Judicial AI: AI Judge V. Judge's AI Supporter
A narrative literature review

Monaya Nasser Eldeen

Abstract

Artificial Intelligence is already taking place in many uses in the legal field. Thus, it already shows its skills and superiority; it is fast, cheap, efficient, and progressing by its own. Such characteristics are exactly what the loaded judicial system needs. However, before rushing to unconsciously integrate AI functions in the judicial decision-making process, it is necessary to consider its shortcomings, limitations, and ability to deal with the certain complications and unique issues of the judicial role. Further, considering the high influence of the judicial decisions and judgements on people's lives and the whole society, the principles required from an AI system involved in the decision-making process should be clear and set. Hence, AI should be used in accordance with its trustworthy capabilities, in a way that ensures the principles set. This paper is written as a narrative literature review. Depending on total 6 previous papers, it discusses the challenges and the requirements to integrate AI in the judicial decision-making process and suggests way to deal with these challenges.

Methods: a search has been carried out on the database Scopus, by searching the terms "AI in court", "Judicial AI", "trustworthy AI", and "AI for justice". 73 documents were found, which were checked according to the following criteria: (a) published since 2018; (b) available as full text in English; (c) general discussion (not specified to the law of a particular country or to one type of judicial system). In this way, 6 documents were selected and used to write this paper.

Introduction

In the last few decades, huge headway towards technology is seen. Among others, the legal world has also adopted many innovative technology-based practices, that are potentially aiming to replace a wide range of human activity in the legal system.[1]

[1] See the first section of this paper; Zichun Xu, *Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities*, in: Applied Artificial Intelligence, Vol. 36(1), 2013652 1025, 1027 (2022).

Yet, integrating AI in the judicial system arises serious challenges and issues that require awareness and consideration. Although AI systems nowadays suggest diverse functions and have been proven as successfully fulfilling them,[2] yet it is still risky to the high standards of the judicial system; where cases are complicated, and mistakes have effects on people's lives. Furthermore, the judicial role requires more than legal knowledge and abstract information. Judges are required to observe the parties' behavior and motives, in addition to the requirement to determine in their trustworthiness and the reliability of evidence. Moreover, the judge's role carries responsibilities on the social and educational values aspect.[3]

These characteristics of the judicial role are essential to ensure justice, stable judicial system, fundamental rights, and desirable social values. However, noticing such factors and considering them in the decision-making process needs human skills and intelligent rather than mechanical capabilities.[4]

Notwithstanding the above, AI systems have developed abilities in data storage, data organization, velocity, low costs, and decision-making in simple cases.[5] All these can be helpful to decrease the human judge's mistakes and workload. However, in which way should these abilities be integrated in the judicial decision-making process? And to what extent should AI activity be involved?

Moreover, even in such simple uses, AI system should be trustworthy. For AI to be trustworthy, it should comply to principles, which should be embedded in its function by designing it according to them. Further, a supervision is required in all stages of the algorithmic function, to identify errors, fix biases, and correct mistakes.[6]

In this paper, I have discussed the challenges of integrating AI in the judicial decision-making process, while dealing with the need and the requirements to a trustworthy AI. Moreover, I've suggested some ways to deal with these challenges for fulfilling the desired principles and requirements.

---

[2] Xu, above n 1, p. 1027.

[3] Tania Sourdin, *Judge v. Robot? Artificial Intelligence and Judicial Decision-Making*, in: University of New South Wales Law Journal, Vol. 41, Issue 4, 1114, 1124 (2018).

[4] Xu, above n 1, p. 1040-1041.

[5] Angela Busacca and Melechiorre Monaca, *Using AI for Justice: Principles and Criteria of the "European Ethical Charter on the Use of AI in Judicial Systems"*, in: Marino, D., Monaca, M. (eds) Artificial Intelligence and Economics: The Key to the Future, Lecture Notes in Networks and Systems, Vol. 523, 157, 159 (2022).

[6] Davinder Kaur, Suleyman Uslu and Arjan Durresi, *Requirements for Trustworthy Artificial Intelligence – A Review*, in: L. Barolli, K. F. Li, T. Enokido & M. Takizawa (Eds), Advances in Networked-Based Information Systems, 105, 106-107 (2021).

This paper reviews 6 previous ones, that dealt each one with some relevant perspectives which contribute to my argument as presented below. This paper combines the previous ones to suggest a complete image, including the present reality, the challenges and problems arise, and suggesting a model to deal with these challenges. The present reality reviews the current practices of AI in the legal field; the problems and challenges part argues that the main issue is that AI is not qualified enough to have the authority and responsibility carried by the judge's role; the suggested solution presents a model for optimally using AI in court, according to it the presence of the human being in the judicial system has a superiority position. In addition, it suggests that AI should be integrated in the judicial system carefully in a supervised way, to use its great advantages, and avoid its shortcomings.

The paper is organized as follows. The first section will review the current uses of AI in the legal field. The second section will discuss the challenges facing the development of an AI judge. The third section will deal with the biases and limitations of AI to be just and trustworthy. The fourth section will present the requirements and principles to get just and trustworthy AI. The fifth section will discuss ways to correctly use AI in court, by partial (rather than full) automation, and by specifying the AI use in a way that decreases the gaps between its limitations and skills. The sixth section will compare, according to the conclusions and insights from previous sections, whether a judge robot is what the judicial system needs, or an AI assistant to the human judge, that wouldn't replace him, but would support him.

AI Practices in Court Nowadays

Organizing Information:

AI systems are able to process a huge amount of data and recognize patterns in texts. The USA and the British legal systems adopted the 'eDiscovery' concept, which is "an automated investigation of electric information for discovery, before the start of a court procedure".[7] This practice is based on machine learning, which desires to the ideal algorithm capable of extracting the relevant parts from a huge amount of data. The parties are the ones determining the search terms and coding, supervised by a judge who assesses and confirms their agreement.

