*Artificial intelligence and criminal procedure: fundamental rights, procedural principles and regulatory challenges**

di **GIULIA CASCONE**

Abstract

Il contributo si sofferma su alcune problematiche regolatorie scaturenti dall'intersezione tra procedura penale e intelligenza artificiale. Prendendo le mosse dall'individuazione delle applicazioni attuali e potenziali dei sistemi di intelligenza artificiale nel procedimento penale, sono presi in esame i conseguenti rischi di compressione delle garanzie e dei principi processuali coinvolti, nell'intento di sottolineare l'impellente esigenza di elaborare un quadro giuridico chiaro, suscettibile di bilanciare i benefici derivanti dall'uso dell'intelligenza artificiale con il rispetto dei diritti e dei principi fondamentali della materia.

1. Artificial intelligence in criminal procedure: the current scenario. Artificial intelligence[1] applications are rapidly spreading their influence across various domains. Fueled by massive amounts of data, these systems are radically changing the way we live, work and interact with each other. In this broader context, criminal procedure is one of the areas in which AI is gaining more and more significant traction, aiding law enforcement agencies and legal professionals to better perform their tasks.

* Il presente articolo è stato sottoposto a revisione e accettato per la pubblicazione in data antecedente all'approvazione, da parte del Parlamento Europeo, del Regolamento europeo che stabilisce regole armonizzate sull'intelligenza artificiale (cosiddetto "Artificial Intelligence Act"), avvenuta in data 13 marzo 2024. Nondimeno, i commenti e le osservazioni esposti nel contributo (e, in particolare, nel par. 4) sulla disciplina risultante dalla proposta avanzata dalla Commissione Europea risultano attuali, in quanto la versione dell'atto normativo da ultimo approvata dal Parlamento Europeo non presenta, in relazione agli aspetti presi in esame, modifiche sostanziali rispetto al testo della proposta commentata nel presente articolo.

[1] On the difficulty of finding a shared definition of Artificial Intelligence (hereinafter, also "AI"), see Finocchiaro, *The regulation of artificial intelligence*, in *AI & Soc* (2023), available at https://doi.org/10.1007/s00146-023-01650-z. In the recent proposal for the adoption of the AI Act (*European Commission Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*, Brussels, 21.4.21), AI is defined as a «software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with». In its general approach published in May 2023 (Available at https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf.), the European Parliament amended the definition of AI systems to align it with the one given by the Organisation for Economic Co-operation and Development (OECD). According to the amendment, AI should be defined as «a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments».

AI-powered systems are currently being used for many purposes[2] in criminal procedure and further applications are highlighted by the ongoing academic debate on the subject[3]. Without claiming to be exhaustive, a brief reconstruction of the current scenario of the possible applications of AI in criminal procedure is necessary in order to assess if and (if so) how the use of such systems can impact on fundamental rights of individuals and affect some basic criminal procedural principles.

Among the wide range of possibilities that AI offers, the most disruptive ones are those that involve replacing human judges with automated systems. Notably, initiatives such as the e-court system in the Netherlands and the project for a robot judge in Estonia exemplify this emerging trend.

The first is a private online court that automates and streamlines certain aspects of legal proceedings. This AI-powered platform handles smaller civil cases, offering a faster and more efficient resolution of the case[4].

An even bolder project is the one proposed in Estonia for the creation of a robot judge. The news concerning the development of such system rapidly spread in 2019[5], prompting the attention of the technical and scientific community. In early 2022, however, the Estonian Ministry of Justice released a statement in which the government's commitment to the development of a robot judge was denied[6]. Nonetheless, we read in the statement that «Ministry of Justice is also interested of AI projects and will look opportunities where AI could be useful and does not exclude the possibility to use the AI solutions in the future to assist judges and court officials». It is unclear whether the news was a hoax or whether the Estonian Government simply took a step backwards on the Country's digital modernisation plan.

