

EJTN THEMIS SEMI-FINAL D – Judicial Ethics and Professional Conduct

The AI is now in session – The impact of digitalization on courts



Team Hungary

Franciska Zsófia Gyurancz - Bernadett Krausz - Dorottya Papp

Tutor: Bálint Kovács

2019

Table of Contents

<i>Abstract</i>	2
1. Introduction.....	2
2. From digitalisation to AI	3
3. Artificial intelligence in courts	4
3.1. Predictive justice tool “COMPAS” in the US	5
3.2. Predictive justice tool ‘Predictrice’ in France.....	7
3.3. Robot judge project in Estonia	8
3.4. E-courts in the Netherlands	8
4. Compliance with the fair trial principle.....	9
4.1. Tribunal established by law and access to court.....	10
4.2. Independence and impartiality.....	11
4.3. Reasonable time.....	12
4.4. Reasoned judgement.....	12
4.5. Legal certainty and divergent case-law	13
4.6. Equality before the law	14
5. Ethical concerns related to the introduction of AI in the judicial system.....	15
5.1. Transparency.....	15
5.2. Responsibility	17
5.3. Non-discrimination	19
6. Conclusions and recommendations	19

ABSTRACT

This paper seeks to map out the problems imposed by the introduction of artificial intelligence (AI) regarding the judicial system. In the era of digitalisation, we are only a step from using AI in our daily work. For this purpose, we carried out a multidisciplinary research related to the possible application of AI in the judiciary worldwide. Our main objective was not only to show that AI requires a different approach both from legislative and executive aspects, but also to offer possible solutions concerning the implementation of AI in the judicial sector. We analysed several aspects of the right to a fair trial with regard to the changes imposed by AI systems. Likewise, we covered the emerging ethical concerns and tackled the issue of transparency, responsibility and non-discrimination. The analysis gradually verified the initial hypothesis of our paper that in spite of the increasing ubiquity of AI, we have to limit its use and adjust both our legal and ethical framework to address these new concerns.

1. INTRODUCTION

We are all familiar with Pandora¹ from Greek mythology, the woman who opened the box containing sickness, death, sorrow, poverty, toil and many other unspecified evils, and unwantedly released them into the world. Pandora got frightened, slammed the lid of the box, with the only remaining item, hope, locked inside. The world has had to endure a prolonged period of suffering and it cannot expect anything to change unless someone comes along to release hope from its captivity.²

AI is one of the hot topics of the 21st century, and although academic legal writers deal with it with a certain reserve and distance, ‘AI is already pervasive on this planet and will continue to spread, deepen, diversify, and amplify’³. AI is changing the way the right to a fair trial is granted and raises the question how some fundamental legal principles can be adequately applied. At this stage of technological evolution, we can differentiate between automated machine learning systems (such as risk assessment tools), and pure AI (the utopia of an AI judge). Undoubtedly, AI is regarded as one of the future’s most effective devices of legal procedures. The major question to be addressed is whether such novel tools of procedural law, applied differently in each State, will ensure fairness in the field of substantial and procedural law. In other words: is there such a phenomenon as artificial justice?

This paper seeks to explore to what extent fundamental principles are underpinned if decision-making processes involve AI solutions. We also attempt to demonstrate how far some

¹ Image: “Pandora Wonders at the Box”, by Walter Crane, In: Victorian Literature and Culture, 2007, 309–326. (https://www.academia.edu/6787667/pandoras_box_walter_crane_our_sphinx-riddle_and_the_politics_of_decoration).

² William Hansen: Classical Mythology. Oxford University Press, 2005. p. 257.

³ Kevin Kelly, „The AI Cargo Cult – The myth of a superhuman AI” (25 April 2019., https://kk.org/wp-content/uploads/2017/05/Myth_Superhuman_Backchannel.pdf)

countries have travelled on the road to implementing AI in their judicial systems. The final question we address is whether the application of AI resulting in more effective law enforcement will keep Pandora's hope alive to safeguard the fundamental concept of fair trial. Our hypothesis is that, beyond the fundamental rules, the setting up of a wholly novel remedial system of individual rules of procedure may pave the way for establishing human control over (the maze of) algorithms.

To discuss the issue of AI in a few pages is a rather futile endeavour because the outlines of the notion itself are somewhat blurred. The scope of this paper does not allow a discussion and an in-depth analysis of definitions on AI. We will use here terms recognised by the European Commission for the Efficiency of Justice (CEPEJ) as the most often used.⁴

2. FROM DIGITALISATION TO AI

Artificial intelligence is one of the major milestones on the way to full-scale digitalisation. The choice of topic for this paper was encouraged by our belief that Hungary has been producing impressive results in the process of court digitalisation. In Hungary, the National Office for the Judiciary conceived projects for the further digitalisation of courts (Digital Court Project) which aims to achieve full e-administration and offer other smart digital choices with a view to facilitating judicial duties and to be client-friendly.

Digital courts can have different manifestations, and they can be established at various levels. In this regard, Hungarian courts offer (and, in some cases, requires) e-procedures which allow clients to submit their documents electronically and the court to communicate with clients. The courts also accept administration complaints via e-forms and to make court administration even more convenient, there is an electronic notification system to remind clients of delivery of important files and procedural steps. Another level of the digitalisation of courts is the e-filing system, which provides judicial access to any case files online anywhere. The “*Via Video*” project provides courts the possibility of remote audition for faster and cost-effective proceedings, and to guarantee the safety of minors and victims. A specific speech to text program also assists the pursuit of the above goals, while the ruling support system collects judicial decisions, lists e-files and public registers for judges with a single click.

Digitalising does not only affect the day-to-day operation of court, there are several external factors to consider. With the development and the availability of big data, some companies offer the services of judge analytics and provide their customers with information

⁴ “European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment”, European Commission for the Efficiency of Justice (CEPEJ), (Council of Europe Portal, 4 December 2018.)

about judges based on the cases they have tried.⁵ It may help lawyers in building their line of argumentation so that they can file a claim which is more likely to succeed in court, and they can make strategic decisions based on a more specific wealth of information. Such assistive software may help courts in making the justification of decisions.