The judge role is significantly related to this use, since the judge's role, mainly in the adversary law system involves a flooding of information submitted by the parties in each case, partially irrelevant or unnecessary, and mainly unorganized. Moreover, judgements written in the basis of laws and previous precedents, which may be a huge amount of data. Searching for relevant previous cases and the law application in each

[7] A. D. (Dory) Reiling, *Courts and Artificial Intelligence*, Vol. 11, Issue 2 1, 3 (2020).

one of them may take long time, which is a precious and rare resource for a human judge. By automating information organization, the whole legal proceeding becomes to efficiency and then to justice, as this increases the capability of the judge to focus on the important details, dealing with less distractions, which enables him to make intelligent and well-founded decisions, based on clear refined information rather than chaos of information.

Advise:

This use utilizes the ability of AI to "provide an answer to a question".[8] Machine has the capacity to contain a huge amount of data along with the ability to extract any piece of data in a minimal time. As a result, it can sum up providing a precise and fast advise according to the data given in its base.

Nowadays, this practice serves attorneys by using 'The ROSS project',[9] that offers its advice service for European and American attorneys. ROSS is an AI law assistant trained by previous cases in all law fields. By asking it a legal question or presenting a legal issue, it provides a precise and fast answer, organized according to relevance. Lawyers inter the legal question, the motions, and the facts of the case, and ROSS gives them the right legal answer. In this way they may know which cases are most likely to get one judicial decision or another, to assess the profitability of submitting a motion or avoiding it, and to base their legal arguments on true, relevant, and contemporary judgmental mindset and tendency.

By acting according to the AI advice, people may solve more problems and disputes without getting to the court.[10] Such a reality may decrease the judicial role in two ways, the one is by decreasing the number of cases; and the other is by integrating such a system as an advisor for the human judge, thus the amount of work that he would need to do for each case would be reduced.

One use of the advisory AI is the Civil Resolution Tribunal (CRT) in British Columbia, Canada.[11] This is the first Canadian online tribunal, whose jurisdiction is increasing by the years. Today it deals with 4 types of disputes: (1) motor vehicle accidents; (2) small claims up to 5,000 dollars; (3) strata property; (4) societies and associations. Parties who choose to resolve their dispute by the CRT, conduct their proceeding absolutely without a human judge intervention. Moreover, The CRT is active by law, and thus its decisions are enforceable in court.[12]

---

[8] Reiling, above n 7, p.4.
[9] See https://www.rossintelligence.com/what-is-ai.
[10] Reiling, above n 7, p. 4.
[11] See https://civilresolutionbc.ca/.
[12] Civil Resolution Tribunal Act [SBC 2012] CHAPTER 25.

Courts in New Jersey use an AI system to "calculate the possibility and extent of bail" since 2017.[13]

Courts in Mexico use nowadays the "Mexican Expertius system", which advises to judges when determining whether a plaintiff is entitled to a pension.[14]

Furthermore, the "206 system", used in Shanghai courts, compares each decision given by a judge to the decisions given in the higher court. When the difference is more than 85%, the system sends the judge an alert advising him to take into consideration the higher court decision. In case that he insists on his decision rather than considering the higher court one, the system sends the case to the court president for supervising and discussing its propriety.[15]

In these days, a new field is being developed, named Online Dispute Resolution (ODR). It suggests an internet platform to parties in disputes, enabling them to get to a resolution in an efficient, simple, cheap, and fast way. The platform suggests the parties some optional solutions, in a process based on AI tools and algorithmic decision-making model, according to data given by the parties by answering a questionnaire.

Courts in the Netherlands making use in the "Rechtwijzer" program, that involves ODR elements. This AI-based program is being used by couples for their divorce or separation processes. According to information given by the parties about their relationship, their history, and difficulties, the program suggests them optional resolutions, links to relevant information and helpful tools and websites. Furthermore, if the parties don't get to a satisfying resolution by the program itself, it provides them information and links to "professional third parties such as mediators, legal representatives, and other dispute resolution processes". In this way, "Rechtwijzer" mostly would cover the whole options enabling the parties to get to an optimal solution with less harms.[16]

The District Court of East Brabant in the Netherlands in collaboration with universities and academic institutions conduct nowadays research which aims to suggest an AI model to be used as a judge assistant in traffic violation cases. The research is based on data from two local District Courts and one Court of Appeal. Yet this study hasn't presented its results or its potential efficiency.[17]

---

[13] Busacca and Monica, above n 5, p. 159.
[14] Sourdin, above n 3, p. 1119.
[15] Xu, above n 1, p. 1039.
[16] Sourdin, above n 3, p. 1121.
[17] Reiling, above n 7, p. 4.

Predictions:

This category deals with the "predictive justice",[18] the ability of AI to predict the decisions of the court in each case according to previous decisions in similar cases.

While the decisions nowadays are given by a human judge, considering the different complexities and information in each case, it is harder and more complicated to an algorithm to precisely predict the actual outcomes of a new case, which has any new element that doesn't exist in previous cases known by the algorithm. Thus, integrating AI systems in judicial decision-making process would deal with this gap by increasing the stability of the decisions given in similar cases, and therefore the accuracy of an algorithm prediction of a decision or outcome would be significantly higher.

One predictive tool is one that has been developed by American academics. It is a machine-learning-based application, which has been proven accurate up to 70.2% in predicting the outcome of a case at the Supreme Court of the United States, and 71.9% for the individual judges' voting behavior. Its predictions are based on the case information and the political preferences and past of voting behavior of the same judge.[19]

Another application is one that has been proven to be accurate up to 79% in predicting whether the judge in European Court of human Rights will rule that there was a violation of a particular provision of the UCHR or not.[20] This algorithm is based on machine learning and natural language processing, and the predictions are based on previous judgements. The way this application works is established on the understanding that the strongest indicator of the outcome of a case is the way that the judge presents the facts of the same case. The algorithm identifies patterns, and thus, it recognizes the predicted outcome.[21]

In the criminal courts of the USA, judges use a tool named "COMPAS". Which has been developed to predict a recidivism of prosecuted defendants or convicted persons. It is based on the defendants' criminal past, life circumstances and facts, in addition to their answers to a questionnaire made up of 137 questions.[22]

A new development in the prediction category is a startup used nowadays in the USA named Ravel.[23] It aims to recognize and analyze patterns in judgements and jurisdictional tendency, which may help to predict the outcome of a case according to

[18] Busacca and Monaca, above n 5, p. 158; Reiling, above n 7, p. 4.
[19] Reiling, above n 7, p. 5.
[20] Reiling, above n 7, p. 5.
[21] Sourdin, above n 3, p. 1125.
[22] Reiling, above n 7, p. 5.
[23] Reiling, above n 7, p. 6.

the contemporary mindset in courts. Deepest information about this startup isn't publicly available, thus the way it works, and the accuracy of its outcomes are not known yet.