To date, artificial intelligence systems designed to replace human judges do not seem to receive wide acceptance and are therefore not particularly popular in Western judicial systems.

---

[2] See Nieva Fenoll, *Inteligencia artificial y proceso judicial*, Marcial Pons, 2018, 23.

[3] One of these is the use of artificial intelligence techniques for biometric recognition of emotions in order to assess the reliability of a witness. At present, there is no evidence of the implementation of such systems across the European jurisdictions, but it cannot be excluded that they may be used in the future. See Cascone, *Emotional biometrics: a preliminary analysis of critical aspects concerning the use of the last AI frontier in criminal procedure,* in VV.AA., *El Proceso en tiempos de cambio. VII Processulus, Encuentro de jóvenes investigadores en derecho procesal*, Colex, 2023, 292.

[4] On this topic, see Nakad-Weststrate, Van Den Herik, Jongbloed, and Salem 2015, *The Rise of the Robotic Judge in Modern Court Proceedings*, in *Conference Paper. The 7th International Conference on Information Technology*, 2015, 59–67, available at http://icit.zuj.edu.jo/ICIT15/DOI/Artificial_Intelligence/0009.pdf.

[5] Niiler, *Can AI Be a Fair Judge in Court? Estonia Thinks So*, in Wired, March 25, 2019, https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/.

[6] Available at https://www.just.ee/en/news/estonia-does-not-develop-ai-judge.

On the other hand, a great deal of software has been developed in order to aid (not replace) the judges and legal practitioners in the performance of their tasks.

Some of these tools seem to be particularly promising in terms of "collateral" support to judicial activity and not so dangerous as regards the respect for fundamental and procedural rights. It is the case of software that provide judges with a selection of previous similar cases, prepare drafts or perform more complex tasks such as analyze the consistency and coherence of testimonies, comparing statements against other available evidence to identify any contradictions or discrepancies, and so on[7].

On the contrary, many concerns have been addressed to risk assessment tools. Within this category, systems such as COMPAS, developed in the United States, and HART, developed in the United Kingdom, are able to assess the risk of recidivism[8]. This kind of tools can prove to be very useful both in the pre-trial phase (for example, to assess the necessity to apply a pre-trial restriction of liberty) and in sentencing[9] (for example, when it comes to quantify the sanction of imprisonment as a consequence of the conviction of the accused person). As is well known, they have generated much debate in recent years as to their reliability and the existence of biases in their functioning that have led to discriminatory outcomes[10].

Although designed for a partially different use, very similar in the functioning are those tools developed to predict[11] the outcome of pending or future court cases[12]. Such systems make it possible to calculate the chance of success of a case and "predict" how a particular court will decide on it. For instance, the SCOTUS system[13] has been applied

[7] Ulenaers, *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?*, in *Asian Journal of Law and Economics,* vol. 11(2), 2020, 10.

[8] See Lupo, *Regulating (Artificial) Intelligence in Justice: How Normative Frameworks Protect Citizens from the Risks Related to AI Use in the Judiciary, European Quarterly of Political Attitudes and Mentalities, EQPAM*, 81.

[9] See Kaspar, Harrendorf, Butz, Höffler, Sommerer & Christoph, *Artificial Intelligence and Sentencing from a Human Rights Perspective*, VV.AA., *Artificial Intelligence, Social Harms and Human Rights*, Palgrave Macmillan, edited by Završnik and Simončič, 2023, 3.

[10] The admissibility of evidence based on COMPAS system was addressed in the famous judgment State of Wisconsin v. Loomis, 881 N. W. 2d 749 (Wis. 2016). A 2016 study from ProPublica demonstrated that COMPAS assigned African Americans a higher (and discriminative) risk rate of recidivism than the one assigned to white people. See Contissa & Lasagni, *When it is (also) Algorithms and AI that decide on Criminal Matters: In Search of an Effective Remedy*, in *European Journal of Crime, Criminal Law and Criminal Justice*, *28(3)*, 2020, 284.