In 2016, a European team of researchers⁶ developed a model that could predict the decisions of the European Court of Human Rights (ECtHR). They examined decisions regarding torture, degrading treatment and privacy. The model could predict the decisions with a 79% of accuracy given the fact that cases which had not passed the admissibility stage were excluded from the research material. The model examined cases as a ‘box of words’, meaning that it was text-based research. The premise was that on the basis of sufficient data, advanced technology could forecast legal decisions. Although, they were of the opinion that AI was not replacing judges or lawyers, but legal professionals would find AI useful for rapidly identifying patterns in cases that lead to certain outcomes.⁷

In 2014, another team from America⁸ made a similar attempt and developed an algorithm that predicted whether the Supreme Court of the United States would uphold or reverse a decision of a lower court and it could boast a 70 % case outcome accuracy.⁹ Later, in 2017 they reported that in the field of judicial prediction, they had made the first model that had a 70.2 % accuracy at the case outcome level and 71.9 % at the justice vote level.¹⁰ They developed an algorithm that predicted how each Justice would vote, defined the justice vote level and, based on these votes, estimated how the case would be decided (i.e. the case outcome level). This model’s accuracy of prediction could decline when an out-of-the-ordinary case occurred, since the model could only “predict” the outcome by relying on past cases. Additionally, when the decisions showed a pattern of inconsistency, the model could not make accurate predictions, either. It must be noted that none of the above models suggested that AI should replace human judges.

3. ARTIFICIAL INTELLIGENCE IN COURTS

The use of digital tools in courts is becoming widespread all over the world. The so-called ‘cyberjustice’ tools have already been implemented in several European judicial systems

⁵ Robert Ambrogi: In Litigation and Legal Research, Judge Analytics is the New Black. (<https://www.lawsitesblog.com/2015/07/in-litigation-and-legal-research-judge-analytics-is-the-new-black.html>)

⁶ The leader of the team was Nikolaos Aletras, a lecturer in natural language processing in the Computer Science Department of the University of Sheffield.

⁷ Thomas McMullan: A.I. Judges: The Future of Justice Hangs in the Balance. (<https://medium.com/s/reasonable-doubt/a-i-judges-the-future-of-justice-hangs-in-the-balance-6dea1540daaa>)

⁸ Daniel Martin Katz, Michael J. Bommarito II, Josh Blackman, A general approach for predicting the behavior of the Supreme Court of the United States. 12 April 2017., PLoS ONE 12 (4): e0174698. (<https://doi.org/10.1371/journal.pone.0174698>)

⁹ David Kravetz: Algorithm predicts US Supreme Court decisions 70% of time. (<https://arstechnica.com/science/2014/07/algorithm-predicts-us-supreme-court-decisions-70-of-time/>)

¹⁰ Daniel Martin Katz, Michael J. Bommarito II, Josh Blackman (2017)

including tools that facilitate access to justice, improve communication between courts and lawyers, and provide direct assistance for the work of judges and court administration.¹¹ Cyberjustice, however, must be distinguished from predictive justice, which appears more recently at the crossroads of AI and machine learning.¹²

According to a report made by the European Commission for the Efficiency of Justice¹³, the use of AI in the judiciary appears to be less popular in Europe than in the US. However, there are several European jurisdictions which are trialling machine learning applications and predictive justice tools.

Here we give examples of the application of machine learning tools introduced to the judiciary, such as predictive justice tools used in the US and tested in France, and smart system applications implemented in Estonia and in the Netherlands.

3.1. PREDICTIVE JUSTICE TOOL “COMPAS” IN THE US

Using algorithms that estimate the risks of recidivism is increasingly common in the US criminal justice system. According to a research centre in Washington¹⁴, risk assessment tools are used across the States to supply judges with information on pre-trial bail, sentencing and parole, suggesting who can be released at each stage of the criminal proceedings. Despite the fact that these kind of modern risk assessment tools are implemented at all stages of the criminal proceedings they were initially designed to support in certain post-conviction decision such as determination of supervision and planning of the most suitable treatment strategies for offenders (e.g. mental health counselling).¹⁵

One of the most widely used assessment tools in the US is the Correctional Offender Management Profiling for Alternative Sanctions, or COMPAS, a risk assessment software developed especially for courts. Its aim is to contribute to rendering data-centric decisions through an evaluation — based on 137 questions answered by the offender during an interview, and on the information obtained from the offender’s criminal history —, and to ultimately reduce recidivism and increase public safety.¹⁶ Through assessing prior rap sheet and criminological factors such as socioeconomic status and stability, family background, employment etc. the algorithm provides a report including a calculated risk score on a scale of 1 to 10, categorising

¹¹ “Guidelines on how to drive change towards Cyberjustice”, (Council of Europe Portal, 7 December 2016).

¹² “Predictive justice: when algorithms pervade the law”, (Paris Innovation Review, 9 June 2017). (<http://parisinnovationreview.com/articles-en/predictive-justice-when-algorithms-pervade-the-law>)

¹³ “European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment”, European Commission for the Efficiency of Justice (CEPEJ), (Council of Europe Portal, 4 December, 2018).

¹⁴ “Algorithms in the Criminal Justice System”, (Electronic Privacy Information Center <https://epic.org/algorithmic-transparency/crim-justice/>).

¹⁵ Liu Han-Wei, Lin Ching-Fu, Chen Yu-Jie, “Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability”, (International Journal of Law and Information Technology, 20 December 2018), p.5. (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3313916)

¹⁶ Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner, “Machine Bias”, (ProPublica Portal, 23 May 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>).

the offender as high/medium/low risk on three bar charts, namely pre-trial, general and violent recidivism.

