Fundamental rights and the use of artificial intelligence in Court

Van Gyseghem, Jean-Marc

Published in:
The Cambridge handbook of lawyering in the digital age

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

14

Fundamental Rights and the Use of Artificial Intelligence in Court*

Jean-Marc van Gyseghem

14.1 INTRODUCTION

The modern world increasingly integrates information and communication technologies into more and more digital services. This involves the use of artificial intelligence (AI) and its algorithms with, necessarily, a transfer of data between various stakeholders, whether through networks or devices.

When introducing its report on algorithms and human rights, the Committee of Experts of the Council of Europe explained that:

> Automated data processing techniques, such as algorithms, do not only enable internet users to seek and access information, they are also increasingly used in decision-making processes, that were previously entirely in the remit of human beings. Algorithms may be used to prepare human decisions or to take them immediately through automated means. In fact, boundaries between human and automated decision-making are often blurred, resulting in the notion of "quasi- or semi-automated decision-making."[1]

The move from human to algorithmic justice implies multidisciplinary interactions between multiple actors processing data. It also involves new actors, such as the developers of software and algorithms. This multitude of stakeholders can make it difficult for citizens to have a real understanding of the algorithm or system beneath the AI. But what do we mean by AI?

The Council of Europe considers that "in the broadest sense, the term refers indistinctly to systems that are pure science fiction (so-called 'strong' AIs with a form of self-awareness) and systems that are already operational and capable of performing very complex tasks (face or voice recognition, vehicle driving – these systems are described as 'weak' or 'moderate')."[2] In other words, "algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations. The procedures

* This work has been done with the financial support from the European Union's Horizon 2020 general MGA program under Grant Agreements no. 830892 (SPARTA) and FEDER dans le cadre du portefeuille de projets WAL-E-CITIES (2017–2020) pour la Région Wallonne This publication reflects the views only of the authors and the European Commission cannot be held responsible for any use which may be made of the information contained therein.

[1] Committee of Experts on Internet Intermediaries (MSI-NET – CoE), "Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications" at 3, https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html (accessed July 10, 2020).

[2] Council of Europe, "What's AI," www.coe.int/en/web/artificial-intelligence/what-is-ai (accessed July 10, 2020).

name both a problem and the steps by which it should be solved." Algorithms are thus perceived as "a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome."[3]

Reading these two statements, the question arises whether the definition of "intelligence" as "the ability to learn, understand, and make judgments or have opinions that are *based on reason*" is still valid.[4] Obviously, AI needs humans to exist. Even if AI could develop itself autonomously by getting information and data, human intelligence is still needed to make it work. AI will either be an expert-level system receiving data and rules/models to deliver a response, or it will be a machine learning system receiving results and data and delivering rules/models, or both. But in each case, the human is at the base of the way in which AI operates. AI does what humans tell it to do, or use the knowledge provided at its creation, with all the potential biases that will be discussed in this chapter.

The use of algorithms in justice raises many questions, such as the ones about the transparency of data processing and decisions but also about further processing, impartiality/presumption of innocence, and equal access to justice. All these issues have a significant impact on fundamental rights, which states cannot divest themselves of. Citizens are entitled to dignity and respect, and the use of AI in court will necessarily have to take these rights into account. These questions will be addressed in this chapter.

The OECD stated in May 2019 that "AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society."[5] This summarizes the issues raised by the use of AI in any processing and, even more, when using it in the course of justice.

First, it is necessary to highlight the fact that technologies are more and more integrated in the professional world, in decision-making, and in the legal environment. AI does not change this paradigm, but triggers multidisciplinary interactions between computer scientists, lawyers, policy makers, sociologists, etc. In other words, AI is not only a technical tool, but it also involves legal and social issues.

With AI, decision-making is transformed from non-autonomous systems, characterized by full human control, to autonomous systems with no or very limited human control. The rise of algorithmic governance entails a lessening of human control and self-empowerment over matters that involve decision-making. Needless to say, this less human approach entails numerous legal and ethical dilemmas.

About the use of AI in a judicial context, the Committee of Experts on Internet intermediaries (MSI-NET) of the Council of Europe[6] highlights the fact that the use of AI in crime prevention and criminal justice might generate some benefits, such as facilitating the processing of large amounts of data faster. In the past, terrorist attacks have put the use of AI under the spotlight, as states asked social networks to use algorithms to track potential terrorists. However, AI is not harmless in terms of "freedom of expression, it also raises concerns for fair trial standards contained in Article 6 of the European Charter of Human Rights (ECHR), notably the

[3] T. Gillespie, "The Relevance of Algorithms" in T. Gillespie, P. Boczkowski, and K. Foot (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (Cambridge, MA: MIT Press, 2014) 167.

[4] Cambridge Dictionary, emphasis added, https://dictionary.cambridge.org/dictionary/english/intelligence (accessed July 10, 2020).

[5] OECD, *Shaping the Digital Transformation in Latin America* (Paris: OECD 2019) 92, with reference to the OECD Principles on Artificial Intelligence.

[6] MSI-NET – CoE, n. 1.

presumption of innocence, the right to be informed promptly of the cause and nature of an accusation, the right to a fair hearing and the right to defend oneself in person."[7]

This chapter will explore the use of AI as an actor/instrument of justice with respect to various fundamental rights and guarantees, such as the right to respect for private life and a fair trial, which might be significantly impacted by AI.

The division of this chapter in sections and subsections is obviously arbitrary and not an easy task. Indeed, the concepts analyzed are common to various fundamental rights. However, the chapter attempts to structure the analysis in a linear way.