[11] But see Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion*, Springer, 2020, 119, on the improper use of the term "prediction". According to the Author, what the AI programs can do is to provide a calculation of how a court or a judge decided in previous similar cases.

[12] See Ulenaers, *The Impact of Artificial Intelligence*, cit., 5. Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., 117.

[13] Nikolaos Aletras and others, *Predicting judicial decisions of the European Court of Human Rights:* a Natural Language Processing perspective, in *PeerJ Computer Science 2:e93*, 2017.

to the decisions of the European Court of Human Rights and has proven to be particularly accurate.

As regards the area of evidence, an example of AI tool that could reveal particularly useful in this field is face recognition systems[14], that employ AI algorithms to analyze facial features and match them against a database of known individuals. This technology can be used to identify suspects or to verify the presence of individuals on the crime scenes not only during the investigation, but also as evidence at trial. By comparing surveillance footage, photographs, or other visual evidence with existing databases, AI-powered face recognition systems offer potential insights to investigators and to the judges and, at the same time, pose relevant concerns for privacy and respect for private and family life[15].

The last frontier in the field seems to be represented by emotional recognition systems based on AI algorithms to analyze facial expressions, voice inflection, and other biometric patterns to infer emotional states. These systems could be used to verify the credibility of witnesses, assessing the consistency between their emotional responses and their declarations by measuring factors such as micro-expressions or changes in vocal pitch[16].

Finally, although they fall outside the scope of this analysis, it is worth mentioning the currently most widespread predictive policing tools, enabling authorities to identify potential crime hotspots and allocate resources. By analyzing historical crime data and patterns, AI algorithms can guide law enforcement agencies in their proactive efforts to prevent crimes[17]. Such software is mainly used in an area (crime prevention) that lies outside the perimeter of criminal proceedings and criminal trials and in which the procedural guarantees are not supposed to operate[18]. Nevertheless, legal scholars have

[14] Buolamwini, Vicente Ordóñez, Morgenstern, & Learned-Miller, *Facial recognition technologies: a primer*, May 29, 2020, https://global-uploads.webflow.com/5e027ca188c99e3515b404b7/5ed1002058516c11edc66a14_FRTsPrimerMay2020.pdf; Faraldo Cabana, *Technical and Legal Challenges of the Use of Automated Facial Recognition Technologies for Law Enforcement and Forensic Purposes*, in VV.AA., *Artificial Intelligence, Social Harms and Human Rights*, Palgrave Macmillan, edited by Završnik and Simončič, 2023, 35.

[15] Neroni Rezende, *Facial recognition in police hands: Assessing the 'Clearview case' from a European perspective*, in *New Journal of European Criminal Law* 11(3), 375-389.

[16] EDPS-TechDispatch # 1_2021, *Facial Emotion Recognition*, European Data Protection Supervisor, 1; Buolamwini, Vicente Ordóñez, Morgenstern, & Learned-Miller, *Facial recognition technologies: a primer*, cit., 8. On the concerns deriving from the use of this technology in criminal proceedings, see Cascone, *Emotional biometrics: a preliminary analysis*, cit., 292.

[17] On algorithmic crime prediction and its criticalities, see Sommerer, *Algorithmic Crime Control between Risk, Objectivity, and Power*, in VV.AA., *The Law between Objectivity and Power*, edited by Bender, Nomos, 2022, 274 ff.

[18] Signorato, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo,* in *Riv. dir. proc.*, 2020, 2, 607. See, however, Quattrocolo, *Artificial Intelligence, Computational Modelling*

highlighted the many concerns that may arise from their use insofar as they compromise citizens' privacy, subjecting them to more or less hidden forms of surveillance. Not to mention that they could lead to distorted outcomes by subjecting the same territories to continuous surveillance due to the increased discovery of crimes facilitated by the use of the software[19].

