

Artificial intelligence in the Criminal Justice System

The Role of decision-makers and how big data tools support them

Federico Boggia

02/06/2022

Abstract. *This essay analyzes the major social and ethical issues of big data and how to legitimize the use of big data algorithms and AIs in the criminal justice system. Focusing on the procedural justice aspect of these tools and analyzing the different regulations currently in force, this essay investigates whether such algorithms can be delegated and replace judges in their role as decision-makers. Here it is argued that decision makers cannot be replaced and that in these fields some technologies should only be used to support decisions.*

Keywords: big data ethics, AI bias, procedural justice

Introduction

Many innovative technologies have changed the procedures and methods of operation of different systems in our society. Nowadays Big data-based algorithms are widely used in different fields, and very often the analysis operations on these data are carried out by AI systems such as neural networks.

One of the much-discussed fields of application of these technologies is the criminal justice system. In recent years, especially in the United States, several news headlines were talking about “Algorithms that Predict Future Criminals”.¹ They got a lot of public attention since the usage of risk assessment software to predict the likelihood of recidivism of individuals was new to the public. These systems are becoming more common, and they are already used to provide guidance to judges at sentencing.² Like many new features, these also leads to different benefits, but also various problems: the most discussed are due to data biases, which can harm the neutrality of judgment in a trial. For instance, in *Loomis v. Wisconsin* (2016) in the USA, the Supreme Court of Wisconsin considered the legality of using the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) risk assessment software in criminal sentencing. In the same year ProPublica published a report³ in which the results of the algorithm used by the COMPAS system were analyzed and the algorithm biases that could lead to a consolidation of discrimination phenomena were highlighted.

¹ Chiel, Ethan. *Secret Algorithms That Predict Future Criminals Get a Thumbs Up from Wisconsin Supreme Court*, FUSION (July 27, 2016).

² Slobogin, Christopher. *Risk Assessment*, in *The Oxford Handbook of Sentencing and Corrections* 196, 203-05. Joan Petersilia & Kevin R. Reitz eds. (2012).

³ Angwin, Julia; Larson, Jeff. *Machine Bias*, ProPublica. (2016).

Another issue concerns the role that these technologies can play in the criminal justice system. Will they only remain as tools to support decision-making, or will they replace the decision makers?

In this essay, through the analysis of current legislation, and using a typical point of view of procedural justice, four key factors, necessary to ensure that a given procedure is perceived as fair by people, are analyzed. The conclusion deriving from this analysis leads to consider it difficult to use certain technologies to replace decision makers.

Changing the decision maker in the criminal justice system

Throughout the history of mankind, various criminal justice systems have always relied on people to play the role of decision-makers. Judges, kings and queens, legislators, druids, priests, bureaucrats: they all cover roles that are legitimized by the peoples. Although these figures have always changed over time and space, the role of the decision-maker has always been played by individuals.

The trust placed in decision makers by people guarantees the perceived fairness of the processes.

Not only that, but some research also reveals that public trust in the courts is an essential component for maintaining good governance.⁴ Furthermore, people expect judges to behave fairly and impartially and in addition to that, they expect a certain predictability in the decisions made by them which are legitimized by a stable system of laws.⁵

We could consider perceived trust as a combination of court user's perception of predictability, fairness, trustworthiness, and unbiasedness of a judicial decision-maker. The more people trust criminal justice systems, the more they obey the law. Many studies show how people obey the law mainly because they believe the system it is legitimate; not because they fear sanctions if they transgress.⁶

Hence, if someone wanted to delegate the role of decision-maker to an AI, they should first consider the need to make everything as transparent and understandable as possible in order to increase people's trust in that tool.

There are systematic differences in individuals' perceptions of algorithmic versus human decisions.⁷ There are several positives and negatives to delegating AI, algorithms could be perceived as more consistent and objective than humans. However, even if these tools can provide forecasts which are fairer, more efficient, and more accurate than clinical judgments

⁴ Jackson J et al. *Developing European indicators of trust in justice*. Eur J Criminol 8(4) (2011).

⁵ Resnik J, *The democracy in courts: Jeremy Bentham, 'Publicity', and the privatization of process in the twenty-first century*. No Found 10 (2013).