14.2 TRANSPARENCY

14.2.1 *Principles*

We will analyze the concept of transparency from two perspectives, namely, the right to a fair trial and the right to data protection.[8] Transparency is a cornerstone of these two fundamental rights: is found it in the data protection legislation as well in Article 6 of the ECHR (in the case of the right to a fair trial). With reference to justice, it is an element of democracy; indeed, it makes it possible to differentiate between a democratic regime and a dictatorship.

The European Court of Human Rights (ECtHR) considers that the requirement of transparency and the right to information deriving from it are fundamental.[9] The lack of transparency may give rise to a violation of Article 8 of the ECHR, as the ECtHR stated in a judgment of January 17, 2019.[10] The case was about administrative proceedings in which a transgender Macedonian national, registered as female, had introduced a request of modification of the sex/gender marker on his birth certificate. After a diagnosis of transsexuality and an adequate hormonal treatment, he succeeded in modifying the first name to a clearly male one. However, the sex/gender marker and numerical personal code remained the same (female). The reason for this was that no official document showing the change of gender was produced. The applicant complained, with no success, of the absence of a regulatory framework for legal gender recognition and the arbitrary imposition of a requirement for genital surgery. The ECtHR considered that the lack of any regulatory framework led to a lack of transparency ensuring the right to respect for the applicant's private life. The ECtHR concluded that the "legal framework in [the former Yugoslav Republic of Macedonia] does not provide 'quick, transparent and accessible procedures' for changing gender on birth certificates for transgender people."[11]

The ECtHR's decision can be easily transposed to the framework of justice. Transparency is a fundamental right of every individual, which must be adequately protected even in the context of the administration of justice.

Brought to the field of justice, the right to transparency extends, among other things, to the reasoning of judicial decisions. In a judgment of November 16, 2010, the ECtHR had the opportunity to reiterate this principle in a case relating to a decision handed down by a Belgian assize court.[12]

[7] Ibid. at 10.

[8] This is linked with Article 8 ECHR.

[9] ECtHR (Grde Ch.), 17 October 2019, no. 1874/13 and 8567/13, *Lopez Ribalda and others* v. *Spain*, § 131. See also Jean Herveg and Jean-Marc Van Gyseghem, "La protection des données à caractère personnel en droit européen: chronique de jurisprudence 2019" (2020) 1 *Journal européen des droits de l'homme / European Journal of Human Rights* 30.

[10] ECtHR, 17 January 2019, X v. *the former Yugoslav Republic of Macedonia*, no. 29683/16.

[11] Ibid.

[12] ECtHR (grand chamber), *Taxquet v. Belgium*, no. 926/05, 16 November 2010.

The applicant had been convicted by the assize court in Liège of murder and attempted murder. At that time, sitting juries had to answer yes or no to questions asked by the president of the assize court. There was no reasoning for the decision, only an arithmetic calculation of the answers given to the various questions that lead to a decision of guilt or acquittal. The applicant therefore brought an action before the ECtHR on the ground that the judgment of the assize court violated Article 6 §§ 1 and 3 (d) of the ECHR, inter alia, on account of the failure to give a reasoned judgment. The ECtHR considered that "the questions, which were succinctly worded and were identical for all the defendants, did not refer to any precise and specific circumstances that could have enabled the applicant to understand why he was found guilty."[13] Prior to that, the ECtHR pointed out that "the national courts must indicate with sufficient clarity the grounds on which they base their decisions"[14] and that such a statement of reasons obliges "judges to base their reasoning on objective arguments, and also preserve the rights of the defence,"[15] "it must be clear from the decision that the essential issues of the case have been addressed."[16] It should be noted, however, that the ECtHR also made it clear that the absence of a statement of reasons does not automatically entail a violation of Article 6 ECHR. Indeed, it is also necessary to ascertain whether other elements of the procedure could make up for the lack of a statement of reasons.

Another aspect of transparency lies in the public nature of the hearing: the hearings before any court must be public – with some exceptions. This publicity "contributes to the achievement of the aim of Article 6(1) ECHR, namely a fair trial, the guarantee of which is one of the fundamental principles of any democratic society."[17]

Consequently, whether we are at the level of Article 6 or Article 8 ECHR, transparency is required.

14.2.2 *Transparency and AI*

It is necessary to question the compatibility of AI with the principle of transparency. While AI is basically the result of human creation, it subsequently develops its own knowledge. This development takes place using the algorithm that underlies its relative "intelligence," as we saw in Section 14.1. AI works in secret, and no one is able to assist in its internal "deliberations." How can such secrecy be compatible with the transparency required by both Article 6 and 8 of the ECHR? Furthermore, if it is not compatible, how can AI be used in court?

The OECD published a set of recommendations on the use of AI. One of them highlights the fact that AI actors commit to transparency regarding AI.[18] This being said, the question remains

[13] Ibid. at § 96.

[14] Ibid. at § 91. See also ECtHR, *Hadjianastassiou* v. *Greece*, no. 12945/87, 16 December 1992 at § 33.

[15] ECtHR, n. 12 at § 91.

[16] Ibid. See also *Boldea* v. *Romania*, no. 19997/02, § 30, 15 February 2007.

[17] ECtHR, "Guide on Article 6 of the European Convention on Human Rights (Criminal Limb)" (April 30, 2020) at 48. See also ECtHR, *Riepan* v. *Austria*, no. 35115/97, § 27; *Krestovskiy* v. *Russia*, no. 14040/03, § 24; *Sutter* v. *Switzerland*, no. 8209/78, § 26.