2. From the impact on fundamental rights to the respect of criminal procedural principles. In the last decade, legal scholars from all over the world have been addressing the many concerns for fundamental rights implicated in the use of AI in judicial systems.

From a general point of view, the same operational characteristics of AI tools presuppose the risk of a violation of the rights to private and family life and to protection of personal data (artt. 7 and 8 CFREU and art. 8 ECHR). On the one hand, it is widely known that the functioning of such systems presupposes the use of huge amount of data in order to train the algorithm to perform its task and this entails massive treatment of (also) personal and sensitive data of individuals[20]. On the other hand, as already mentioned in the previous paragraph, AI (especially the tools which are used in preventive policing) allows the tracking and analysis of daily habits of people, realizing forms of State surveillance[21].

While the risks related to the violation of privacy have a cross-cutting dimension, many negative implications of the use of artificial intelligence have been highlighted with specific reference to criminal proceedings. They range from the respect for certain procedural rights considered to be “fundamental”, to the same preservation of the intimate nature and structure of criminal procedural models of Western judicial systems.

Before briefly reviewing some of the principal concerns, two preliminary remarks are necessary.

First, each of these problematic aspects entails a regulatory challenge for the legislators. As is well known, the European Commission presented a proposal for the

and Criminal Proceedings, cit., 41 on the interference between preventive policing and criminal investigation.

[19] Signorato, *Giustizia penale e intelligenza artificiale*, cit., 608.

[20] See European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, 56, available at https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

[21] European Commission’s white paper on AI, 2020, 11, available at https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf. Also see Algeri, *Intelligenza artificiale e polizia predittiva*, in *Dir. pen. e proc*., 2021, 6, 731.

adoption of a regulation on artificial intelligence (the so-called AI act) in April 2021. It is the world's first systematic regulation in this field. The proposal classifies many of the AI technologies used in the field of law enforcement as "high risk". This means that, in order to circulate in the internal market, they must comply some technical requirements that ensure reliability, traceability and verifiability of the results. However, there are no specific procedural rules and guarantees related to the use of similar systems in criminal proceedings. As a result, it will be up to national legislators to regulate specifical aspects concerning the use of AI in criminal investigations and criminal trial. Such regulation will be aimed at adequately balancing the (expected) benefits of the use of AI in criminal proceedings – in terms of increased efficiency – and the risks of affecting fundamental rights of individuals.

Secondly, as we will see, many of these problematic aspects depend on certain structural and operational features of AI systems. This entails the need to seek technical solutions that can make the functional characteristics of the AI systems compatible with the requirements of criminal proceedings.

One of the first concerns related to the use of AI tools in criminal proceedings consists in allocating responsibility in case of failure[22]. Judges and public prosecutors could rely on AI tools in order to perform complex and sensitive tasks that could have serious consequences on individuals. Although this aspect could concern any application of AI in the justice system, when it comes to criminal proceedings the consequences of an error on the fundamental rights of citizens are likely to be more serious, as they may affect the personal freedom.

The issue becomes even more problematic when considering that, currently, the development of such systems is almost entirely in the hands of private entities, which may not have any interest in developing software capable of reaching sufficient levels of reliability and robustness. As already stressed by legal scholars, this situation entails the necessity to introduce specific regulatory solutions in order to hold private entities accountable for their participation, even if indirect, in public functions[23].

[22] For further references, see Lupo, *Regulating (Artificial) Intelligence in* Justice, cit., 83.
[23] Gascón Inchausti, *Desafíos para el proceso penal en la era digital: externalización, sumisión pericial e inteligencia artificial*, in Conde Fuentes, Serrano Hoyo, **La justicia digital en España y la Unión Europea**, Ed. Atelier,, 2019, 193.

Moving to the specific realm of criminal proceedings, the use of artificial intelligence could infringe multiple procedural guarantees, many of which are linked to the right to a fair trial protected by the European Convention on Human Rights[24].