The case of Eric Loomis is a good example of the need of cautious use of the COMPAS risk assessment. Several concerns have arisen when the 34-year-old offender, Eric Loomis was sentenced to a six-year imprisonment and a five-year extended supervision partially because of his high risk score of reoffending on all three bar charts as COMPAS indicated.¹⁷

Out of five criminal charges Loomis pleaded guilty to was only an attempt to flee a traffic officer and operate a motor vehicle without the owner's consent, while he denied his role in a drive-by shooting.¹⁸ The circuit court ordered a pre-sentencing investigation report (PSI) which included COMPAS risk assessment score. After having been sentenced, Loomis lodged an appeal arguing that by referring to a COMPAS risk evaluation the court had infringed his right to due process. Firstly, as the software's methodology was a trade secret, he could not assess its accuracy (how the factors weighed), therefore his right to a decision based on accurate information had been violated.¹⁹ Secondly, he argued that the court had deprived his right to an individualised judgment as COMPAS provided a prediction on the basis of a similar data group instead of predicting the specific likelihood of his personal reoffending risk. Finally, he also complained that COMPAS had also took gender into account.²⁰

Upon appeal, the Wisconsin Supreme Court rejected all these arguments. It held that since COMPAS was a proprietary instrument, its mechanism could not be disclosed to the public. Besides, as the software used only publicly available data and information provided by the defendant, Loomis could have reviewed and challenged the accuracy of the information his risk score was based upon. Moreover, the circuit court had based its decision only partly on the COMPAS assessment, as it had taken into account other circumstances as well to provide an individualised judgment. In addition, the court had the discretion to disregard the report if necessary. Furthermore, the Supreme Court found that the defendant failed to provide sufficient evidence showing that the judgment had taken into account gender.²¹

However, the Supreme Court drew attention to the cautious use of the COMPAS risk assessment. It stipulated that this tool cannot be used at the determination of release or incarceration and of the severity of the sentence; therefore judges must give an adequate explanation to the factors other than the assessment taken into account in their decision-making. The PSI containing a COMPAS assessment must involve four cautions for judges²²:

¹⁷ State v Loomis 881 N.W.2d 749 (Wis. 2016), Section 16.

¹⁸ Ibidem, Section 12.

¹⁹ Ibidem, Section 46.

²⁰ Ibidem, Section 34.

²¹ Ibidem, Section 85.

²² Ibidem, Section 66.

- firstly, specifying the ‘proprietary nature of COMPAS’ which characteristic prevents disclosure of how factors are weighed and how risk scores are calculated,
- secondly, despite the fact that COMPAS is based on a national data sample, there has been “no cross-validation study for a Wisconsin population” completed,
- thirdly, “studies have raised questions about whether COMPAS scores disproportionately classify minority offenders as having a higher risk of recidivism”²³, and
- finally, the application of risk assessment tools must be continuously assessed due to their evolving nature.

Additionally, the Supreme Court concluded that courts must consider that COMPAS risk assessment scores are unable to identify particular high-risk individuals as they rely on group data therefore predict group behaviour.²⁴

3.2. PREDICTIVE JUSTICE TOOL ‘PREDICTRICE’ IN FRANCE

Predictive assessment tools are generally used in common-law countries. However, in 2017, two French appellate courts in Rennes and Douai were asked to test a predictive software referred to as ‘Predictice’. Louis Larret-Chahine, one of the founders of the software program, vowed to put an end to the current practice with justice being unpredictable, random and disparate across the country, and said he would move towards “something more logical, scientific or, at the very least, a little more controllable”²⁵. The software was only applied in calculating the amount of redundancy payments in dismissals without a “real and serious cause”, but the stated objective of the predictive tool was to reduce excessive variability in judicial decision-making in order to guarantee the equality of citizens through creating a decision-making tool. Despite the narrowly defined set of cases, the French Ministry of Justice declared the absence of added value of the tested version of the software for the decision-making process of judges.²⁶ The software examined only one part of the court decisions and “it was unable to gauge the subtle differences in the headnotes or to reckon with the compensation awarded under out-of-court settlements”²⁷.

²³ Ibidem.

²⁴ Ibidem, Section 74.

²⁵ “Predictive justice: when algorithms pervade the law”, (Paris Innovation Review, 9 June 2017).

²⁶ “European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment”, European Commission for the Efficiency of Justice (CEPEJ), (Council of Europe Portal, 4 December, 2018), p.42.

²⁷ Roseline Letteron, “Digital Access to the Law”, (Les Annales des Mines, 3 September 2018, <http://www.annales.org/enjeux-numeriques/2018/resumes/septembre/14-en-resum-FR-AN-septembre-2018.html>).

3.3. ROBOT JUDGE PROJECT IN ESTONIA

Estonia is one of the rising stars among Europe's AI players. It made a huge step forward in modernising its government functions by introducing an 'e-residency program'²⁸ and a national ID smartcard²⁹. Furthermore, deep-learning algorithms have automated several government functions. They are scanning, for example, satellite images with algorithms determining if subsidised farmers are following the rules. The CVs of Estonian laid-off workers are also scanned by a machine learning system to find them a job matching their skills.³⁰

Besides deploying a rising number of AI or machine learning systems in government services, the Estonian Ministry of Justice has officially requested Estonia's chief data officer to design a robot judge for presiding over small claims disputes of less than 7,000- euros to clear a backlog of such cases. The project is still in progress, but in theory the two parties will upload all relevant information to the software's database and the AI software will analyse them to render a decision based on pre-programmed algorithms and previous training.³¹ The software's decision would be legally binding but could be appealed to a human judge. The country's economic ministry is considering granting AI and robot judges a legal status to help the legislation allocate responsibility for decision-making involving an AI-controlled software.³² Estonia's effort is not the first attempt to put AI at the service of legislation, though it may be the first one to assign decision-making authority on the basis of an algorithm.