[18] OECD, "Recommendation of the Council on Artificial Intelligence," OECD/LEGAL/0449 (May 22, 2019, https://oecd.ai/assets/files/OECD-LEGAL-0449-en.pdf (accessed August 5, 2020). "AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art:

(i) to foster a general understanding of AI systems,
(ii) to make stakeholders aware of their interactions with AI systems, including in the workplace,
(iii) to enable those affected by an AI system to understand the outcome, and,
(iv) to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision."

whether the use of algorithms in the work of the judiciary meets the requirement of transparency at the level not only of decision-making, but also of the data processing that is carried out.

The European Commission for the Efficiency of Justice (CEPEJ) published the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, in December 2018. The CEPEJ highlights that the lack of transparency might come from intellectual property issues. Indeed, and as already mentioned, the creation process behind each AI is likely to imply patents or copyright, trade secrets, etc. This situation leads to a protection of the creation (source code, etc.) that is likely to be in conflict with transparency. CEPEJ recommends that "a balance must be struck between the intellectual property of certain processing methods and the need for transparency (access to the design process), impartiality (absence of bias), fairness and intellectual integrity (prioritising the interests of justice) when tools are used that may have legal consequences or may significantly affect people's lives."[19] The Charter thus points out the issues raised by the tension between the use of AI and the duty of transparency required by fundamental rights.

There is a delicate balance to be struck between the right of the designer of the algorithm to keep his creation secret and the right to know what the algorithm hides. This is even more true when the right to a fair trial is at stake. This balance will not be easy to find, as the holder of the intellectual property right will be extremely reluctant to disclose the codes of the algorithms. In trying to find a solution, CEPEJ highlights options ranging from a total technical transparency to an audit of the system by independent authorities or a certification granted by public authorities with regular reviews.[20]

Calls for "open source" algorithms seem to be misleading: it is hard to imagine a developer making an algorithm completely transparent, after having invested time and money in its development. It also seems useless to demand such transparency, which clashes with other principles relating to intellectual property. Developers usually rely on intellectual property rights (IPR) or other legal and technical protections for their licensing strategy. Transparency may trigger tensions between the need to create new applications and the need to protect investments. However, such transparency might be reached when the public authority is the source of the algorithm (which is rarely the case).

It can be observed that many companies are not in favor of licensing their product under an open scheme such as open source. This trend finds its justification in the fact that innovation needs IP protection to remunerate investments. The reluctance seems even stronger when dealing with algorithms such as those used in the field of AI, where competition is strong. Indeed, such inventions are at the core of the business model of many companies. As discussed above, this reluctance to license under open schemes affects the transparency principle of privacy protection and a right to a fair trial.

Loomis v. *Wisconsin*[21] is an example of this lack of willingness to be transparent. In that case, the US Supreme Court refused to consider Mr. Loomis' appeal. Mr. Loomis applied to the US Supreme Court to gain access to the source code of the software named COMPAS, on fair trial grounds. Mr. Loomis had been sentenced to a prison term by the Supreme Court of Wisconsin, which had based its decision on the results of COMPAS. This software calculates the risk of a person reoffending within two years on the basis of 137 analytical criteria. Before the Wisconsin

[19] CEPEJ, "European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment," adopted at the 31st plenary meeting, Strasbourg (December 3–4, 2018), https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (accessed July 13, 2020).

[20] Ibid. at 11.

[21] *Loomis* v. *Wisconsin*, cert. denied, 137 S.Ct. 2290 (2017).

Supreme Court, Mr. Loomis argued that "COMPAS reports provide data relevant only to particular groups and because the methodology used to make the reports is a trade secret" he "asserted that the court's use of the COMPAS assessment infringed on both his right to an individualized sentence and his right to be sentenced on accurate information."[22]

Furthermore, the lack of openness might also impact the availability of the results created by AI, and the availability of the data reduces the possibility to improve algorithms. Indeed, algorithms need to be fed with data to improve; if the amount of data is reduced, the evolution of algorithms is curtailed and, consequently, there could be a reduction in the quality of the results, as well as the competition between developers. This, in other words, could mean that only big companies would have the ability to improve algorithms; by reducing the competition, there is a high risk of creating a monopolistic position, with a reduction of the quality of services, rising costs, and so on. For these reasons, there are now growing demands for the use of open data (and not open source), which would allow smaller developers to create AI systems with fewer constraints. It should incidentally be noted that the European Union promotes open data as an instrument for research.[23]

While it seems unrealistic to demand open-source AI, it seems desirable to require more transparency on the way the algorithm works. In other words, the developer should provide the public with "key subsets of information about the algorithms . . . for example which variables are in use, which goals the algorithms are being optimized for, the training data and average values and standard deviations of the results produced, or the amount and type of data being processed by the algorithm."[24] This would likely meet the requirement of transparency of Article 8 ECHR, as well as Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR hereafter) and Directive (EU) 2016/680 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offenses or the execution of criminal penalties, and on the free movement of such data (the Directive hereafter), and Article 6 ECHR.

There must be transparency about the source of the data, as implicitly demanded by both Article 6 and 8 ECHR. Indeed, Article 6 ECHR requires equality of arms, and respect for the adversarial process. This, in turn, implies that the data being processed must be subject to a contradictory control by the parties, as regards both the data's quality and lawfulness. This transparency of the source of the data is also required in terms of privacy. Thus, the transparency principle contributes to the guarantee of the informational self-determination of the data subject and acts as a control on the elements on which the judge bases the analysis of the case. This is concretized by an obligation to provide information, access, etc.