⁶ Tom R. Tyler, *Why People Obey the Law* (1990).

⁷ Helberger N et al. *Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making*. Comput Law Secur Rev 39 (2020).

made by human beings, people believe that algorithms do not consider the uniqueness of cases of individuals and above all they feel they cannot take an active part, not being able to have an interpersonal relationship with the algorithms.

Big data algorithms and AIs are already getting used from law firms to read documents, prepare case files, and predict the win rate of court cases.⁸ With the increasing use of these technologies, even in the courts and law firms, it is quite predictable how these algorithms will find more and more space in criminal justice systems.

There are various specific risks of adopting AI systems to replace judges in their role as decision-makers: for example, the current statistical model of prediction used by many algorithms based on machine learning takes into account parameters such as accuracy, precision, recall. These parameters are needed in order to keep track of the false positives generated by the predictive model. The problem with these systems is the domain of applications of such predictive models: in the criminal justice system even just a false positive can correspond to an unjustly condemned person.

Furthermore, predictive models generated by AIs are the result of a dataset-based training process. These datasets are usually made through data-mining operations. This data can incorporate different types of bias which can lead to the consolidation of certain prejudices or discrimination. Two main problems therefore emerge: the first is that the data and the algorithms that use them are human artifacts, the second is that, due to the very nature of these algorithms based on static models, the results could lead to different paths compared to the initial objectives.

Big data ethics and social issues

Considering the data, it is necessary to analyze the collection processes that lead to the creation of databases and to the concept of big data.

Big data are characterized by the increased volume, velocity and variety of data being produced. Such characteristics bring major concerns raised by community stakeholders that fear that large-scale analysis could “*warp or misinform narratives, particularly those relating to marginalized populations*”.⁹

For many fields of study, the collection of big data is made easy by the immeasurable amount of information available on the Internet and on social networks. Unfortunately, the mere collection of data excludes possible starting biases from the analysis. For example, several studies show that there are different socio-demographic and economic factors that lead individuals to use one site or a social network rather than another. The time spent and how to use these platforms also depends on social factors.¹⁰ While in other fields these biases could

⁸ Faggella D. *AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications*. Emerj (2020).

⁹ Popham, J., Lavoie, J., & Coomber, N. *Constructing a public narrative of regulations for big data and analytics: Results from a community-driven discussion*. Social Science Computer Review. (2018).

¹⁰ Hargittai, E. *Potential biases in big data: Omitted voices on social media*. Social Science Computer Review. (2018)

lead to simple noise in the generation of predictive models, in the criminal justice system, data analysis and tuning processes are necessary in order not to run into predictions and interventions based on out-of-context interpretations that can lead to the punishment of innocent people.

Moreover, in criminal justice, data is not something exact: the same crimes are man-made normative phenomena that may change according to place and time. The big data used by predictive models usually come from larger groups and tend to ignore minorities and special cases and thus be constituted by the most widespread and regulated types of crimes.

The predictions of these models will consequently be the more accurate the closer the reality of the observed facts is to the reality perceived by the AI in the training phase. The problem in providing big data to AI in this context is the difficulty of obtaining predictive models capable of balancing competing values.

In modern democracies the codes of criminal procedure are the result of years of legal doctrines in constitutional and criminal law which lead to the perception of the different legal systems more or less equitable. Understanding how to balance decisions is what makes the perception of one legal system more legitimate than another. If the data underlying the training of AI to use in criminal justice systems are biased, it is clear that such systems can never be perceived as legitimate by the population.

European regulations and discrimination in data-supported decision making

Even if the use of algorithms based on big data and AI systems can lead to the reinforcement of certain biases and possible discrimination, it is undeniable that the use of these tools as a support to the decision-making process can help to make decisions that are as objective and informed as possible. A data-supported decision could also help limit existing discriminatory treatments in human prejudices.

It is also clear that the main difficulties in understanding how to best implement these tools are due to their recent introduction. An important role should be played by institutions which should consider every possible damage coming from a poor regulation.

At EU level, an important step forward at the regulatory level was taken with the introduction in 2016 of the General Data Protection Regulation (GDPR).¹¹

The GDPR analyzes these technological innovations with particular attention to fundamental rights that could be infringed by technologies that use people's data.

