independency will be in dispute if more cases will be published or an AI system will be built outside of the Dutch judicial system. One of the components we have discussed was whether an AI system will be able to replicate the human judge, especially because it lacks *humani*ty. The interview was ended with the following quote:

*"Law is alive, except for the robot judge."*

Their input is used throughout all the chapters and this interview is recorded. I have received verbal consent to use their input in this research.

**Interview with the WRR**

On the 18th of April 2022, I sent an email to the WRR to request for an interview. I received a reply on the 10th of May 2022 and planned a (phone) interview on the 24th of May 2022 between 5:00PM-5:30PM.

The WRR is in favour of implementing AI within the judicial system whether it *as* court or *in* court. Rather than focussing on what AI cannot do, we should focus on what AI can do. In this, a reference was made to one of the articles that discusses the possibilities of supportive roles of AI (Prins, 2018).[43] The WRR specifically mentions that an AI judge is as transparent as any judge because no one can look into the brains of judge. Additionally, according to the WRR, we focus too much on the challenges while we should focus on the potentials of such system, especially because quality of case law will improve. The WRR specifically advised me to investigate how AI will mitigate biases.

The WRR referred me to two professors at the University of Tilburg. Even though numerous e-mails have been sent, an interview with either has never took place.

The input is used throughout all the chapters and this interview is not recorded. I have received verbal consent to use the input in this research.

---

[43] It is the same article that was also referred to in the interview with the Council.

**Interview with Moonlit**

The last interview for this research was conducted with Moonlit. A colleague of my working at the Tax Law Department of Erasmus University also works for Deloitte, the company that is started the pilot of Moonlit. On the 16th of June 2022, between 4:00PM and 4:30PM, I had an interview with the lead of Moonlit.

In this interview, we have discussed the way Moonlit works and how the accuracy of 70% is determined. Even though it can be interpreted as a good accuracy, Moonlit also mentioned that the guessing-rate should be considered. Additionally, Moonlit asked the question who is wrong if the predictive system and real case are not aligned. Furthermore, Moonlit's cases are easier to assess than a complete AI judge because there are only two parties, and more or less two outcomes (i.e. taxpayer wins or tax authority wins). The future of Moonlit, however, is not as a predictive system but as a legal database that will analyse different laws and regulations within the European Union. To do so, it will become a company on its own (i.e. it will not be part of Deloitte anymore). On the question whether an AI judge will become reality in the near future, Moonlit believes AI *in* court provides a lot of possibilities. However, AI *as* court is maybe, at least for now, one step too far.