When considering the use of AI in the course of justice, substantial weight should be given to the risk of lack of fairness (including transparency concerns), but also the benefits that AI can bring about. Benefits include faster justice (speedier decisions) and more consistency across cases and decisions. These are the two major points highlighted by the various reports on the use of AI in the justice system. The Council of Europe states that

[22] *State* v. *Loomis*, 881 N.W.2d 749 (Wis. 2016) (2017) 130 Harvard Law Review 1530; see also Ellora Israni, "Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis," *Jolt Digest* (August 31, 2017), https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1 (accessed May 5, 2020).

[23] See, e.g., the Health Programme Database, https://data.europa.eu/euodp/fr/data/dataset/health-programmes-database (accessed July 13, 2020).

[24] MSI-NET – CoE, n. 1 at 38.

> the trend towards using automated processing techniques and algorithms in crime prevention and the criminal justice system is growing. Indeed, there may be some benefits in such use as massive data sets may be processed more speedily or flight risks assessed more accurately. Moreover, the use of automated processing techniques for the determination of the length of a prison sentence may allow more even approaches to comparable cases.[25]

However, any judgment is built around the elements brought by the parties, including the prosecutor and investigators, in compliance with the applicable legislation. The parties must respect the applicable legislation, including the one governing privacy. This means that the parties must have the opportunity to check the legality of the evidence and, more specifically, judges have to base their decisions on these elements combined, as the case may be, with his or her own perception of the elements. However, and as stated in the ECHR, "the question which must be answered is whether the proceedings as a whole, including the way in which the evidence was obtained, were fair. This involves an examination of the 'unlawfulness' in question and, where a violation of another Convention right is concerned, the nature of the violation found."[26] And the ECtHR added that: "In that context, regard must also be had to whether the rights of the defence have been respected, in particular whether the applicant was given the opportunity of challenging the authenticity of the evidence and of opposing its use, as well as the opportunity of examining any relevant witnesses."[27]

When it comes to AI, the algorithm may have access to a large amount of data available on the Internet, creating a "data lake." This brings us to the question of big data, which is typically summarized by reference to the so-called five V's:[28]

- Volume: the amount of data processed over an extremely short time is enormous;
- Velocity: the processing of data is extremely fast;
- Variety: the data is available in many different forms (structured, text, images, etc.);
- Truthfulness: this concerns the credibility or veracity of the data;
- Value: the data must bring an added value in regard to user-defined goals.

Antoinette Rouvroy points out some issues raised by big data from a privacy perspective. One of the issues that can impact the use of AI in the courts is that "in the context of Big Data, it is the exponential quantity, and not the quality of the processed data that makes automated processing potentially problematic for the rights and freedoms of individuals."[29] Rouvroy argues that big data focuses more on the quantity than on the quality and concludes that:

> by definition, big data are massive amounts of data, a phenomenon that is in direct opposition to the major European principles of data protection, including the principles of minimization (only data necessary for the purpose) and purpose (data only collected for an identified, declared

[25] Ibid. at 10.

[26] ECtHR, *Allan* v. *United Kingdom*, no. 48539/99, 5 November 2002 at §42.

[27] Ibid. at §43.

[28] The Council of Europe defines big data as "the growing technological ability to collect, process and extract new and predictive knowledge from great volume, velocity, and variety of data. In terms of data protection, the main issues do not only concern the volume, velocity, and variety of processed data, but also the analysis of the data using software to extract new and predictive knowledge for decision-making purposes regarding individuals or groups." CoE, "Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data," T-PD(2017)1 at 2.

[29] A. Rouvroy, "Homo juridicus est-il soluble dans les données?" www.researchgate.net/publication/321193294_Homo_juridicus_est-il_soluble_dans_les_donnees (accessed July 13, 2020); see also A. Rouvroy, "Des données et des hommes: Droits et libertés fondamentaux à l'ère des données massives," Conseil de l'Europe, T-PD-BUR (2015) 09 REV, January 2016; and D. Gray and D. Keats Citron "The Right to Quantitative Privacy" (2013) 98 *Minnesota Law Review* 62.

and legitimate purpose), time limitation (data must be erased once the purpose has been achieved, and may not be used, with some exceptions, for other purposes than those initially declared) .... Big Data, on the contrary of minimization, is the maximum collection, automatic, by default, and unlimited storage of everything that exists in digital form, without there necessarily being a purpose established a priori: the usefulness of the data only becomes apparent along the way, thanks to the statistical practices of data-mining, machine-learning, etc. A priori useless data may prove extremely useful in the long run for profiling purposes, for example, and become more useful as the data sets grow larger.[30]

In the context of justice, the two values of volume and truthfulness raise issues. The issue of volume was discussed above, the problem of truthfulness will be discussed here of data. Assuming that having mass data does not pose a problem, it is still necessary to have quality data, especially when such data is being used as a basis for a judicial decision – a decision that will necessarily have effects, positive or negative, on the concerned individual. "Data analysis algorithms are applied to large amounts of data to find patterns of correlation within datasets without necessarily making a statement on causation .... The use of data mining and pattern recognition without 'understanding' their correlation or causal relationships may lead to errors and raise concerns about data quality."[31] Is the data source reliable? Is the data continuously updated? These are the questions that must necessarily be asked when AI is used in the administration of justice. The quality of the data is also a question for the existing legal databases used by the algorithms. Let us imagine that Mr. X appears in court for assault and battery. He acknowledges the facts and will therefore be convicted with a moderate sentence due to his confession. However, the algorithm processed by AI finds, in the databases that it has access to, a previous judgment rendered in another country, convicting Mr. X based on similar facts. With this new element, the AI system could recommend the sentencing of Mr. X to a heavier penalty, on the grounds that he is a recidivist. However, it turns out that the judgment found by the algorithm and used to set the sentence had been overturned on appeal, but that decision was not accessible. In this scenario, the data had obviously not been updated. This example shows the need for the citizen to know where the data comes from and whether it is current. In sum, although AI has a large volume of data at its disposal, this does not mean the AI outcome is reliable. The AI can and should be, at most, only an aid to the decision, but not the decision-maker itself. It must, moreover, be accompanied by transparency.