France, as well, has developed an AI tool based on previous cases from the whole French courts, in addition to information about the judges, their tendency, habits and opinions. By using this data, it predicts the outcome of a given case.[24]

Table 1: practices conclusion

The practice	What is it used for?	What is it?	In which countries is it used?	Notes
E-Discovery	Organizing information	"An automated investigation of electric information for discovery, before the start of a court procedure"	USA and Britaine.	The parties determine the search terms and coding, supervised by a judge who assesses and confirms their agreement.
ROSS	Advise	An AI law assistant that provides answers to legal questions based on the law and previous cases.	USA and Europe.	This practice serves lawyers.
Civil Resolution Tribunal (CRT)	Advise	An online tribunal that deals with 4 types of disputes: (1) motor vehicle accidents; (2) small claims up	British Colombia, Canada.	This platform is active by law, and its decisions are enforceable in court.

[24] Busacca and Monaca, above n 5, p. 159.

		to 5,000 dollars; (3) strata property; (4) societies and associations.		
Bail calculator	Advise	An AI system that calculates the possibility and extent of bail.	New Jersey.	
Mexican Expertius system	Advise	It advises to judges when determining whether a plaintiff is entitled to a pension.	Mexico.	
206 system	Advise	It alerts the judge when his decision is different than a higher court decision in more than 85%.	Shanghai.	If the judge doesn't consider the alert, the system sends the case to the court president for supervising and discussing its propriety
"Rechtwijzer" program (an ODR tool)	Advise	It is used by couples for their divorce or separation processes.	Netherlands.	According to data given by the parties, the program suggests solutions, links to

				professionals, and relevant information to get to resolve the dispute.
The American academics predictor	Predictions	It is used to predict the outcomes of cases in the supreme court. Moreover, it is used to predict the judges' voting behavior.	USA.	Its predictions are based on the case information and the political preferences and past of voting behavior of judges. It is proven as accurate up to 70.2% in predicting the outcomes of cases and up to 71.9% in predicting the judges' voting behavior.
The UCHR violation predictor	Predictions	It is used to predict whether the judge in the European Court of human Rights will rule that there was a violation of a particular provision of the UCHR or not.	European Court of human Rights.	The predictions are based on previous judgements. It is proven as accurate up to 79%.

COMPAS	Predictions	It is used in the criminal courts to predict a recidivism of prosecuted defendants or convicted persons.	USA.	It is based on a questionnaire and data given about the defendant.
Ravel	Predictions	it recognizes and analyzes patterns in judgements and jurisdictional tendency, to help predicting the outcome of a case according to the contemporary mindset in courts.	USA	
The French predictor	predictions	It is used to predict the outcomes of given cases.	France	The predictions are based on previous judgements and data about judges' tendency, opinions and habits.

Challenges of Developing AI Judge

The automation of judicial decision-making process is known as "Digital Justice".[25] Throughout the years, the notion of "robot judge" has been taking place in the science

[25] Busacca and Monaca, above n 5, p. 159.

fiction and the dreams and desires of so many legal-tech experts. In the current reality, it seems to be closer than ever, though, it still full of risks and challenges to get to the ideal trustworthy model. Experiments on decision-maker AI, while assessing its performance and improving its weaknesses are already taking place and heralding constant progress. Yet, it still far from fully replacing the human judge's role.

An AI system is developed by many factors, each one of them has his own essential role to get the final model. In the algorithmic decision-making process, many factors are involved, such as the model creators and programmers, the model trainers, the model developers, the human judges who have written the judgements used to train the model, the policymakers, the supervisors on the model's function, and the AI system itself. All these factors are responsible for the outcome given by the AI system. Thus, who should have the authority to make such a decision?[26] The answer for this question should be legislated clearly. Such a legislation has an impact also on the accountability question, when dealing with wrong decisions.

Another challenge is the digitalization required to law documents and legal language, transforming them into codes.[27] Legal language is based on semantics and nuances, it is very often context-dependent, thus requires wider linguistic understanding.[28] Words are the most dominant and significant tool for jurists. For that reason, translating legal language into code would be challenging and critical.

Translating legal language into code is the task of programmers and IT experts, who, in most cases, don't have the legal knowledge required to translate precisely. Furthermore, these systems store the whole data back in time, while for it, the whole data stored is always relevant and useful in the relevant cases. However, new precedents are set all the time, as well as changes in laws. The newer precedents are often added to previous ones and doesn't stand on their own, but sometimes they totally replace them. Such changes change the right answer to a legal question. Storing opposite answers to the same legal question may lead to disruption in the algorithm's function and to incorrect results. These diagnoses would be challenging to an AI system; while it requires the system to apply the relevant law on the relevant case, considering the certain point of time the action took place in, and identifying the relevant law to apply on it.

One more challenge is the limited ability of the machine to create decisions based on its own discretion. AI systems are programmed and trained on an existing data, such as past judgements given by human judges. Furthermore, their outcomes are predetermined, while all the possible outcomes are limited to the options derived from the data that the

[26] Sourdin, above n 3, p. 1126-1127.
[27] Sourdin, above n 3, p. 1127.
[28] Sourdin, above n 3, p. 1130.

algorithm has trained on.[29] Though, the judicial decision-making process requires more than that. It requires creativity, understanding the human nature, behavior, motives, and emotions, evaluating social and educational values, and having a sense of justice, conscience, and ethics. All these are desired factors that are being considered in judgements and decisions, yet an AI system can't provide. Thus, discretion is an inherent need for the judge's role; since AI systems are still limited to such abilities, using them for judicial decision making would cause serious damages to the judicial system, in systemic, ethical, and social aspects.