First and foremost, it has been highlighted that the high level of technical complexity of AI systems can affect the equality of arms. The introduction of scientific knowledge into criminal proceedings often results in an imbalance between the parties, favoring the one (generally, the public prosecutor) that has greater resources to interpret and challenge the cognitive results of the technology at stake. According to this analysis[25], algorithmic evidence exacerbates this imbalance due to the inaccessibility of the source code (for reasons of protecting industrial property) or the inherent opacity[26] of the algorithm.

For the same reason, the use of AI tools in criminal proceeding could impact the right to effective remedy and to access to information which are relevant to perform an effective defense before and during the trial[27].

Another technical characteristic that may have concerning consequences in the procedural realm is linked to the phenomenon known as the "datification"[28]. AI algorithms, as mentioned before, operate on the base of the processing of enormous amounts of data to provide output results through correlations that - in machine learning systems - are created by the AI systems themselves. This can lead to certain distortions.

[24] For an overview of the ECHR rights affected by AI tools in criminal proceeding, see more specifically Ulenaers, *The Impact of Artificial Intelligence*, cit., 17. The rights which could be affected range from the access to court (that could be infringed in case of implementation of AI-judges) to the presumption of innocence, that can be violated in the case of algorithms created for decision-making whose operational parameters do not reflect the distribution of the burden of proof that the principle imposes.

[25] Quattrocolo, *Equità del processo penale e automated evidence alla luce della Convenzione europea dei diritti dell'uomo,* in *Revista Italo-Española de Derecho Procesal*, 2019, 118.

[26] Even if the companies that develop AI software made public the algorithms that the machines use to operate their calculations, it is very often the own IT architecture that does not allow, not even programmers, to reconstruct the statistical associations on the basis of which a certain result was provided. As explained by Bonsignore Fouquet, *Sobre Inteligencia Artificial, decisiones judiciales y vacíos de argumentación,* in *Teoría & Derecho. Rev. pens. Jur.*, 2021, n. 29, 264, the opacity can be of three types: first of all, it can derive from the industrial secret that could cover to the algorithms that support the functioning of the machine; secondly, it could depend on the judge's lack of technical skills in understanding the inferences conducted by artificial intelligence, given its high technological character; finally, it could be an opacity in the strict sense. This is the most marked form of opacity, since it concerns the same functioning of the tool, which is based on the development of new statistical criteria which are directly developed by the machine to match data and whose reconstruction could be impossible even for same programmers. On the topic, also see Pérez Estrada, *La inteligencia artificial como prueba científica en el proceso penal español*, in *Revista Brasileira de Direito Processual Penal*, 2021, 7, 1392; Contissa & Lasagni, *When it is (also) Algorithms and AI that decide on Criminal Matters*, cit., 281.

[27] Contissa & Lasagni, *When it is (also) Algorithms and AI that decide on Criminal Matters*, cit., 290 and 297. The Authors stress that «even where humans formally retain control over the final decision, the possibility of effectively contesting its merits remains at best a remote hypothesis».

[28] See Završnik, *Algorithmic justice: Algorithms and big data in criminal justice settings*, in *European Journal of Criminology*, 18(5), 2019, 633.

The most evident, and consequently the most discussed in literature, is the risk of discrimination due to the use of input data that may be biased since they pertain to past cases that may have been decided with discriminatory criteria from a social, economic, or cultural perspective. For this reason, in the proposal for the adoption of the AI Act, the European Commission emphasizes the importance of adequate data selection for the training of systems[29].

It appears to be related to the same technical characteristic under discussion the concern that relying on AI systems for complex decision-making activities may lead to losing sight of the necessary "individualization" of justice, especially in the criminal field. The homogenization of decisions resulting from the standardized use of the algorithm could potentially suppress the need to distinguish one case from another, which is particularly relevant when determining the punitive treatment following the establishment of a subject's criminal responsibility[30]. And the same distortion could be observed regarding the application and interpretation of the law.