The above hypothetical raises the issue of the integrity of automated decision-making. The AI will process data received or taken from databases and will deliver a decision (even under the form of a suggestion). Both the GDPR and the Directive address the issue of automated decisions by recognizing the principle of prohibition.

Automatic decision-making also includes profiling, which is defined as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements."[32] Because of the dangers of profiling, the Directive prescribes a prohibition on automated decision-making, unless the member state establishes appropriate safeguards. The same holds true for the GDPR, which also lays down a principle of prohibition, with exceptions that must be interpreted restrictively.

[30] Rouvroy, n. 29.
[31] MSI-NET – CoE, n. 1 at 6.
[32] Art. 4(4) GDPR.

The rationale behind these two provisions is to prevent individuals from being profiled without their knowledge and without any rules to protect their data or, more generally, their privacy. From this point of view, we can link this principle to the notion of fairness in Article 6 ECHR.

It seems important to note that if AI is introduced in the context of court proceedings, it must remain under the control of the user who is, for the purposes of this contribution, the judge. The CEPEJ[33] expresses this concern stating that the AI must help the user to gain autonomy, instead of reducing it. This also means that a judge must be able to control the automatic decision, without being bound to it. However, as highlighted by the CEPEJ, this requires an education of the users with respect to legal tech (LT), so as to allow them to understand how to control the decisions generated by these technologies and the limits of AI. These arguments support the view that AI should be seen as a decision-making tool, but not as a decision-maker.

In order to guarantee transparency, the citizen and the parties must necessarily have access to the data that had been processed by AI, in order to have the opportunity to challenge its veracity and bring counterarguments. However, is it possible for the citizen or even his or her lawyer to analyze the large volume of data processed by AI? In order to reduce the amount of data to be challenged by the citizen, the judicial decision must be very clear about the elements that the judge used to arrive at that decision. Consequently, this means that the work performed by the AI must be clearly identified and controlled before a binding decision. In other words, a human (judge) must validate the AI processing, as requested is rendered by both the Directive and GDPR.

In sum, legislative initiatives will have to be taken to ensure a transparent and fair trial as required by Article 6 ECHR, and subsequently by the Directive and the GDPR, with respect to AI. These initiatives will need to ensure transparency by providing key subsets of information about the algorithms to the public, the source of the processed data, and how decisions are reached. Besides, the concept of empowerment requires that the data subject be given more control over the subject's personal data.

14.3 IMPARTIALITY AND PRESUMPTION OF INNOCENCE

14.3.1 *Principle*

Article 6(2) ECHR stated that "Everyone charged with a criminal offence shall be presumed innocent until proved guilty according to law." The ECtHR in various cases viewed that:

as a procedural guarantee in the context of a criminal trial itself, the presumption of innocence imposes requirements in respect of, amongst others, the burden of proof (*Telfner* v. *Austria*, § 15); legal presumptions of fact and law (*Salabiaku* v. *France*, § 28; *Radio France and Others* v. *France*, § 24); the privilege against self-incrimination (*Saunders* v. *the United Kingdom*, § 68); pre-trial publicity (*G.C.P.* v. *Romania*, § 46); and premature expressions, by the trial court or by other public officials, of a defendant's guilt (*Allenet de Ribemont*, §§ 35–36, *Nešťák* v. *Slovakia*, § 88).[34]

In other words, the defendant is presumed innocent as long as no definitive conviction has been pronounced.

[33] CEPEJ, n. 19.
[34] ECtHR, n. 17 at 58.

The concept of impartiality is also a major element to the right to a fair trial as set by Article 6 (1) ECHR, providing that "in the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law."

The ECtHR has stated in various decisions that:

> Article 6(1) ECHR requires a tribunal falling within its scope to be impartial. Impartiality normally denotes the absence of prejudice or bias and its existence or otherwise can be tested in various ways (*Wettstein* v. *Switzerland*, § 43; *Micallef* v. *Malta* [GC], § 93; *Nicholas* v. *Cyprus*, § 49). The concepts of independence and impartiality are closely linked and, depending on the circumstances, may require joint examination (*Ramos Nunes de Carvalho e Sá* v. *Portugal* [GC], §§ 150 and 152 – see also, as regards their close interrelationship, §§ 153–156; *Sacilor Lormines* v. *France*, § 62). The defects observed may or may not have been remedied during the subsequent stages of the proceedings (*Helle* v. *Finland*, § 46; *Denisov* v. *Ukraine* [GC], §§ 65, 67 and 72).[35]

14.3.2 *Impartiality and Presumption of Innocence and AI*

If, at first glance, AI gives the impression that it can only be fair given the absence of feelings; in fact, however, it remains a human creation. Behind all AI, there is human work. The Council of Europe has rightly pointed out that the "algorithms replicate the functions previously performed by human beings but involve a quantitatively and qualitatively different decision-making logic to much larger amounts of data input."[36] It also raised a major point about human intervention in the creation of the algorithm, by pointing out that:

> In the field of crime prevention, the main policy debates regarding the use of algorithms relate to predictive policing. This approach goes beyond the ability of human beings to draw conclusions from past offences to predict possible future patterns of crime. It includes developed automated systems that predict which individuals are likely to become involved in a crime, or are likely to become repeat offenders and therefore require more severe sentencing. It also includes systems meant to predict where crime is likely to take place at a given time which are then used for prioritizing police time for investigations and arrests. Such approaches may be highly prejudicial in terms of ethnic and racial backgrounds and therefore require scrupulous oversight and appropriate safeguards. Often the systems are based on existing police databases that intentionally or unintentionally reflect systemic biases.[37]

The question of bias is crucial because it can lead to discrimination grounded on, for instance, gender, race, ethnic or sexual orientation. This would bring justice back to the darkest years of the European continent, such as 1930–1945. The question is unfortunately not only theoretical, since situations of algorithms corrupted by bias have already been discovered. For example, the aforementioned COMPAS software has been criticized by authors[38] who found that some of the criteria taken into account by the algorithm were, albeit indirectly, linked to race. This, of course, opens the door to racial prejudice.

[35] Ibid. at 48.
[36] MSI-NET – CoE, n. 1 at 6.
[37] Ibid. at 11–12.
[38] J. Larson, S. Mattu, L. Kirchner, and J. Angwin, "How We Analyzed the COMPAS Recidivism Algorithm," Propublica (May 23, 2016), www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (accessed August 21, 2021).

The stories that gave rise to the analysis by Issac and Lum[39] are as follows:

- An eighteen-year-old girl, who already had a criminal record for acts committed while a minor, was arrested for attempting to steal an unlocked bicycle and scooter worth $80 on the street with another teenager of the same age. Her data was entered into a computer program that determined that she was at high risk of re-offending.
- A forty-one-year-old man was arrested for stealing $86.35 worth of tools from a store. This man had previously been sentenced to five years in prison for armed robbery and attempted armed robbery. His data was encoded in a software program that determined that the risk of recidivism was low.

The results obtained are troubling given the criminal background of each of them. After analysis, it turned out that there was a difference between the two individuals – the color of their skin: the teenager was black, and the man was white. Ironically, a review of the records two years later showed that the teenager had not been charged with any new crimes, while the man was now serving an eight-year sentence for breaking into a warehouse and stealing thousands of dollars' worth of electronic equipment.

With reference to COMPAS, Larson, Mattu, Kirchner, and Angwin[40] reveal that "black defendants were often predicted to be at a higher risk of recidivism than they actually were."[41] *A contrario*, "white defendants were often predicted to be less risky than they were."[42] Their analysis also "showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 45 percent more likely to be assigned higher risk scores than white defendants."[43]

This confirms Kraemer, van Overveld, and Peterson's opinion that:

> some algorithms clearly produce genuine value-judgments. Consider, for example, algorithms used in decision support programs, i.e. systems that help decision makers to make better decisions by ranking a set of alternative actions with respect to some predefined criteria. A typical outcome of an algorithm used in such a program is a verdict like "Alternative X is the best option" or "Alternative X is better than alternative Y with respect to criterion Z." It would be pointless to deny that these sentences express genuine value-judgments.[44]

The authors conclude that "a strong case can be made for the claim that some algorithms are essentially value-laden. Some algorithms, such as those used for classifying cells as diseased or non-diseased, forces the designer of the algorithm to take a stand on controversial ethical issues, e.g. whether it is more desirable to prefer false positive errors over false negative ones."[45]

[39] William Issac and Kristian Lum, "To Predict and Serve? Significance," The Royal Statistical Society (October 10, 2016), http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2016.00960.x/epdf (accessed May 15, 2020).
[40] Larson, Mattu, Kirchner, and Angwin, n. 38.
[41] Ibid. The authors found that "black defendants who did not recidivate over a two-year period were nearly twice as likely to be misclassified as higher risk compared to their white counterparts (45 percent vs. 23 percent)."
[42] Ibid. The authors found that "white defendants who re-offended within the next two years were mistakenly labeled low risk almost twice as often as black re-offenders (48 percent vs. 28 percent)."
[43] Ibid. The authors found that "Black defendants were also twice as likely as white defendants to be misclassified as being a higher risk of violent recidivism. And white violent recidivists were 63 percent more likely to have been misclassified as a low risk of violent recidivism, compared with black violent recidivists" and that "the violent recidivism analysis also showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 77 percent more likely to be assigned higher risk scores than white defendants."
[44] F. Kraemer, K. van Overveld, and M. Peterson, "Is There an Ethics of Algorithms?" (2011) 13(3) *Information & Communications Technology Law* 251.
[45] Ibid.

The same type of bias can be encoded in relation to, for instance, ethnicity or geographical location (e.g., place of residence), in predictive criminal software. Unchecked, these biases can lead to unacceptable injustices in our democratic society. The risk of bias is high, and can lead to biased decisions that are not compliant with Article 6 ECHR. Indeed, these biases jeopardize the presumption of innocence: as mentioned above, the study by Larson, Mattu, Kirchner, and Angwin highlighted major violations. An AI – if created with bias, intentionally or not – might determine in advance that someone is at risk of committing a crime, on the basis of elements whose quality has not been demonstrated. This shows again the necessity of having human control over the way AI works.