3.4. E-COURTS IN THE NETHERLANDS

On 11 January 2010, the first online private court in the Netherlands was set up offering fully digitalised court proceedings, but its decisions were the products of human reasoning. Since 2011, one specific type of decisions – e-Court judgements adopted by default in debt collection proceedings – are rendered as the sole result of AI.³³

By developing a machine learning application software engineers have found a method how a court case in debt collection can be modelled by identifying the essential key parameters assessed during decision-making. A study set out three different areas in determination of these parameters: the claim (claim amount, due date, interest with a due date, percentage and proportionality with regard to the claim amount, personal data on the parties), the costs of debt

²⁸ It's a status enabling digital entrepreneurs to start and manage an EU-based company completely online regardless their current residency providing access to Estonia's transparent digital business environment. (<https://e-resident.gov.ee/>)

²⁹ Estonia employs a national ID card system used for several services, for instance e-voting and digital tax filing.

³⁰ Eric Niiler, "Can AI be a fair judge in court? Estonia thinks so" (Wired Business, 25 March 2019, <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>)

³¹ Joe Pinkstone, "Estonia creating AI powered judge", (Daily Mail Online, 26 March 2019, <https://www.dailymail.co.uk/sciencetech/article-6851525/Estonia-creating-AI-powered-JUDGE.html>)

³² Ibidem.

³³ Henriette Nakad-Westrate, Ton Jongbloed, Jaap van den Herik, Abdel-Badeeh M. Salem, „Digitally Produced Judgements in Modern Court Proceedings”, (International Journal of Digital Society, Colume 6 Issue 4, December 2015), p.1102.

collection (costs of the writ of summons, court fee of both private and public court, costs of representation in court), and the course of the proceedings (court's competence, defendant's right to be duly notified of the oncoming court proceedings, and to invoke the competence of the public court, observing the court's arbitration rules, whether the defendant appeared in court, and whether the claim should be rejected because of unlawfulness or unreasonableness).³⁴

Examining the characteristics of the e-Court system, the study shows that the robotic judge has not yet exploited its full potentials for two reasons. On the one hand, the Dutch public court system makes only a limited use any of the intelligent system's reasoning methodology in legal decision-making. On the other hand, the types of cases this machine learning system can handle involve neither the weighing of arguments, nor assessment tasks, such as the competence of the e-Court. The latter is performed by human intervention at an earlier stage which means a rather cautious policy in applying new technologies in the court system. Furthermore, after the digital judge has rendered the judgment it must be handed to the public court for implementation. This means that digitally signed decisions are sent to the public court system, where the court's clerk will manually insert the data into the system and recalculate the awarded amounts manually.³⁵ The study also shows that there has not been a single case where clerks were able to improve the decision of the digital judge. However, human errors have occurred during the process of entering the data into the public court's system.

4. COMPLIANCE WITH THE FAIR TRIAL PRINCIPLE

Modern societies have a handful of deeply rooted expectations and concepts of court proceedings. When defining these expectations, our moral compass leads us to the definition of a fair trial (or due process). We all agree on the fact that courts should conduct their procedures in compliance with the right to a fair trial. In this section, we investigate a few selected components of the principle of a fair trial by analysing the possible challenges on the basis of which the necessity of a new approach related to AI systems may be recognised.

Article 6 of the European Convention on Human Rights (ECHR, Convention)³⁶ recognises a right to a fair trial with the core components of a fair hearing, a reasonable time and the independence and impartiality of the court. Although the case-law of the ECtHR gives us detailed guidance on the requirements deriving from these principles, we must bear in mind that the interpretation of the Convention – thus of Article 6 – shall always be subject to present-day conditions. This concept of the Convention as '*a living instrument*' entails that the principles

³⁴ Ibidem, p.1103-1104.

³⁵ Ibidem, p.1108.

³⁶ European Convention on Human Rights

and standards of the ECHR are not static, their interpretation shall reflect on social, economic changes.³⁷

The examples in part 3 of our paper neatly illustrated that the presence and the expansion of AI or other algorithmic decision-making systems are forewarning also from this aspect. We must answer the questions that how the different AI systems will comply with the concept of fair trial, and whether we are facing new challenges. Shall we interpret the fair trial principle in different ways? Will the introduction of an AI judge change how we see the procedural guarantees and the requirements for a judge?

The Council of Europe has recently adopted the European Ethical Charter on the use of AI in judicial systems in which the respect for human rights prevails as a key principle.³⁸ This Charter also provides that we must ensure the guarantees of the right of access to the judge and the right to a fair trial even when AI is applied to provide or assist decision-making.

4.1. TRIBUNAL ESTABLISHED BY LAW AND ACCESS TO COURT

Article 6 of the Convention requires tribunals to be established by law. As far as AI tools are concerned, which assist judicial proceedings and support decision-making processes, the parties or the accused ought to be informed when an ‘AI judge’ will handle their affairs by either assisting in the proceedings, or even making a final decision. In order to comply with this principle, it should be clearly prescribed by law in which cases and to what extent AI systems are used.

It is anticipated that with the help of AI systems more frequently applied in the legal field, more people will gain access to court as the efficiency of courts is expected to improve. In addition, it will be easier and cheaper to seek and receive legal advice. However, the more widespread use of AI shall not exclude any citizen and restrict their access to court. It should be guaranteed that all parties have equal opportunities in bringing their affairs to a judge, and their inadequate digital skills (i.e. their inability to apply AI tools) should not prevent them from doing so. All parties should display a certain amount of IT knowledge regarding AI in order to understand how a device using AI functions. Also, the introduction of an AI judge would raise questions regarding the right to an effective remedy, as the decision might be challenged on the grounds of software dysfunctions as well.³⁹

³⁷ See Case of Tyrer v United Kingdom, Case of Marckx v. Belgium, ECHR

³⁸ European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018), Council of Europe, February 2019.