In this regard, some authors have pointed out that the use of algorithms for decisions that must be taken during criminal proceedings could result in a paralysis in the interpretative evolution of law, establishing a normative correlation between how rules have been interpreted in the past and how they will be interpreted in the future[31]. Thus, the multiple nuances of the law[32] and their necessary permanent adherence to the social context would be attenuated, leading to «nullify the virtuous bottom-up spin that induces changes in legal interpretation, in any legal systems. Discouraging potential litigants to avoid going to court; pushing lawyers to stick to arguments that proved successful in previous cases; inducing judge to respect the precedent, even if it is not part of the legal culture, would reduce, extremely, any advancement of the legal culture»[33].

[29] In the *considerandum* 44 of the AI Act proposal, the European Commission stresses that «high data quality is essential for the performance of many AI systems […] Training, validation and testing data sets should be sufficiently relevant, representative and free of errors and complete in view of the intended purpose of the system. They should also have the appropriate statistical properties, including as regards the persons or groups of persons on which the high-risk AI system is intended to be used […] In order to protect the right of others from the discrimination that might result from the bias in AI systems, the providers should be able to process also special categories of personal data, as a matter of substantial public interest, in order to ensure the bias monitoring, detection and correction in relation to high-risk AI systems»

[30] Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., 124.

[31] See, again, Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., 121.

[32] Ulenaers, *The Impact of Artificial Intelligence*, cit., 18.

[33] Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., 212.

The same author underscored that the use of artificial intelligence tools for judicial decision-making could even undermine the foundations of civil law systems, altering the relationship between higher and lower courts. In particular, it could erase any distinction between them since they would rely on the same algorithms, trained with the same data sets[34], with a consequent standardization of the decisions.

It is evident that changes of this type transcend the subjective dimension of respect for the rights of the individuals involved in criminal proceedings[35], reaching the very foundations of the judicial system as we know it.

3. Transparency as principal concern and minimum requirement for AI's application in criminal proceedings. The scenario that unfolds before the interpreter is that of a true paradigm shift. The inclusion of artificial intelligence systems in the administration of justice is much more disruptive phenomenon than that - widely discussed in legal doctrine in recent decades - of scientific evidence. Ultimately, the question is whether and to what extent we should relinquish certain structural characteristics of criminal investigation and prosecution in order to exploit the multiple benefits that AI promises to bring.

According to a part of the legal scholarship, indeed, significant advantages could derive from the use of artificial intelligence systems in terms of efficiency and more transparent and impartial administration of justice, starting with the greater legal certainty resulting from increased uniformity in the application of the law. It has been pointed out that technology-driven decisions could decrease discrepancies in case adjudication, reducing situations where similar cases are decided differently[36]. It has also been emphasized that, thanks to predictive justice, the administration of justice could become more transparent, democratic, egalitarian, and objective[37]. The risks of discriminatory outcomes linked to biases and prejudices of human judges would be attenuated[38] and there would be an improvement in the efficiency of legal practitioners' work, since they

[34] Quattrocolo, *Per un'intelligenza artificiale utile al giudizio penale*, in *BioLaw Journal*, 2021, 2, 393.
[35] For an overview of the changes of practitioners' roles and competencies due to the diffusion of AI tools in the field of law, see Ben-Ari, Frish, Lazovski, Eldan & Dov Greenbaum, *Artificial Intelligence in the Practice of Law: An Analysis and Proof of Concept Experiment*, 23 RICH. J.L. & TECH., 2017, 2.
[36] Prins and Roest, *AI en de rechtspraak: Meer dan alleen de 'robotrechter'*, in *Nederlands Juristenblad* , 2018, 93 (4), 267, available at https://pure.uvt.nl/ws/portalfiles/portal/20232594/NJB_1804_ART_1.pdf.
[37] Larret-Chahine, *Le droit isométrique: un nouveau paradigme juridique né de la justice prédictive*, in *Archives de philosophie du droit*, no. 60, 2018, 287. Similarly, Canzio, *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sist. pen.*, 8 gennaio 2021.
[38] Di Giovine, *Il judge-bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cass. pen.*, 2020, 3, 594.

would be assisted by artificial intelligence in performing their functions[39], being AI capable of analyzing in short time much more data and information than every human being.