Limitations and Biases of AI

As we aim to integrate the AI systems in "serious" tasks and roles, we need to be aware to the challenging limitations of AI abilities in this point of development. The judge's role has wide implications on the public and the citizen's rights, it bears the responsibility of guaranteeing the citizens' rights by law, and it is the supervisor on the law getting to its crowd. Thus, the integration of AI in the judicial system should be done carefully and mindfully, while decision-maker AI is not yet perfectly precise, and its outcomes show biases and unjust results. "Some forms of AI that are currently in use have already demonstrated that there can be significant risks in using AI in terms of bias and that programmers and others can replicate bias without intending to do so. These issues have suggested that algorithms can produce unwanted results and promote racism and inaccurate outcomes".[30] Hence, indifference to these gaps might undesirably lead us to unjust judicial system, a huge damage to citizens' trust in the judicial system and to their ability to obtain their rights, in addition to the valent harm of actualizing the law to each case getting to the court.

One example of the unpreparedness of AI to be just and trustworthy is "COMPAS" that is used in the criminal law system of the USA. This machine-learning system works by analyzing facial features and information to decide the recidivism chances of a defendant or an arrested person. It has been trained by data from previous cases from the USA courts, as well as data of previous defendants or convicted persons, and their criminal future after their release. After years of use by the police and the judicial systems, COMPAS turned out to be racist towards black people, while it used to tag innocent black people with an "innocent look" as criminals, as well as tagging criminal white people with "criminal look" as innocent ones.[31] On its results base, innocent people have been arrested and convicted, while criminals were released.

[29] Sourdin, above n 3, p. 1128.
[30] Sourdin, above n 3, p. 1129.
[31] Xu, above n 1, p.1035.

These results are caused by a biased data sets given to the machine in its basis. Though an essential insight is that for dealing with this kind of machine bias, data should be neutral, representative, and qualitative. Furthermore, the relevant factors to make a judicial decision should be clear to the algorithm, by programming it in a way that irrelevant considerations, such as the plaintiff or the defendant color, shouldn't affect the decision. Learning machine would conclude that some factors are relevant according to random statistics on the base of previous cases; in this way, in the example above, if black people were convicted in more cases than white people, the algorithm concludes that the defendants' color should be considered and affect the judgement. Thus, more than training the AI system according to a neutral data, consistently confessing is required as well, to ensure that the data given and used in each decision is neutral and unbiased.

Though, human judges are also influenced by their worldviews, backgrounds, beliefs, and their own agendas; even though they are not necessarily aware of that, these biases and effects are undesirably part of the judgements, and thus the results of the cases are undesirably not neutral. These same results are used as the database of the AI system and given to it as the examples to imitate and be directed by. In this ground, training AI systems in this manner, the way to biased judgements of the AI judges is inevitable.[32]

Flaws like these, while trying to improve the efficiency, impair the quality of the results.[33] The legal AI experts has impelled to integrate the current technological practices that shown above (section 1), in the name of efficiency. But this "rushing" comes at the expense of the quality of the results, which are perceived as reliable and trustworthy to those who are unaware of the current undeveloped points of AI. As discussed above, AI is already given to use of judges around the world, while its outcome is biased; the results supervised in retrospect are more often seem to be unjust. One reasonable explanation may be that it is more comfortable to human judges to believe the smart machine than themselves. This obviously leads them to adopt undesired and unjust decisions and is leading the whole judicial system to undesired reality where the lawful principles, the court promise of justice and guaranteeing rights, and the significance of the judicial entity and its characteristics, all may be emptied of content.

The quality of the results is also influenced by the extent the data given to the machine is presentative and balanced.[34] If most of the results in a certain sort of cases given to the machine are showing coincidental sided, the AI system wrongly infers that these sided results are intended, and this is the right way to decide in such cases; which leads the

[32] Xu, above n 1, p. 1036.
[33] Xu, above n 1, p. 1032.
[34] See Xu, above n 1, p. 1032-1033.

machine itself to be biased. This process occurs usually without awareness, and thus the reason of the biased results received by the AI system are not understandable, and even very hard to be fixed. Moreover, the process of data assimilation in the AI system involves dealing with the serious gaps between the machine language and the natural language. ″The polysemy, contextualization, and vagueness of natural language make it difficult for machine language, which is based on grasping core morphemes and semantics through the word vector transformation and word segmentation technology, to comprehensively and accurately recognize and understand complex semantics in cases″.[35]

AI system is also limited in dealing with problems that require to innovate resolutions that don′t grow necessarily from the base program written by its human designer in advance. Though real cases that get usually to the court are complex disputes, which accordingly, need complex resolutions. ″Justice is a very complex mechanism to determine division and settle disputes, involving people, property, time and place, subjective and objective state, behavior mode, involved tools et al″.[36] Getting to an appropriate resolution to such a complexity requires perceptual and mental flexibility, a skill that AI can′t yet provide, although its deep learning ability; since even though, its deep learning depends on the previous data given by the designer, thus, yet limited.[37]

Considering these limitations of the AI, it is still necessary to keep the human supervising together with partial manual actions; instead of full automation which may lead to incorrect results. Thus, in this point of time, the technology still not ready to be independent from the human accompany. Its intelligence is not full yet, its consideration is not promised to be proper, thus it is still not trustworthy. Trusting such an intelligence by integrating it without full awareness in the judicial role implicates the fundamentals of the judicial system and the justice values. It is the human responsibility to consider the very certain developed points and abilities of the AI nowadays, and to integrate it in the right way and use.