Also in Europe, some jurisdictions make use of predictive software. Namely, this kind of software has been set up by the Durham police to predict the risk of an individual committing an offense within a certain period.[46] Whether or not the individual will be included in a reintegration program will depend on the result obtained from the process conducted by the software. The algorithm is supposed to predict an offending act based on thirty-four factors such as gender, criminal record, age, place of residence, etc. It should be noted that twenty-nine of these thirty-four factors are related to the individual's criminal record. Oswald, Grace, Urwin, and Barnes' analysis of the system concludes that "there is a sub-set of decisions around which there is too great an impact upon society and upon the welfare of individuals for them to be influenced by an emerging technology; to an extent, in fact, that they should be removed from the influence of algorithmic decision-making altogether."[47]

With respect to criminal justice, Leroux aptly points out that:

> criminal litigation encompasses diverse realities, not all of which are quantifiable or objectifiable. Thus, the reasoning followed by the judge in concluding the guilt or innocence of a suspect, while it is certainly based in a decisive manner on the objective elements revealed by the investigation and included in the file, can also be nourished by considerations that are not all likely to be brought together in an equation, because they are linked to feelings or emotions. In this respect, the calculation of probability delivered by analytical justice seems to us to be ill-suited to the decision-making process relating to the guilt or innocence of a suspect.[48]

Justice, in other words, is not a simple matter that can be dehumanized and entrusted exclusively to a software. Even in a "simple" traffic accident, the assessment of responsibilities can be delicate. Predictive software, hence, could be contrary to the principle of the presumption of innocence.

So far, we have analyzed software used in predictive analyses of recidivism, when the court is already seized of the accused's case file. However, the same type of software could also be developed to predict offenses by individuals who are not being accused and standing trial yet. In this "big brother" scenario, individual behavior would be analyzed outside of the context of criminal litigation, to predict any indictable offense. Needless to say, this use of the software would have an even greater impact on human rights, as well as privacy. In the worst-case scenario, an individual may be arrested and convicted not for what he or she did, but for what the AI claims he or she will do. If the presumption of innocence is already widely violated with software such as COMPAS and HART, this use of predictive technology is even more

[46] M. Oswald, J. Grace, S. Urwin, and G. Barnes, "Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality" (2018) 27(2) *Information & Communications Technology Law* 233.

[47] Ibid.

[48] O. Leroux, "Justice pénale et algorithme" in J. B. Hubin, H. Jacquemin, and B. Michaux (eds.), *Le juge et l'algorithme: Juges augmentés ou justice diminuée* (Brussels: Collection du Crids, 2019) 61 (loose translation).

problematic in predicting criminal acts. Furthermore, in light of the aforementioned risk of bias, entire categories of people would risk being charged with intent to offend, or being put under surveillance, even when they have nothing improper or illegal. In sum, the use of predictive software in criminal justice entail risks for the principle of fairness and presumption of innocence, which are difficult to accept in democratic societies.

14.4 EQUAL ACCESS TO JUSTICE

This section will consider legal analytics, that is, AI analyzing the jurisprudence of courts, or individual judges. This technology can be used by courts to reach a decision; the use of this type of software may entail gains in consistency of the case law and avoid disparity from one court to another. Many will see this as a major step forward in the search for an egalitarian justice. However, this development also entails risks. For instance, these analytical tools may be used not only by the courts, but also by individuals committing criminal acts, who would have an opportunity to adapt their criminal behavior based on decisions rendered in similar cases. Individuals, in other words, would be facilitated in their cost-benefit analysis, while undertaking criminal activities.

To be sure, legal analytics should not be prohibited; however, we need to be aware of the deviations to which it may be subject. Leroux points out that "these applications ... make it possible to determine which courts are likely to take a more favorable decision and, within these courts, which judges (identified by name) could be more lenient or stricter."[49] Leroux therefore notes that legal analytics may encourage a propensity to "forum shop". This raises not only ethical doubts, but also legal questions concerning equal access to justice: not all parties will have the same weapons, since the more affluent could benefit from the help of such software, to the detriment of the less affluent.

Indeed, the switch toward an algorithmic justice, autonomous from any human intervention, could create an effect of inequality of arms between parties. As Mougenot and Gérard point out, "it seems obvious that digital modes of dispute resolution are a priori accessible only to people who have the appropriate equipment and who are computer literate, i.e. who have sufficient skills to use these systems. Clearly, such a situation leads to a widening of the digital divide."[50] As a consequence, the switch from a human to algorithmic justice risks of marginalizing a whole category of litigants who are entitled to fair justice, as guaranteed by Article 6 ECHR. Moreover, the ECtHR considered that, based on Article 6 ECHR, governments should take positive measures to ensure access to justice, and the fulfillment this duty requires that countries undertake positive action to ensure that access to justice is effectively guaranteed. For these reasons, the ECtHR has held that litigants suffer a violation of their right of access to justice if the state fails to implement sufficient measures necessary for such access, such as the access to a lawyer.[51] We can, quite logically, draw a parallel between this case and the move from human justice to an algorithmic one. The use of AI as a means of "choosing" one's judge, or adopting one's criminal behavior to evade the justice system, could create a significant inequality of arms

[49] Ibid. at 58–59 (loose translation).

[50] D. Mougenot and L. Gérard "Justice robotisée et droits fondamentaux" in Hubin, Jacquemin, and Michaux (eds.), n. 48 at 41 (loose translation); see also B. Custers, K. La Fors, M. Jozwiak, E. Keymolen, D. Bachlechner, M. Friedewaldand, and S. Aguzzi, "Lists of Ethical, Legal, Societal and Economic Issues of Big Data Technologies," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3091018&download=yes (accessed July 13, 2020).