In light of this, it is more urgent than ever to answer the question appropriately defined as “ethical” by authoritative doctrine, «whether we need to realize 'predictive justice' tools. This question must be approached, bearing in mind that efficiency in justice administration is a 'balanced' concept. However, that same ethical dilemma could also be reframed in the opposite way: “can we avoid exploiting all the benefits derived from such instruments?”»[40].

It is not the purpose of this brief article to provide a definitive answer to such a complex question, which is only recently gaining the attention it deserves within the legal community. I believe, perhaps wrongly, that the ongoing debate on the regulatory challenges posed by the use of artificial intelligence can lead to legal solutions that achieve a balanced equilibrium of interests at stake, avoiding the transformation of the administration of justice, especially in criminal matters, into a “bureaucratic”, sterile, and mathematical process.

The first fundamental obstacle in this regard, in my opinion, is related to algorithm transparency or “opacity”. We have already touched upon this topic earlier, but it deserves further exploration, at least from two perspectives. On the one hand, on this characteristic of AI algorithms depends the most disruptive effect of the employment of AI in criminal proceeding, even in case of use as mere support for decision-making: I am referring to the potential delegitimation of the exercise of the judicial function. On the other hand, the concept of algorithm “opacity” has been addressed, until now, from a purely technical standpoint. Anyway, it acquires another (*rectius*, a further) meaning when viewed from the perspective of those who rely on it to make decisions that can be crucial during the criminal proceeding, starting with the judge and the public prosecutor. Both issues deserve further explications.

Starting with the first one, it has been suggestively stressed that «no deja de producir cierta sensación de vértigo el hecho de que el resultado de un programa de inteligencia artificial se halle entre los ingredientes que han de conformar decisiones judiciales susceptibles de proyectarse sobre la situación personal del sujeto pasivo del

[39] In this sense, *ex multis*, Lasagni, *Difendersi dall'Intelligenza artificiale o difendersi con l'intelligenza artificialle? Verso un cambio di paradigma*, in *Riv. ita. dir. e proc. pen.*, 2022, 4, 1559.
[40] Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., 123.

proceso»[41]. This statement strikes at the heart of the problem. As in common experience the "feeling of vertigo" is usually associated with something unknown, when it comes to the use of AI in criminal procedure it is related to the fact that «although the machine-induced rules may lead to accurate predictions, they do not refer to human expertise and may not be as intelligible to humans as an expert's manually constructed rules. Since the rules the ML algorithm infers do not necessarily reflect explicit legal knowledge or expertise, they may not correspond to a human expert's criteria of reasonableness»[42]. This lack of intelligibility becomes problematic as it could make it difficult for citizens to understand and accept judicial decisions, perceiving them as fair. Moreover, it has been stressed that the use of AI algorithms for decision-making can result in a loss of independence and autonomy of the judge, primarily because he will naturally be inclined to adhere to the machine's outcome, being this last difficult to contest as it works as a sort of "black box"[43]. Secondly, this fideistic approach may weaken the cognitive and rational character of judgment, in which, similarly to independence and autonomy, lies the foundation of the judge's legitimacy[44]. This ultimately could lead to a form of "codified justice" that favors standardization at discretion and, for this very reason, can dispense with "strong" forms of legitimacy for those who perform it[45].

One could argue that the lack of transparency in algorithms is similar to the opacity of the human mind, but this objection overlooks a crucial distinction. Individuals are aware of mental processes by which other people reason and get to certain decisions: not the same can be said when it comes for a human being to understand the very functioning of an AI algorithm.