Moreover, even though the decision-maker AI has been more accurate than the human judge in certain cases, it is still not enough for the public to make the AI system trustworthy. This is for the public expectations from the AI to be even more accurate. The standards expected from AI in the accuracy terms is significantly higher than the expectations from human judges. This is one more obstacle to tag the AI as trustworthy to integrate the judge′s role. This attitude expresses the insistence of the public on the

[35] Xu, above n 1, p.1033.
[36] Xu, above n 1, p. 1033.
[37] Xu, above n 1, p.1033, 1040.

quality of the results produced by the AI system, while not settling for the efficiency that it can provide.[38]

The judicial role is based on **experience**; the judicial role requires more than the sort of experience that an AI system can provide. It is not the **data** or the information of each case,[39] but the built discretion, the built wisdom, the ability to understand the human actions and interactions, to identify the trustworthiness of people, and to feel the sense of just. All these are still far from the current technological reality. While the judicial AI is based only on data, which can't express the whole nuances of the case, it is rather limited to stiff patterns and clear answers. Such a conduct can't handle the complexity of judicial cases and furthermore, can't guarantee a just result.[40]

One more difficulty facing the worthy integration of AI in the judicial role is the Black Box system. This is an algorithmic system that shows its input and output, while the algorithmic process made to get to the output is totally unknown. Using such a system in the judicial decision-making process has heavy risks on the trustworthiness of the judicial system, for having no reasoning to the decision or the judgement given. In this way, it invades the transparency, and the ability of the human judges and officials to supervise and monitor the correctness of the decisions. Moreover, it may harden or block the parties' ability to appeal on such a judgement, for the lack of reasoning.

Requirements for Trustworthy AI

In the light of the many limitations and obstacles the current judicial AI faces, and in the way to find the right way to use it while using its advantages and avoiding its shortcomings, it is essential to define the requirements from an AI system in the judicial role. When requiring integrating AI in a system with a wide public influence, especially in the judicial system, were morality and justice are the leading values, there are though ethical principles that apply on the human judges, and the more so apply on the AI that inters [fully or partially] the human judge's role, or the one that has any influence on the judicial decisions. The followings are some of these ethical principles presented in the literature about AI ethics:

Respect for Fundamental Rights:[41]

Just and desired decisions are based on the fundamental rights and stem from them. Thus, AI tools which are involved in the judicial decision-making process need to be

[38] Xu, above n 1, p. 1033-1034.
[39] Xu, above n 1, p. 1034.
[40] Xu, above n 1, p. 1034.
[41] Reiling, above n 7, p. 6; Busacca and Monaca, above n 5, p. 162-163.

built and designed based on the fundamental rights, such as privacy, fair trial, the right of access to justice, and equal treatment. For guaranteeing these rights in the AI system, they should be assimilated already in the algorithmic design stage.[42] Which means that this criterion should be taken into consideration in the very first phases of planning and creating the algorithmic basis.

The Commission for the Efficiency of Justice (CEPEJ) adopted in December 2018 the "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment".[43] Which aims to set guidelines and rules to administrate the use of AI in the judicial system. "The CEPEJ's view as set out in the Charter is that the application of AI in the field of justice can contribute to improve the efficiency and quality and must be implemented in a responsible manner which complies with the fundamental rights guaranteed in particular in the European Convention on Human Rights (ECHR) and the Council of Europe Convention on the Protection of Personal Data. For the CEPEJ, it is essential to ensure that AI remains a tool in the service of the general interest and that its use respects individual rights".[44]

Nevertheless, an unsupervised process of training the AI system with random amount and quality of decisions may lead to a "different kind of justice".[45] Thus ensuring results based on the fundamental rights by the law and its right interpretation means selecting precisely the kind of decisions for training the AI system, that reflect the sense of justice that we want to preserve, while minding the possible interpretations that may result by the decision-making process.[46]

Equal Treatment and Non-Discrimination:[47]

This principle is found as one of the most discussed principles in the AI justice debate.[48] Discrimination is one of the well-known human judges' behavioral biases; whether because of their worldviews, backgrounds, beliefs, or agendas, or because of the time of the day, when they've eaten, workload that day, or momentary mood and emotions.[49] Allegedly algorithms should be the solution for this kind of discrimination, since they

[42] Busacca and Monaca, above n 5, p. 162.
[43] European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*, Strasburg, Council of Europe 2019.
[44] CEPEJ, above n 43.
[45] Busacca and Monaca, above n 5, p. 163.
[46] Busacca and Monaca, above n 5, p. 163.
[47] Kaur, Uslu and Durresi, above n 6, p. 107; Reiling, above n 7, p. 6; Busacca and Monaca, above n 5, p. 162-163; Giampiero Lupo, *The Ethics of Artificial Intelligence: An analysis of ethical frameworks disciplining AI in justice and other contexts of application*, Oñati Socio-Legal Series Vol. 12, ISSUE 3, 614, 627 (2022).
[48] Lupo, above n 47, p. 627.
[49] Sourdin, above n 3, p. 1128-1129.

shouldn't act by biased or personal considerations, but by objective ones. However, it is not the case. Algorithmic answer for a question is based on the data given in the system's input or base. The amount and quantity of this data have a direct significant effect on the result in the output; since this data may indicate, for example, that individuals who belong to a certain ethnic group were more likely to be convicted in a certain kind of cases. According to these data, the algorithmic result may wrongly convict an innocent person just for belonging to that ethnic group. Thus, the results of the algorithmic systems are "oriented", not "neutral".

To solve this problem, strict control over the data is required, in a way that guarantees standards such as quality, diversity, and representation.

Data Quality and Security:[50]

As noted above, the quality of data given as input and for training the AI system has a decisive effect on the quality of the decisions made by the algorithm. That since the decisions are based on the input data, and compared to the human judge, the algorithm is limited to that data, and has no other sources which can fix or make accurate the perception the decision given according to. Ensuring qualitative data and sources in the input guarantees qualitative decisions in the output, accordingly; inter alia, unbiased decisions.

Moreover, data security is essential especially in the judicial system since it maintains a highly sensitive data with a real and serious effect on people's lives. Thus "model should protect the data on which it is trained on and the identity of the users using it" to protect their privacy and secure their data.[51]

Data security comes along with a multidisciplinary way programmed models, which integrate experts of technical legal and social sciences. Furthermore, to ensure security, it is essential to conserve the system in a safe and secure environment, to avoid system integrity and unwanted changes in the decision-making process.

Thus, to guarantee data security, we need to deal with three points: to ensure the security of data sources (input sources); to program the AI models according to a multidisciplinary approach; and to conserve the algorithmic systems in safe technological environments.

[50] Reiling, above n 7, p. 6; Busacca and Monaca, above n 5, p. 164; Kaur, Uslu and Durresi, above n 6, p. 108.
[51] Kaur, Uslu and Durresi, above n 6, p. 108.