³⁹ Alina Pastukhova: Artificial Intelligence as a Judge: Can We Rely on a Machine? LLM paper, Ghent University, Faculty of Law, 2017. p.29. (https://lib.ugent.be/fulltxt/RUG01/002/376/194/RUG01-002376194_2017_0001_AC.pdf)

4.2. INDEPENDENCE AND IMPARTIALITY

The independence of courts and judges is based on the theory of separation of powers and adds a high value to public confidence in courts. The principles of independence and impartiality are closely related; for this reason, the ECtHR commonly considers the two requirements in combination.⁴⁰ It has also established a two-limb test on impartiality: a subjective test, where regard must be had to the personal conviction and behaviour of a particular judge, i.e. whether the judge holds any personal prejudice or bias in a given case; and also an objective test, which ascertains whether the court itself and, among other aspects, its composition, offers sufficient guarantees to exclude any legitimate doubt in respect of its impartiality.⁴¹

As far as the independence is concerned, both the appointment of judges and their term of office need adjustment in national law before the introduction of an AI judge. The guarantees against outside pressure require a thorough evaluation as AI systems face different threats than human judges. A human judge can be put under pressure through existential, private and institutional channels. Internal judicial independence requires to be free from instructions or pressures from fellow judges or from court administration.⁴²

Another form of pressure that tribunals are facing is prejudicial publicity. The principle of public hearing is subject to pressure exerted by the press and social media (Twitter, Facebook) which adds another layer to this already complex problem. Courts have to ensure the publicity of proceedings in order to guarantee the freedom of expression, but in case of a virulent press campaign surrounding a trial, the parties and the judges may feel under pressure due to the opinions and expectations expressed in the media. As for an AI judge, we can exclude public pressure as a decisive factor in decision-making as these systems do not take the expectations of the press or the public into consideration.

Furthermore, there is no possibility of promotion or any other kind of remuneration for AI judges which could be used as leverage against them. However, the most common challenge we are facing is AI systems' vulnerability to hackers. These systems need to be sufficiently safeguarded from digital criminals and any outside influence.

Also, these programmes are thought to be unbiased and operating without emotions (as compared to human judges), but we often tend to forget about the possibility of fraudulent use. As it was highlighted in the first part of this paper, human supervision should be established to control these systems at this stage of the technological evolution. The question is whether we should set new rules for the supervisors as well.

⁴⁰ Guide on Article 6 of the ECHR – Right to a fair trial (criminal limb), European Court of Human Rights, p.20. (https://www.echr.coe.int/Documents/Guide_Art_6_ENG.pdf)

⁴¹ Guide on Article 6 of the ECHR – Right to a fair trial (civil limb), European Court of Human Rights, p.45.

⁴² Guide on Article 6 of the ECHR (criminal limb), p.21.

As for the impartiality, people in general tend to have a (indirect) subjective attitude based on their life experiences, or even their daily routine. There are no two humans who would think completely identically, and this difference may be reflected in their choices and decisions as well. However, taking the case of AI, is it feasible to identify any subjective attitude? Or, more succinctly, can AI tools be designed to function more objectively and independently than a human judge? At this point it is impossible to state that an AI judge would be more objective, fair, or impartial in its decision.⁴³ The data sets on the basis of which AI judges work may also be biased and right now there is no widely accepted solution to bypass these biases.

4.3. REASONABLE TIME

One of the procedural aspects of a fair trial is the reasonable time requirement. Judicial proceedings shall be conducted in a timely manner, and judgments shall be delivered in a foreseeable timeframe. To determine whether the length of proceedings can be considered reasonable, all circumstances must be taken into consideration.⁴⁴ In fact, an introduction of automated judgement processes should contribute to the observance of the reasonable time requirement and to fast decision-making. One of the main disadvantages of the ordinary court proceedings is their lengthiness. Judges are overloaded with cases and administration tasks that can be partially passed to automated systems already in place nowadays.⁴⁵

The introduction of AI may improve the efficiency of the judiciary, as it can take over a huge workload and enable human judges to focus on fewer cases. It is predicted that AI judges will shorten the length of proceedings as they can take more cases simultaneously, without fatigue, holidays or other human factors. This is, without doubt, a positive impact on the right to a fair trial.

4.4. REASONED JUDGEMENT

Article 6 of the Convention requests domestic courts to give reasons for judgment in civil and criminal proceedings. The case-law of the ECtHR is quite clear on this matter, and domestic laws also prescribe which decisions need a reasoning and to what extent. Generally, we expect a detailed reasoning for court decisions and that should be the case for an AI judge as well. But in practice, we must take into consideration the limitations of an AI as it is seen as a '*black box*': we send our inputs and we get an output via an unknown process. Is machine reasoning the same as human reasoning, or does it contain an undecipherable process? What explanation can we expect from an AI in a court decision? Some argue that machine reasoning is not more than

⁴³ Pastukhova, 2017, p.36.

⁴⁴ Case of Boddaert v. Belgium, ECHR

⁴⁵ Pastukhova, 2017, p.29.

“algebraically manipulating previously acquired knowledge in order to answer a new question”.⁴⁶

At this point, we must clarify that transparency and explanation are two different requirements. To obtain full transparency, we should know the mathematical algorithms that drive an AI. Here we focus on reasoning and the right to explanation which has been taken up recently in discussions⁴⁷ about the General Data Protection Regulation (GDPR). Articles 13–15 of the GDPR define the rights to ‘*meaningful information about the logic involved*’ in automated decisions which provisions the possible future requirements for AI systems in the legal field as well.

First, we should define ‘*explanation for a decision*’. This concerns the reasons or justifications for a particular outcome, rather than a description of the decision-making process in general. It should be understandable for humans and should give an insight as to which input was determinative or influential on the output. It should include the main factors of the decision-making and provide a reason why two similar looking cases have different solutions.⁴⁸ Regulation of AI reasoning is necessary, and AI should be particularly obliged to give at least the same detailed reasoning for its decision as human judges.