[51] ECtHR, *Airey* v. *Ireland*, no. 6289/73, 9.10.1979, para 25.

problems. Consequently, justice based on AI would not provide access to justice for all, which mean that governments would fail in their duty to take adequate measures to provide access to justice. A violation of Article 6 ECHR would therefore take place. The use of such software should be regulated by law, so that it can usefully assist judges and courts, rather than as a means to elude of the law.

14.5 FURTHER PROCESSING

Another aspect of AI is the further processing of personal data in the sense of the GDPR. Judicial actors, such as judges and attorneys, have at their disposal databases containing judicial decisions. These decisions might contain personal data, such as surnames, first names (parties, witnesses and judges) and, where appropriate, sensitive data such as health data, sexual life data, etc. Even when the data is apparently anonymized, the advent of big data might make re-identification possible. Anonymity thus might be illusory: research has demonstrated the possibility of re-identifying by using only fifteen attributes of an individual who had previously been anonymized.[52] In other words, the various elements contained in a decision may make it possible to identify the parties, as well as the witnesses or judges. As Mougenot and Gérard rightly point out, "the creation of databases of case law, their conservation and their subsequent use by artificial intelligence systems present a risk not only for the privacy of the litigants, but also for that of the members of the court and third parties."[53]

The data can thus be used for a new purpose: if the initial purpose being the rendering of justice, the new one is to create a database, often with a commercial objective. The question raised by this new purpose is its compatibility with the original one. The answer to this question depends on the obligations incumbent on the managers of these databases, who can be classified as data controllers in the sense of the GDPR. The GDPR adopts a principle of prohibition of further data processing for purposes not compatible with the first processing. De Terwangne reminds us that "the notion of 'compatible' use has given rise to many questions in practice and the authors of the GDPR have been concerned to further define it. Article 6(4) GDPR thus sets out a series of criteria for establishing whether the processing of data for another purpose is compatible with the purpose of the original collection or not."[54] These criteria include the link existing between the two purposes, so that it is possible to "admit all subsequent uses that are linked to, logical and consistent with the stated aims."[55] Besides this, "the context in which the personal data were collected, in particular with regard to the relationship between persons data subjects and the controller"[56] has to be taken into account. The nature of the data is also relevant, in light of "the increased risk of processing sensitive data"[57] together with the

[52] L. Rocher, J. Hendrickx and Y. A. de Montjoye, "Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models" (2019) 10 *Nature Communications* 3069.
[53] Mougenot and Gérard, n. 50 at 48 (loose translation)
[54] C. de Terwangne, "Les principes relatifs au traitement des données à caractère personnel et à sa licéité" in C. de Terwangne and C. Rosier (eds.), *Le Règlement général sur la protection des données (RGPD/GDPR: Analyse approfondie* (Brussels: Larcier, 2018) 97–98 (loose translation).
[55] Ibid.
[56] Ibid. at 97. The author states that "in order to be correctly identified and this criterion should be read in the light of the recital 50, which states: 'the context in which the data in question have been collected, in particular the reasonable expectations of the data subjects, depending on their relationship with the controller, as to the further use of such data.' This criterion of reasonable expectations of the data subject is particularly relevant, since limiting what is done with the data to this which enters into the forecasts of this subject, it allows the latter to retain controls the fate of its data." (loose translation).
[57] Ibid. at 97–98 (loose translation).

"possible consequences of the envisaged further processing for the persons concerned."[58] Furthermore, one must be careful to verify "the existence of appropriate safeguards, which may include encryption or pseudonymisation."[59]

There is no doubt that court decisions entail the processing of sensitive data (belonging to special categories), and that the processing of such data for purposes other than the rendering of justice may have an impact on the data subjects. Often, this makes further processing incompatible with the initial purpose, that is, the rendering of justice.

14.6 CONCLUSION

AI must be surrounded by the best safeguards to ensure that it does not infringe fundamental rights, especially in the area of justice. While triggering many questions, the rise of AI also offers new opportunities for the administration of justice. But does that mean that, in the future, AI will work autonomously in the place of a human judge? This is unlikely due to the problems noted by Irsani relating to the use of predictive software such as COMPAS in the field of criminal justice:

> morally troubling precisely because sentencing should not be easy. Actors in the criminal justice system should lose sleep over the fact that they are systemically depriving people of their life, liberty, and property. That should be hard. It is a serious, unimaginable thing. Anyone who has a hand in this system should have to grapple with the consequences of their work; as algorithms become a part of the criminal justice system, that 'anyone' should include technologists.[60]

If AI is used in the administration of justice, it must be under the supervision and control of the judge. Furthermore, any decision taken based on the results provided by such software must be motivated in comprehensible and clear words, to enable any litigant, whatever his or her level of education, to understand it. The judiciary cannot simply state that "it is the AI that made the decision," as this would be contrary to Article 6 ECHR.

In reality, the use of software could complicate judicial reasoning. Indeed, transparency should be provided on the following aspects of AI systems deployed in the administration of justice:

- key subsets of information about the algorithms;
- the way the algorithm works and how it arrives at the solution;
- the origin of the data;
- the quality of the data (e.g., reliability ratio).

Despite the benefit of using AI in terms of speed and amount of data processed, moving from human-made justice to AI-made justice raises many problems linked to the ECHR, the GDPR and Directive (EU) 2016/680 of 27 April 2016, which guarantee the respect of the fundamental rights of individuals. In conclusion, we should be in favor of the use of AI as an aid to decision-making, but certainly not as a judge itself.[61]

[58] Ibid.
[59] Ibid. at 98 (loose translation).
[60] Israni, n. 22.
[61] Mougenot and Gérard, n. 50 at 14.