The GDPR also explicitly introduce a right to explanation. It is expressly mentioned that data subjects should be provided with “meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject”.¹²

¹¹ General Data Protection Regulation (GDPR) Compliance Guidelines: <https://gdpr.eu/> (last visited 04/06/2022)

¹² General Data Protection Regulation, Recital 71, Art. 13 (2) (f) and Art. 15 (1) (h).

Article 21 of the EU Charter of Fundamental Rights prohibits discrimination based on several grounds. All data relating to characteristics that can lead to discrimination (sex, ethnicity, age, etc.) are intrinsically linked to the individual who possesses them and are therefore by definition personal information. As personal information they are protected under the data protection legal framework.

More specifically when it comes to the use of big data for law enforcement purposes, the resolution of the European Parliament of 16 February 2017 “warns that [...] maximum caution is required in order to prevent unlawful discrimination and the targeting of certain individuals or groups of people defined by reference to [...] as well as individuals who happen to be defined by particular characteristics”.

But the European Union has already recognized the importance and benefits of using AI in the justice sector¹³: “the European approach to artificial intelligence (AI) will help build a resilient Europe for the Digital Decade where people and businesses can enjoy the benefits of AI”.¹⁴ On the other hand EU also points out the importance of trust and transparency given by algorithmic systems to minimize risks and maximize benefits.

On 8 April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence followed by the Assessment List for Trustworthy Artificial Intelligence (ALTAI) (17 of July 2020). ALTAI is a practical tool that help developers to develop a Trustworthy AI based on several requirements.¹⁵ Now EU is also working on a Regulatory framework proposal on AI¹⁶ which consists in a risk-based approach that defines 4 levels of AI risk: unacceptable, high, limited, minimal or no risk. For example, the use of AI classified as unacceptable would be prohibited while those classified as high risk will have special requirements for data governance, document and record keeping, transparency requirements, and human oversight. AI used in court is classified as high risk.

Consequently, the European Union does not ban the use of AI in the criminal justice system and is also trying to regulate its implementation. The adoption of AI will affect the role of judges and citizens' trust in the justice system. As sooner or later we will see more and more algorithms involved in the current decision-making process in the courts, therefore it is necessary to understand what factors are necessary for the legitimacy of these tools.

¹³ European Union, *Artificial Intelligence for Europe*. European Union (2018).

¹⁴ A European approach to artificial intelligence | Shaping Europe's digital future (europa.eu): <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> (last visited 04/06/2022).

¹⁵ Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment: <https://digital-strategy.ec.europa.eu/it/node/806> (last visited 04/06/2022).

¹⁶ Regulatory framework proposal on artificial intelligence | Shaping Europe's digital future (europa.eu): <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (last visited 04/06/2022).

Algorithms, data, and procedural justice

Several studies show how the attitude of individuals towards algorithmic decision-making depends on the type of task assigned to the algorithm. In general, algorithms are perceived as more reliable when dealing with facts concerning reason and not emotion.¹⁷

The main idea that pushes AI proponents to replace judges in the role of decision-makers is the possibility of eliminating any bias and any human heuristics from the decision-making process. This would serve to limit the criminal justice system in a direction of pure scientific method: this idea was already present in Cesare Beccaria's pamphlet *On Crimes and Punishments* (1764) which concluded that criminal justice should be confined to reason. In addition to having paved the way for the removal of the death penalty, Beccaria's work made it possible to humanize criminal trials and to give a role to the emerging social class of those times and to contribute to the birth of modern democracies. Similarly, today, the transition to new types and processes of decision-making serves to legitimize the new digital social class that pushes the adoption of these tools to have fairer and impartial processes.

While there are huge differences between humans and algorithms, it is necessary to understand how one can legitimize the delegation of an AI as a decision maker instead of a judge. The perception of fairness by individuals of the criminal justice system is different from real fairness. Many psychological studies show that individuals tend to value interaction with authority as more important than possible outcomes (Tyler 1990).

Using principles of procedural justice, it is possible to identify some key factors necessary for the legitimation process on the part of individuals. As the name implies, procedural justice focuses more on processes and thus on interactions rather than on process outcomes. Hence, if the procedures are considered legitimate and the criminal justice system is perceived as fair, then unfavorable outcomes for individuals are not a threat to the legitimacy of legal authorities. If we consider two alternative procedures that serve a particular criminal justice system task, it is necessary to monitor the procedural justice aspect to understand which of the two to rely on. Therefore, if someone wants to delegate an AI, they should focus on increasing the procedural justice of that tool.