It is important for all reasoning to give a concrete explanation (not describing the general behaviour or process, but the specific decision) and to show counterfactual faithfulness (the outcome should match the factors on the input side). Naturally, it will be a technological challenge to convert all these inputs and types of facts into an AI system without setting biases. In addition, the different decision-making paths of an AI and humans need to be addressed, and the trade-off between accuracy and explanation need to be properly balanced. Some authors argue that to make AI explanations understandable for humans, we may have to sacrifice some accuracy and best solutions.⁴⁹ It is also argued that machines may find the best possible outcome by some undecipherable, mathematical way and identify patterns that humans cannot follow but their design obliges the system to find an eligible reasoning, and therefore an AI may render a decision which is explainable but not the most optimal.

4.5. LEGAL CERTAINTY AND DIVERGENT CASE-LAW

Decisions and proceedings should have a certain stability and predictability in order to obtain public trust in courts. This is part of the principles of legal certainty and the rule of law.

⁴⁶ Bottou, L. (2011). From Machine Learning to Machine Reasoning. (<https://arxiv.org/abs/1102.1808>)

⁴⁷ E.g.: Andrew D. Selbst, Julia Powles: Meaningful information and the right to explanation, International Data Privacy Law, 2017, Vol. 7, No. 4., (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039125)

⁴⁸ Doshi-Velez et al.: Accountability of AI Under the Law: The Role of Explanation, Berkman Klein Center Working Group on Explanation and the Law, 2017. (<https://arxiv.org/pdf/1711.01134.pdf>), p.3.

⁴⁹ Doshi-Velez, 2017.

However, contradictory or inconsistent court decisions may cause uncertainty which may have a chilling effect on public trust for the judicial system.⁵⁰

In this respect, AI-based decision-making could result in improving the legal certainty of tribunals. AI is thought to be objective, reliable and eliminating regional differences and therefore providing a consistent case-law which could help building trust towards the judicial system whilst protecting legitimate expectations as well. However, this concept has its dangers as well. If an AI is stable and delivers the same judgement in similar situations, how can it adapt its decisions to social, economic changes which are often rapid but it takes a considerable length of time to rewrite the relevant codes? We must be cautious as stability does not always coincide with fair or '*right*' decisions. Sometimes case-law needs to be changed, which – to our current knowledge – requires human supervision and intervention.

4.6. EQUALITY BEFORE THE LAW

The current generation of automated risk-assessment tools has the potential to positively impact the rights of '*low-risk*' criminal defendants and offenders to life, liberty, and security of the person. If such tools are more accurate than humans at predicting the risk of recidivism, then low-risk offenders will end up being imprisoned at a lower rate and for shorter periods of time. It is difficult to estimate, however, whether the current automated risk assessment tools have a negative or positive impact on the equality and non-discrimination rights of criminal defendants from groups that have historically been discriminated against, such as ethnic minorities and the mentally ill.⁵¹

The efforts to develop '*fair*' algorithms have not removed the biases yet. On the contrary, even if a ground for discrimination (e.g. race, poverty) is discounted from the data sets these patterns may be reintroduced by other proxies, for example postcode or number of children in a certain size of house.⁵² Bias or prejudice may not be recognised as such by the police when integrated into an automated computer program which is deemed neutral. As a result, bias may become standardised and may then be less likely to be identified and questioned as such.⁵³ Also, a recent study came to a conclusion that a commercial software that is widely used to predict

⁵⁰ Guide on Article 6 of the ECHR (criminal limb), p.44.

⁵¹ Raso et al: Artificial Intelligence & Human Rights: Opportunities & Risks, Research Publication No. 2018-6, September 25, 2018, Berkman Klein Center at Harvard University, p.25.

⁵² A.I. Judges: The Future of Justice Hangs in the Balance. <https://medium.com/s/reasonable-doubt/a-i-judges-the-future-of-justice-hangs-in-the-balance-6dea1540daaa>

⁵³ Algorithms and Human Rights - Study On The Human Rights Dimensions Of Automated Data Processing Techniques (In Particular Algorithms) And Possible Regulatory Implications. Prepared By The Committee Of Experts On Internet Intermediaries, Council of Europe, March 2018. (<https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10>), p.30.

recidivism is no more accurate or fair than the predictions of people with little to no criminal justice expertise who responded to an online survey.⁵⁴

5. ETHICAL CONCERNS RELATED TO THE INTRODUCTION OF AI IN THE JUDICIAL SYSTEM

As presented above, there are steps to introduce automated decision-making structures in the judicial system, e.g. as a tool for risk assessment. We also tackled the issue of the right to a fair trial in connection with AI. Based on our research, we can expect that AI will be present in more and more fields in the judicial system. However, this new era of AI raises a few ethical concerns and predicts a change in the traditional judicial roles and rules of conduct.

To achieve a ‘trustworthy’ AI and address the challenges imposed by automated decision-making (notably the risk of violation of human rights), it is vital to offer a complex framework in which we interpret and discuss AI. Two major components of this framework are the ethical rules and the human rights law. The ethical sub-framework focuses on transparency and accountability while guaranteeing the fairness of the process, whilst human rights law framework offers a holistic and efficient analysis method to tackle the harms induced by AI systems. Some authors argue that international human rights law can define harm, it imposes specific obligations on States and it can be mapped onto the overall algorithmic life cycle from conception through implementation to evaluation.⁵⁵

For the purpose of this paper, within the ethical framework three major factors have been identified and taken into consideration, both before setting up AI systems and also as a follow-up evaluation formula. These factors are transparency, accountability and non-discrimination. These are basic principles of any decision-making process, but any of the AI decision-making techniques can easily violate them. In the following, we will explore the adaptation of the above principles to AI programmes, and will also attempt to offer some solutions.