Transparency and Explainability:[52]

Transparency is one of the significant principles of the reformed law systems. A dedicated study has shown transparency as the most discussed principles in the AI ethics documents, with 83.3% of the investigated and analyzed documents.[53] Decisions and judgements are being established and made accessible to the public. Transparency allows criticism of the judgements and judicial decisions and guarantees fair trial, by requiring reasoning for each decision. The reasoning requirement promotes transparency due to the judgements being exposed to the parties and the public, while giving the parties the option of appealing in justified cases.

Accordingly, explainability enables understanding the outcomes of the algorithmic decision-making process, in a way that make it possible to predict the outcome that will be given based on a certain input. In this way, it is more guaranteed that the decisions given by the AI system are desirable and explainable, in a manner that at least enables the human supervisor to understand the decision's logic.

Though, as mentioned above, in the algorithmic systems, the black box challenges the transparency of the decision-making process, since the only observable parts of the process are the inputs (the data given) and the outputs (the decisions). In this way, the principle of reasoning might be emptied of its content and may lead the judicial system to a world of arbitrary decisions, where there is no access to the logic and manner they were given according to.

Given the above, to protect fundamental rights and ensure just decisions, it is essential to avoid a black box model in the judicial algorithmic system. Moreover, technical transparency required, by providing explanations in a clear language to the computational language, which probably won't mean to the human law expert much.

Furthermore, conserving the reasoning requirement for transparency, including its all-significant consequences, requires presenting the reasoning of each decision given by the AI system. In this way, parties and the public get, like a human decision, a reasoned and understandable decision to each case.

[52] Kaur, Uslu and Durresi, above n 6, p. 108; Busacca and Monaca, above n 5, p. 164; Reiling, above n 7, p. 6.
[53] Lupo, above n 47, p. 627.

AI Under User Control:[54]

This principle keeps the control in the human "user" hands. Which is essentially important to the judicial system, where human judges should ensure justice and right decisions, even in these given by algorithms. That emphasizes the importance of keeping the human in the control position, while delineating the algorithmic role for support and optimize the judicial role rather than replacing the decision maker in a way that limits the human's ability in making these decisions.

One of the factors that directly affect the user control is transparency, mentioned above as a separate principle. Transparency guarantees clarity which make the algorithmic actions understandable, among others, to the user, who is in this case the decision-maker (judge), who should understand the algorithmic decision-making process to approve its justice and rightness.

Such characteristics ensure explainability, that "enable the users of the model to correctly predict the outcomes of given input and the reason that could lead to model failure".[55]

Thus, keeping the human judge (the "user") in the center and in the controller role, can ensure just and desirable decisions, while avoiding situations as the human judge cannot understand the decision given by the AI system, and thus can't supervise on it, fix it or making it more precise or relevant to the certain case. "[T]hat enforces that AI system should always be in control of humans to prevent harm".[56]

Accuracy and Robustness:

These principles ensure that the decisions made by AI are precise, relevant to the certain case and based on the right data. A model is accurate if it can "correctly predict the outcomes by generating less false positives and false negatives".[57] Further, the model should comply with the robustness requirement, in a manner that it should can "perform accurately under uncertain conditions".[58]

Accountability:

Decisions made by AI system may not always be right and precise, either because of external effects or internal unexpected errors. An accountable model should take

[54] Busacca and Monaca, above n 5, p. 165; Reiling, above n 7, p. 6; Kaur, Uslu and Durresi, above n 6, p. 108.
[55] Kaur, Uslu and Durresi, above n 6, p. 108.
[56] Kaur, Uslu and Durresi, above n 6, p. 108.
[57] Kaur, Uslu and Durresi, above n 6, p. 107.
[58] Kaur, Uslu and Durresi, above n 6, p. 107.

responsibility for each decision, while operating an internal control system to justify the decisions to the users.[59]

Furthermore, the judicial system deals with sensitive issues, that have a serious impact on people's lives and on the whole procedure and law values. Judgements, especially on judicial systems with a wide interpretive judges' role, can change the law system's base, such as citizens' rights and validity of laws. Thus, the question of accountability of the AI system, dealing with the responsibilities in cases of failure and wrong outcomes, has a huge importance, and providing a model that has an aspect of accountability and report on errors is essential for a just and desirable system.[60]

Integrity:

This principle captures the algorithmic decision-making process within defined parameters, set by the user according to the desirable characteristics to the outcomes. "These parameters can be operational, ethical or technical and can be different for different applications".[61] Thus, integrity of the AI system should express the fundamental rights, the desirable procedural rules, and the requirements of transparency and accountability. These parameters define the limits which within are placed the desired decisions and outcomes.

Reproducibility:

Reproducibility requires the system to be able to reproduce outcomes based on certain input, in a way that serves for the same input in other cases.[62]

Regulation:[63]

Integrating AI in the judicial system has many unique complications, which need to be regulated clearly. Guidelines and laws are essential for integrating AI in the judicial system in a way that effect the values of justice and human rights positively, avoiding harming these values for trial-and-error manner.

Ways to Correctly Use AI in Court:

As it is currently seen, AI is still limited and not developed enough to comply with the authority and accountability given in the judges' role; mainly given the risks in its use

[59] Kaur, Uslu and Durresi, above n 6, p. 108.
[60] Lupo, above n 47, p. 618.
[61] Kaur, Uslu and Durresi, above n 6, p. 108.
[62] Kaur, Uslu and Durresi, above n 6, p. 108.
[63] Kaur, Uslu and Durresi, above n 6, p. 108.

dealt above, such as biases, unjust decisions, and unsecure data usage. Therefore, using AI in court should be made cautiously and mindfully, with awareness to the complications and unique issues. This should be done while using the AI advantages and abilities in which it surpasses human judges' abilities, in one hand; and in the other hand to be aware to its risks and shortcomings, by supervising its functioning, not using it in functions that it is limited for, and specializing its use for functions that are adapted to its proven advantages.

For all these to be guaranteed, the clear conclusion is that AI shouldn't fully replace the decision-making process of the human judge, but rather assisting him by using its superhuman abilities, for optimizing the decision-making process and making it more efficient, for neutralizing human biases and inaccuracies, and for facilitating the process of extracting the relevant data for each case, such as similar past cases and precedents. Thus, bellow presented the suggested elements to correctly use AI in court:

Partial Automation and Automation in Particular Cases:

In the first chapter above, one of the AI uses dealt is organizing information. This use is one way to express partial automation.