Obviously, there is always the danger that authorities might be able to produce a procedure that appears fair but in fact is not, thus creating a "false consciousness (Tyler 1990). The focus, however, is on the possible legitimization by individuals: it has been shown that there are four key factors that contribute to the perceived fairness of the procedures:

1. Interpersonal aspect: how the individual is treated during the interactions with authorities.
2. Neutrality: whether the decision-maker is perceived as neutral.
3. Trustworthiness: whether the decision-maker is perceived as trustworthy.
4. Participation in the process: how and if the individual was able to actively participate.

¹⁷ Rai TS, Diermeier D. *Corporations are Cyborgs: Organizations elicit anger but not sympathy when they can think but cannot feel*. *Organ Behav Hum Dec* 126:18–26 (2015).

Interpersonal aspect

The first factor must deal with how the individual is treated during the interactions with authorities. More specifically if an individual is treated with respect and dignity by the decision maker.

Obviously if an AI were to fill the role of decision-maker, the individual would continue to interact with people of the criminal justice systems such as police officers. Surely this factor could be the least relevant as the interpersonal aspect would not be outdone but knowing that the decision was made in a predetermined way by an algorithm could make the individual feel harmed in his dignity.

Neutrality

One of the most crucial factors is certainly neutrality. On this factor, supporters of big data algorithms and AI are trying to ensure that these tools are increasingly adopted in courts. But can an AI decide in a truly neutral way? In this context, neutrality refers to whether the decision maker follows rules impartially and is making objective, unbiased decisions based on facts. The software running the algorithm follows the same rules in the same way for every decision, this would seem to indicate neutrality.

However, one of the main criticisms concerns the aforementioned biases. If the big data used for AI training were biased, then many or all decisions could be influenced by these biases. Nevertheless, these could be easily removed from a computer algorithm than from a human. Not only that, if someone were to become aware of specific cases of discrimination, a decision-making algorithm could be implemented that aims precisely at reducing such existing discrimination, making these algorithms the perfect tool to bring procedural justice to neutrality. This factor depends a lot on the AI implementation rules used.

Trustworthiness

In the field of procedural justice there are two characteristics that contribute to trustworthiness: the first is whether the decision maker has a link with the individual (same social group, city of origin, age, etc.) and secondly whether the decision maker explains and argues his decision.¹⁸ Certainly, the AI will never achieve reliability through some kind of personal connection. Therefore, an algorithm will only be deemed trustworthy if its calculations are transparent and understandable to the affected person.

The main problem here is that AI's that involve a large amount of data are often Black box AI. These types of systems can be viewed in terms of its inputs and outputs without any knowledge of its internal workings. Opacity of decisions is not an exclusive problem of AIs. As of now, the courts have generally refused to impose a duty on trial court judges to explain sentencing decisions. In this factor, AI could improve by adding transparency to processes and analyzing the correlations between inputs and outputs to better understand the internal processes of the AI itself.

¹⁸ E. Joh, *The New Surveillance Discretion: Automated Suspicion*, Big Data, and Policing, 10 HARV. L. & POL'Y REV. 15, 40 (2016).

Another important feature would be the ability of AI to explain decision-making processes in natural and easily understandable language.

“Explainable AI should be able to communicate the result naturally to humans, but also the reasoning process that justifies the result.”¹⁹

Participation in the process

Individuals feel legitimate a process the more they can express their point of view to the decision maker.

The "process" consists also of the individual telling their story to the decision -maker.

Although this part may be irrelevant to the outcome of the process, active participation makes the perception of the justice system by individuals more legitimate. Furthermore, in almost all modern criminal justice systems, individuals are represented by lawyers who take their place. Hence rarely an individual has a real chance to speak to the decision-maker. It is the mere fact of being able to speak that makes the individual feel an active part of the process. Clearly with an AI this factor would completely fail.

With an AI the defendants will have no real chance of participation in the trial and no opportunity to present their point of view to the decision-maker.