5.1. TRANSPARENCY

A lack of transparency is also called the ‘black box’ problem. A black box, by definition, is a system whose inputs and outputs are known, but the system by which one is transformed to the other is unknown. We can differentiate between a legal and a technical black box. Due to the proprietary nature of intelligent systems they are legally protected by trade secret statutes. Since they are developed by private companies, their key mechanism is usually kept confidential. We

⁵⁴ Julia Dressel and Hany Farid: The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 17 Jan 2018, Vol. 4, no. 1. (<https://advances.sciencemag.org/content/4/1/eaao5580>)

⁵⁵ Lorna Mcgregor, Daragh Murray and Vivian Ng: International Human Rights Law as a framework for algorithmic accountability. *British Institute of International and Comparative Law* 2019, p.325. (https://www.cambridge.org/core/services/aop-cambridge-core/content/view/1D6D0A456B36BA7512A6AFF17F16E9B6/S0020589319000046a.pdf/international_human_rights_law_a_framework_for_algorithmic_accountability.pdf)

call this phenomenon a legal black box.⁵⁶ While the technical black box feature covers the lack of transparency in AI algorithms, the developers do not disclose the mathematical codes, and users are not trained to understand them. Consequently, here is a key question: even if AI software is used only as a support tool, how does a judge weigh the validity of the software's mechanism if s/he has no understanding of its decision-making process?

Regarding transparency, we propose the following two solutions. In the first scenario, the possible applications of AI and all the potential problems arising from the application of algorithms in the judicial system should be identified. It is also recommended to establish joint teams with several experts (e.g.: process engineer, legal knowledge engineer, data scientist, information expert) from a broad variety of disciplines such as IT, law, sociology, engineering. This interdisciplinary group would be in a position to holistically evaluate the effects on society that AI systems may implicate. Also, the standards should be set worldwide as effective as possible, these joint teams should communicate with each other globally. The precondition of starting a discussion on these topics is that the users understand how these tools operate from data input through processing and computation to prediction output.

Our second proposal reflects on unpacking the legal black box by offering three distinct solutions. Firstly, the definition of trade secrets could be redefined by lawmakers by excluding the mechanism of machine learning from the legal concept of trade secret. Secondly, judges must be empowered to make software development private companies reveal the weighing code of these assessment tools. The third solution could be to put only public agencies in charge of these legal software designed to enable judicial systems to rely on AI in the decision-making process. This last solution is present in Pennsylvania where algorithms developed by public agencies are available for the public to analyse.⁵⁷

Our general recommendation to cut down on the risks arising from the black box phenomenon would be that parties should have the right to be fully informed of the effects of the software and therefore be able to opt out of automated decision-making. Similar efforts are emerging worldwide. The UK House of Lords AI Committee highlighted the importance of a full and satisfactory explanation for the decisions made by an AI system which could have a substantial impact on an individual's life. The European Parliament also tackled the issue of disclosure on its 2016 report on AI, while Articles 13-15 of the GDPR establish the right to obtain 'meaningful information about the logic involved, as well as the significance and the

⁵⁶ Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability. *International Journal of Law and Information Technology*, Forthcoming, 2019, p.17.

⁵⁷ Pennsylvania Commission on Sentencing, Risk Assessment Project II Interim Report 2 Validation of a Risk Assessment Instrument by Offense Gravity Score for All Offenders, February 2016. (<http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-ii-reports/interim-report-2-validation-of-risk-assessment-instrument-by-ogs-for-all-offenses-february-2016/view>)

envisaged consequences of such processing for the data subject'. The GDPR also addresses the opt out mechanism in its Article 22 (1) which enables the data subject to opt out of the use of automated decision-making by providing 'the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'.

The existing use of AI already has controversial effects that warn us to be precautionous. Consequently, in these cases we should place a moratorium on the extensive application of AI solutions while preparing their legal and technical framework to give the opportunity for courts to create rules governing how these tools should be applied and supervised in practice. In the meantime, policy makers could set standards and educational organisations could start to teach how lawyers can handle AI based systems. Education is key in this aspect, as all involved parties and operators should be able to understand the technological background in order to trust automated systems.

5.2. RESPONSIBILITY

Even with the framework set above, there are questions regarding responsibility in automated decision-making processes. At this stage of technological progress, AI systems are only mimicking human decision-making; they cannot understand the social-ethical dimensions in rendering a decision. The question then arises: who would be responsible when a decision is supported or made by an automated system. Would it be the judge, the lawmaker or the system developer? Who is responsible for the consequences of intelligent machine (dis)functioning? Also, low accuracy of AI can be traced back to insufficient data sets which induce bad weighing processes and poor connections between input and output data.

If we implement automated decision-making such as risk assessment tools designed by private companies into judicial decision-making processes, we should be aware that by doing so, public powers are somewhat delegated to private companies. By allowing them to design the data sets and algorithms, we also empower them to influence decisions affecting such fundamental human rights as e.g. the right to freedom. In this scenario, it should be clear on what legal basis those public powers can be exercised by private actors, if at all.

Setting standards in law has never been perfect. Humans make mistakes, but over time and with practice we accumulate knowledge on how to avoid errors – e.g. by constantly refining the system.⁵⁸ As AI only mimics the human mind, we may lose this attribute in transition and even miss out the ability of courts to shape the law. H.L.A. Hart said that 'machines may know what it is, but not what ought to be', which requires a broader understanding of the world,

⁵⁸ Christopher Markou: Why using AI to sentence criminals is a dangerous idea. (<http://theconversation.com/why-using-ai-to-sentence-criminals-is-a-dangerous-idea-77734>)

meaning that they can apply the law but cannot push it forward.⁵⁹ Could or should AI systems be using equity and make discretionary decisions even when they lack a moral-ethical backbone? Decisions should be made accurately to treat individuals with dignity while preserving the principle of individualisation. We should also take the trust of parties into consideration. By using AI, we might be eliminating the human touch from the process, thus undermining the trust of the parties in the judicial system.