The input is (mainly) used in chapters two, four and five and this interview is recorded. I have received verbal consent to use the input in this research.
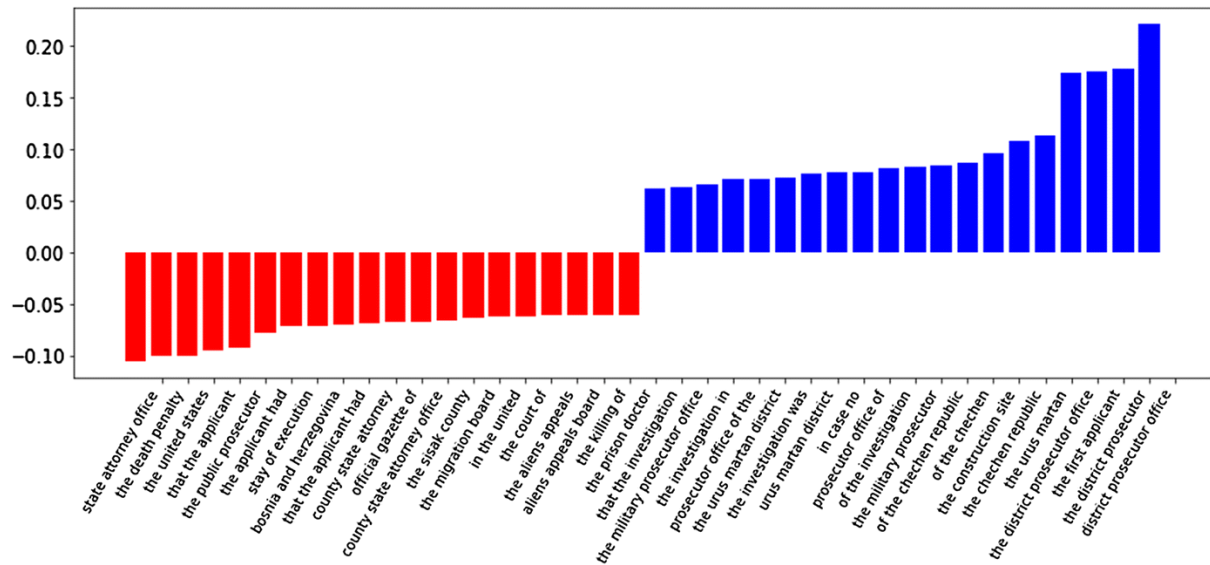
**Figure 8:** predictions on the relationship between words in a case at the ECHR and a violation of a human right, with an accuracy of 75% (Medvedeva, 2020). The colour blue represents a violation and red a non-violation.
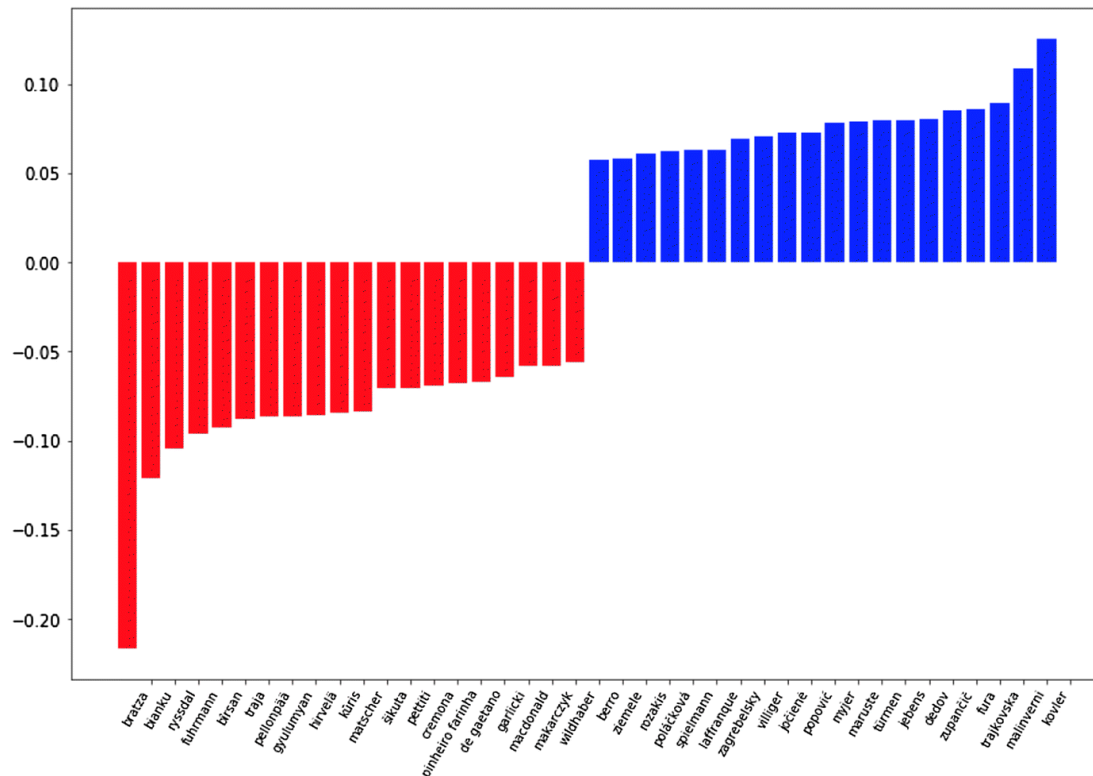
**Figure 9:** predictions on the relationship between the judge and a violation of article 13 ECHR in the underlying case (Medvedeva, 2020). The colour blue represents a violation and red a non-violation.

# Appendix C

**Article 6 ECHR**

**Clause 1**

In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. Judgment shall be pronounced publicly but the press and public may be excluded from all or part of the trial in the interests of morals, public order or national security in a democratic society, where the interests of juveniles or the protection of the private life of the parties so require, or to the extent strictly necessary in the opinion of the court in special circumstances where publicity would prejudice the interests of justice.

**Clause 2**

Everyone charged with a criminal offence shall be presumed innocent until proved guilty according to law.

**Clause 3**

Everyone charged with a criminal offence has the following minimum rights:
(a) to be informed promptly, in a language which he understands and in detail, of the nature and cause of the accusation against him;
(b) to have adequate time and facilities for the preparation of his defence;
(c) to defend himself in person or through legal assistance of his own choosing or, if he has not sufficient means to pay for legal assistance, to be given it free when the interests of justice so require;
(d) to examine or have examined witnesses against him and to obtain the attendance and examination of witnesses on his behalf under the same conditions as witnesses against him;
(e) to have the free assistance of an interpreter if he cannot understand or speak the language used in court.