Organizing information attends the pre-court part of the legal process, as automating the extraction of the relevant parts of the information submitted as case documents. This use takes advantage of the ability of AI to recognize "patterns in text documents and files",[64] in a way that make it easier to the human judge to identify the relevant law for a certain case and reduces the possibility that the human judge wrongly bases his decision. For that to be accomplished, "the legal information must first be made machine processable".[65] That means that both the AI system and the judicial system should make adjustments to use AI to organize data effectively. For that to happen, the judicial system should digitize documents and enrich them by adding structure and legal meaning, "with textual readability, document structures, identification codes and metadata all available".[66] These make the judgements and decisions accessible to the world of AI terms and use. In addition, judges and workers of the court should learn the way AI works and understand its logic as knowing the ethics guidelines and principles. At the same time, AI system should be able to explain transparently the way that its outcomes come according to.

Furthermore, the major missions of the judge when dealing with a case are determining in the reliability of the parties, determining the legal questions arise from the case,

[64] Reiling, above n 7, p. 3.
[65] Reiling, above n 7, p. 7.
[66] Reiling, above n 7, p. 8.

determining the facts of the case according to his impression and consideration, and implementing the law on the case's facts. All the other technical missions that assist him to get to a judgement are not complicated for AI abilities. Thus, actions such as organizing data, extracting the relevant data according to given parameters, identifying similarity between the current case and past cases, or suggesting corrections to the decision draft, all these are automatable actions, in a way that significantly promotes efficiency and enables judges to deal with the major legal missions, that AI still not ready to take part in.[67]

Simple recurring cases, that don't require wide consideration or dealing with new questions in new cases, can be assigned to AI system that can deal with them successfully. However, complicated new cases, that deal with new legal questions and can't be determined simply according to previous decisions, AI can't be held accountable to handle these cases and give just and right decisions. Such cases need the human judge sense and experience, especially when the law sources are silent or ambiguous, when a human perspective is needed to fill the legal gap according to the legislator's intention and objectives. Therein "Judicial adjudication is a complex professional activity that requires both specialized legal expertise and the cognitive and emotional abilities of judges. Many key concepts in the judicial application, such as "justice," "reasonable attention" and expression of meaning," are deeply rooted in the rich and colorful life of humans".[68]

Therefore, partial automation significantly increases the efficiency of using AI system in court, and significantly decreases risks of incorrect uses of it for its limitations. For answering the question of which part of the decision-making process should be automated, the decision-making process may be divided into 4 stages: (1) information acquisition; (2) information analysis; (3) decision selection; (4) decision implementation.[69] An important point to mention, is that people do care about which part of the decision-making process is automated. Most people believe that the results of automating the information acquisition process are fairer than the results of automating other stages. Moreover, legal experts believe that automating the decision implementation stage is less fair than individuals who don't have a legal expertise.[70] Accordingly, people believe that "low level of automation would ensure the fairest outcomes in judicial decision-making".[71]

---

[67] Xu, above n 1, p. 1039.
[68] Xu, above n 1, p. 1037.
[69] Algorithms in the court, p. 9.
[70] Algorithms in the court, p. 9-10.
[71] Algorithms in the court, p. 14.

Suggestions to Deal with the Gap Between Limited AI and Just AI:

Human Control:

Human control is still needed and essential in all phases. That guarantees supervising on the outcomes and the process of AI and allows fixing incorrect results. In addition, it ensures the independence of the human judge and avoids over-reliance on the algorithmic system. Furthermore, the judicial system won′t and shouldn′t give up the human judge′s presence in the system and the decision-making role, here is ″[d]rawing experience and wisdom from social life and fully considering individual cases in the social life background are the basic conditions for human judges to gain the public′s trust and approval″, thus, human accompaniment is essential ″to confirm and maintain the ethical order and good feelings that the society hopes for, to fully care for the dignity and value of humans″.[72]

Human control takes place additionally by guaranteeing the human judge independency and avoiding his over-reliance on the AI assist while decision-making. For that to be ensured, the responsibility for each decision should mainly remain on the human judge, rather than sharing it with the AI system. In this way, judges should supervise each stage and use the AI outcomes cautiously, while considering their relevance and accuracy.[73]

Tools such as ″Rechtwijzer″ that presented above should be used wider and should be adopted in more countries. It suggests smart resolutions to the specific kind of disputes (separation or divorce). This kind of disputes is mostly simple hence dealing with it is automatable. The program directs the parties to relevant information, links, websites, and tools. The way the program works keeps the process in a human control in tow main aspects. the first one is that the results given by the program are suggestions, without any element of obligation; still considering that these results are outcomes for information entered by the parties themselves. In this way, the process remains in the couples′ control. The second aspect is the default of the program when parties don′t get to resolution. It links them to human professionals who have the skills and the expertise required to get them to resolution.

Data Quality:

One of the most important elements to deal with the gap between the limitations of AI and the desired outcomes is to ensure qualitative data. ″It has long been known that bad data, such as legally incorrect decisions, reduce the quality of the AI results″.[74]

---

[72] Xu, above n 1, p. 1041.
[73] Xu, above n 1, p. 1040.
[74] Reiling, above n 7, p. 7.

An AI system should be trained by representative, diverse, neutral, right, and relevant data. In any other case, the algorithm would conclude that wrong and undesired factors, should be considered to get to the "right" decision. Such a case has already occurred, in the COMPASS example given above.

Decisions of the judicial system, especially, should meet high standards of neutrality and fair trial, avoiding ulterior motives and irrelevant considerations. Furthermore, one of the significant reasons for integrating machines in the decision-making process is to neutralize the human judge's possible biases. For that reason, qualitative data is essential, for eliminating effects of bad data on the justice and neutral requirements from a judicial decision.

Collaborations Between Legal and Technological Experts, and the Requirement for Human Judges to Understand Technology and AI Systems Function:

Translating legal language into code, supervising the AI system function, and understanding the decision-making process that the AI system made to get to the given decision. All these require deep understanding and expertise in both fields of law and technology. Thus, in all stages of integrating and using AI in court, both kinds of experts should be involved.