What we suggest is that by drawing a line between discretionary and non-discretionary decision-making we can control to what extent AI should be included in these processes. On the one hand, we consider non-discretionary cases as the territory where AI tools can be used excessively as seen in Estonia or in the Netherlands. On the other hand, discretionary matters when rendering a judgment necessarily implies a weighing exercise, cannot be decided on the sole basis of AI therefore they should be a subject to human supervision.

Based on the above, we recommend distinguishing three categories from the aspect of the implementation of AI tools. The first category should be a red zone for AI, the second where the usage of AI is limited, while in the third category AI is allowed without any restriction, although the three key aspects *transparency*, *equality*, *responsibility* must be guaranteed in all three scenarios. In our interpretation, the red zone covers all cases where the right to freedom can be infringed. In criminal justice, bail decisions, pre-trial decisions and sentencing should be dealt with a case-by-case approach taking into consideration the individual characteristics of both the case and the defendant. In the second category, the use of AI in the judicial system is present and should be encouraged but only with limitations. The AI should not be the sole basis of decision-making but only used as an aiding tool. This category covers post-conviction decisions such as determination of supervision, treatment strategies (mental health/drug/alcohol counselling) which have been the original target of the COMPAS risk assessment tool. For the third category, we refer to decisions based on non-discretionary factual background such as small claim disputes or debt collection proceedings.

Apart from the technological preconditions, when deciding to what extent we should use AI in the judicial system, we must consider that making decisions is not the only task of judges. Judges also manage their cases and adhere to the court system. They enable a human framework for courts beside being role models and serving educative purposes for the public. Therefore, when establishing AI systems in courts, not only the question of responsibility is at stake but

⁵⁹ Kaveh Waddell: Can judging be automated? (<https://www.axios.com/artificial-intelligence-judges-0ca9d45f-f7d3-43cd-bf03-8bf2486cff36.html>)

also ethical questions, to decide upon the level of discretion and supervision held by the judiciary.⁶⁰

5.3. NON-DISCRIMINATION

The principle of non-discrimination is deeply rooted in our judicial system. However, when using AI tools, we need to put this requirement under more scrutiny. The idea behind AI-operated risk assessment tools is that if computers can accurately predict which defendants are likely to commit new crimes, the criminal justice system can be fairer and more selective about who is imprisoned and for how long. These tools have a massive influence on judges, and after reading the risk score they may have predetermined views of or even bias against the defendant.

In this aspect, AI systems pose a fundamental problem. The data they rely on are collected by a criminal justice system in which race makes a big difference in the probability of arrest. COMPAS may also have the unintended consequence of framing sentences around recidivism risk in a manner that leads judges to place greater emphasis on incapacitation as a goal of sentencing.⁶¹ In this manner, algorithms have the potential to distort the values underlying laws and policies that (in principle) society has collectively determined to be fair, and without a proper democratic control.

Considering these problems, we suggest that before adopting risk-assessment tools in the judicial decision-making process, jurisdictions should demand that any tool being implemented need to undergo a thorough and independent peer-review process. We should also understand the limitation of AI systems: it is difficult to construct a score that does not include items correlating with race, poverty, joblessness and social marginalisation. They are supposed to be used primarily to determine which defendants are eligible for probation or treatment programs. It would also be essential to pass on the codes of these tools to researchers to evaluate their techniques in terms of internal racial bias.

6. CONCLUSIONS AND RECOMMENDATIONS

As noted above, many questions regarding transparency, responsibility and non-discrimination remain to be discussed by policy-makers, legislators and judicial stakeholders while the extensive use of automated decision-making applications can be predicted. Our research has indicated that, due to the technological revolution in the IT sector, the use of intelligent systems in the judiciary is a necessity. Therefore we propose the following set of recommendations:

⁶⁰ Tania Sourdin: Judge v. Robot? Artificial Intelligence and Judicial Decision-Making. UNSW Law Journal Volume 41(4), p.1133.

⁶¹ Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner, "Machine Bias", (ProPublica Portal, 23 May 2016).

1. An interdisciplinary framework with both legal and ethical dimension should be set up not only to tackle all possible ethical concerns raised by the introduction of automated systems to the judiciary but also to improve our understanding of the operation of these processes.
2. International discussions must be encouraged at both EU and international level in order to identify and adopt 'best practices' regarding the possible use of AI in the judicial system.
3. Such discussions should be supported by the setting up of joint teams with several experts to holistically assess the fundamental aspects of the implementation of AI systems.
4. AI tools to be used should be tested in order to guarantee independence, reliability and non-discrimination before applying them in the judiciary.
5. AI software developers should be obliged to reveal the weighing code to the interested parties or AI applications in the judiciary should be exclusively developed by public agencies. In the latter case, the methods applied should be excluded from the legal protection of trade secret law.
6. Parties should be fully informed about the core characteristics and impacts of using AI tools at the very beginning of their proceedings. Once this phase is completed, parties should be enabled to opt out of the limitless application of AI tools.
7. Judges and other legal professionals must be trained to use AI applications with the understanding of their limitations. This could build public trust in the 'new era' of the judiciary.
8. In non-discretionary cases, AI tools could be freely used in order to make these proceedings faster, more accurate and efficient. Human supervision should be only reserved in discretionary cases. The fundamental right of appeal to a human judge must be granted in both cases.

To sum up, we have revealed some major consequences of the application of AI with regard to ethical considerations and basic fundamental rights. In our view, a comprehensive EU regulation in this field is essential. Beyond the obvious ethical issues there are economic considerations to be taken into account as the development of AI is stimulated by the economy, which makes the subject a key market issue. Besides a basic coordination of fundamental rights there is a need for novel procedural instruments which allow for a human control over both the strong and the weak forms of AI. For we shall not forget: hope still held captive in Pandora's box may only be freed by a human hand.