¹⁹ Opening the ‘black box’ of artificial intelligence | Research and Innovation (europa.eu): <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/opening-black-box-artificial-intelligence> (last visited 05/06/2022)

Conclusion

Als and big data algorithms in general do not appear to achieve an exceedingly high score on the procedural justice scale. Although these tools could potentially increase the perceived neutrality of the criminal justice system, they fare poorly in the areas of participation and trustworthiness.

Certainly, there are many differences between human beings and algorithms, the focus held in this essay was in general the perception by individuals of these differences. In particular, the possibility of introducing AI into the criminal justice system up to the possibility of delegating them in the role of decision-maker. Furthermore, everything was analyzed without considering hybrid situations, but taking into consideration only two alternatives: the current model and a model with an AI as decision-maker. The level of integration of AI and its relationship to human judges can take many forms and can affect people differently.

Surely from this analysis it emerges how the de-subjection operated by the digitalization of our society can lead to undesirable consequences. On the other hand, process automation could increase our knowledge of the criminal justice system itself.

For instance, if in some justice systems the existing structure leads to arbitrary and discriminatory results, AI tools could improve the reliability and neutrality of the system. In conclusion, the introduction of these tools in the courts must take into account many factors, but most likely the decision makers will never have to be replaced by AIs. Criminal procedure has several conflicting goals and many times the outcome requires a mediation that is the result of years of jurisprudential doctrine. This makes it difficult to completely remove the human component from the decision-making part. The impact on society and on the human rights of individuals would be too great to allow such a radical change: as humanity we would lose the power to decide on ourselves.

References

- Angwin, Julia; Larson, Jeff. *Machine Bias*, ProPublica. (2016).
- Aleš Završnik, *Algorithmic justice: Algorithms and big data in criminal justice settings*, *European Journal of Criminology*, Vol. 18(5) 623–642 (2021).
- Chiel, Ethan. *Secret Algorithms That Predict Future Criminals Get a Thumbs Up from Wisconsin Supreme Court*, FUSION (July 27, 2016).
- Christopher Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs*, *NIJ Journal* 280, (2019).
- Faggella D. *AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications*. Emerj (2020).
- FRA – European Union Agency for Fundamental Rights, *#BigData: Discrimination in data-supported decision making* (2018).
- E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 HARV. L. & POL'Y REV. 15, 40 (2016).
- Gizem Yalcin Erilis, Themeli et al., *Perceptions of Justice by Algorithms* (2022).
- Hargittai, E. *Potential biases in big data: Omitted voices on social media*. *Social Science Computer Review*. (2018)
- Helberger N et al. *Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making*. *Comput Law Secur Rev* 39 (2020).
- Jackson J et al. *Developing European indicators of trust in justice*. *Eur J Criminol* 8(4) (2011).
- Pakes, F. *Comparative Criminal Justice: Fourth Edition* (4th ed.). Routledge (2019).
- Popham, J., Lavoie, J., & Coomber, N. *Constructing a public narrative of regulations for big data and analytics: Results from a community-driven discussion*. *Social Science Computer Review*. (2018).
- Rai TS, Diermeier D. *Corporations are Cyborgs: Organizations elicit anger but not sympathy when they can think but cannot feel*. *Organ Behav Hum Dec* 126:18–26 (2015).
- Resnik J, *The democracy in courts: Jeremy Bentham, 'Publicity', and the privatization of process in the twenty-first century*. *No Found* 10 (2013).
- Ric Simmons, *Big Data and Procedural Justice: Legitimizing Algorithms in the Criminal Justice System*, 15 OHIO St. J. CRIM. L. 573 (2018).
- Slobogin, Christopher. *Risk Assessment*, in *The Oxford Handbook of Sentencing and Corrections* 196, 203-05. Joan Petersilia & Kevin R. Reitz eds. (2012).
- Tom R. Tyler, *Why People Obey the Law* (1990).
- Tom R. Tyler, *Social Justice: Outcome and Procedure*, 35 INT'L J. PSYCHOL. 117, 121 (2000).
- Wenhong Chen, Anabel Quan-Haase, *Big Data Ethics and Politics: Toward New Understandings*, *Social Science Computer Revie*, Vol. 38(1) 3-9 (2020).