Integrating AI in the judicial system involves the need for human judges' proficiency in the technological and algorithmic field. The new legal era requires and expects the judge to add technological knowledge to the profession's required skills. Understanding both fields of law and algorithmic technology ensures the control of the human judge on the algorithmic process, in a way that enables him to "participate in the making and revision of artificial intelligence legal system, supervise the fairness of algorithms, timely discover the problems of artificial intelligence algorithms, and effectively avoid the technical risks of algorithm black box, algorithm hegemony and algorithm discrimination".[75]

AI Judge V. AI Assistant for Judge:

AI is getting smarter by the time. It is purchasing new skills and abilities, getting more accurate and resembles itself more to a human judge. However, besides its very developed skills and progress, it is still limited to its mechanical essence characteristics. These limitations risk the judicial system and its highly desired values, such as justice, ensuring fundamental rights, and ensuring compassionate, conscience and ethical

[75] Xu, above n 1, p. 1038.

decisions-maker, who understands humanity in all its characteristics, to be able to judge human beings.

Although AI systems can be decision-makers, but as mentioned above, the cost for the risks in using them instead of human judges is significantly higher than the utility in such a use. Thanks to their developed skills in organizing data, their ability to store a huge amount of data, and their ability to make right decisions in simple cases, they can be used in certain ways that utilize their benefits and avoids their shortcomings. That means that AI shouldn′t replace the human judge but support his decision-making process.

AI system can even suggest a judgement for its decision-making ability, that should be reviewed by the human judge who can use it, fix it, or approve it. In this way, the machine would save time and efforts for the human judge, who would deal with a draft rather than creating the judgement in all its phases. Moreover, it saves him the process of searching for the relevant data, laws, precedents, and sources. Such a use remains the control on the decision-making process on the human judge′s hands, while the AI system wouldn′t generate a judgement by itself, without human supervision.

The conclusion that AI should be used to support human judge rather than replacing him promotes the ″transhumanism″,[76] which means that technology has a significant role in enhancing the human beings′ intellect, abilities, and functions. Using AI in such a way keeps the human judge for tasks that requires his own attribution as a human being. In this way, the judge′s role becomes more refined. Moreover, in a systemic aspect, this would be a better utilization of resources, both the technological and the human ones.

Conclusion:

While the options are an AI system replacing a judge or an AI system supporting the human judge, according to the found above, supporting the decision-making process made by the human judge is the chosen option; ″the goal of the development of AI systems should be to complement current human work, allowing for greater efficiencies, rather than total replacement of humans″.[77] AI system doesn′t have many essential skills for the law field, especially for the judge′s role, that the human judge has. However, human judges lack supportive skills that AI systems have, which can be used for facilitating the decision-making process, and making it more efficient, faster, and more neutral.

In the light of the written above, an AI judge is clearly not the optimal preferred option. Human control is essential, especially in legal proceedings, where data may be highly

[76] Sourdin, above n 3, p. 1131.
[77] Sourdin, above n 3, p. 1130.

sensitive. Furthermore, human leading to these proceedings is essential too, for having the human abilities to feel people, understand human logic and behavior, have the mental flexibility and creativity in resolving cases, evaluate evidence and sayings, and have the full ability to explain their conclusions and decisions. Judges are required to be conscientious and compassionate; they need to understand emotions and motives. These have a direct impact on justice, which is an inherent liability on the judge's role.

Furthermore, judges are not merely a machine outcoming abstract decisions. They have an essential public and civil role, such as an educational role of dictating desirable social behaviors. Moreover, they have a significant role in applying the law in a way that keeps them an interpretation space, which needs the human brain capacities, and still not provided by the AI services.

Indeed, simple cases, such as couples' separation or uncomplicated disputes, can be dealt with an advisory system based on AI models, such as the "Rechtwijzer" program used in the Netherlands. Such a use advises the parties resolutions that may solve their dispute outside the court, or without the need to a human accompaniment. At the same time, such a program refers the parties to human professionals if it couldn't get to solve the dispute. All the same, the parties have the right to choose in any given time, to leave the process and go to court or to any other platform.

Integrating AI system in the judicial decision-making process is inevitable, yet desired. Correct use of it would positively affect the stability and the neutrality of the judicial decisions, the efficiency and the time needed to complete handling cases, judges' workload, and the costs of legal proceedings. For that to happen, the main accountability rests with everyone involved in the craft, both the technological experts and the legal ones. They should be aware to the risks in each use, and to specify the functions they intend to involve the AI activity into the trustworthy algorithmic capabilities.

References

A. D. (Dory) Reiling, *Courts and Artificial Intelligence*, Vol. 11, Issue 2 1 (2020).

Angela Busacca and Melechiorre Monaca, *Using AI for Justice: Principles and Criteria of the "European Ethical Charter on the Use of AI in Judicial Systems"*, in: Marino, D., Monaca, M. (eds) Artificial Intelligence and Economics: The Key to the Future, Lecture Notes in Networks and Systems, Vol. 523, 157 (2022).

Civil Resolution Tribunal Act [SBC 2012] CHAPTER 25.

Davinder Kaur, Suleyman Uslu and Arjan Durresi, *Requirements for Trustworthy Artificial Intelligence – A Review*, in: L. Barolli, K. F. Li, T. Enokido & M. Takizawa (Eds), Advances in Networked-Based Information Systems, 105 (2021).

European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*, Strasburg, Council of Europe 2019.

Giampiero Lupo, *The Ethics of Artificial Intelligence: An analysis of ethical frameworks disciplining AI in justice and other contexts of application*, Oñati Socio-Legal Series, Vol. 12, ISSUE 3, 614 (2022).

Tania Sourdin, *Judge v. Robot? Artificial Intelligence and Judicial Decision-Making*, in: University of New South Wales Law Journal, Vol. 41, Issue 4, 1114 (2018).

Zichun Xu, *Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities*, in: Applied Artificial Intelligence, Vol. 36(1), 2013652 1025 (2022).

https://www.rossintelligence.com/what-is-ai.

https://civilresolutionbc.ca/.