Coming to the second aspect, I mentioned that the concept of "opacity" has so far been addressed and defined only from an eminently technical perspective. That is to say, the discussion on "opacity" has focused on the inability to access the source code of the algorithm (due to industrial property reasons) or the inability to understand, given its structure, how it produces certain results. In my opinion, even if it were possible to

[41] Gascón Inchausti, *Desafíos para el proceso penal en la era digital*, cit., 204.
[42] Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, 2017, 111.
[43] Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., 94.
[44] Ferrajoli, *Diritto e ragione. Teoria del garantismo penale*, 3° ed., Laterza, 1996, 19 ss.; Ferrajoli, *Las fuentes de legitimidad de la jurisdicción, Reforma Judicial: Revista Mexicana de Justicia*, n. 15-16, 2010, 5; Di Bitonto, *Neuroscienze e processo penale*, in *Prova scientifica e processo penale*, edited by Canzio & Luparia, Cedam, 2017, 746. Also see Cascone, *Emotional biometrics: a preliminary analysis*, cit., 304.
[45] Richard M. Re and Solow-Niederman, *Developing Artificially Intelligent Justice*, in *Stanford Technology Law Review*, 2019, UCLA School of Law, Public Law Research Paper no. 19-16, Available at SSRN: https://ssrn.com/abstract=3390854, 246.

understand how an artificial intelligence system generated a particular association, the highly technical nature of the scientific knowledge that underpins its functioning would still prevent judges, other legal practitioners, and citizens from truly understanding the meaning of its result.

Under this point of view, the problem of the judge's uncritical adherence to the machine's outcome and the risk of over-reliance on the machine result are nothing more than another form of "opacity"[46] that does not depend on the structural characteristics of the algorithm but rather on the knowledge and technical skills of those who must interpret its result. In this sense, it can be asserted that "technical transparency" of the algorithm does not solve the problem of its "opacity" and, therefore, the potential delegitimization that citizens may perceive when they are looking at a justice system administered through technically impenetrable solutions.

This perspective presents an additional – maybe the principal – regulatory challenge for lawmakers: it is not just about making the source code accessible; it is not just about making the technical functioning of the algorithm "transparent". It is also about training a class of legal practitioners who can interpret the machine's result, who have the technical and scientific knowledge to understand it, and, if necessary, to dissociate from it. This is the greatest challenge that European legislators face today.

In this regard, the solution proposed by some legal scholars regarding the need for "meaningful human control"[47] in decision-making processes based on the results provided by artificial intelligence algorithms does not seem to be conclusive for at least two reasons. Firstly, it does not solve the issue of the "technical" transparency of the algorithm: until it becomes possible to reconstruct how it processes data to produce certain results, all the critical aspects discussed in previous pages cannot be resolved, starting with the inability to ensure the effectiveness of the right to defense, which presupposes the possibility to challenge a judicial decision based on an automatically generated cognitive result. On this point, it is also worth noting that the solutions recently proposed by experts in the field of explainable AI (XAI), based on a concept of explainability rooted in post-hoc logic, do not appear satisfactory. In essence, post-hoc

[46] Similarly, Bonsignore Fouquet, *Sobre Inteligencia Artificial, decisiones judiciales y vacíos de argumentación*, cit., 264.

[47] Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Dir. Pen. Cont.*, 2020, 4. Similarly, Dinacci, *Intelligenza artificiale tra quantistica matematica e razionalismo critico: la necessaria tutela di approdi euristici*, in *Proc. pen. e giust*. 2022, 6, 1637; *Gialutz, Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Dir. pen. cont*., 29 maggio 2019, 22; Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia,* in *Discrimen*, 15 maggio 2020, 19.

explainability reproduces what happens in communication, providing explanations of algorithm workings that can differ from the actual processes of the algorithms[48]. These solutions do not seem adequate for the simple reason that they do not explain the concrete functioning of the algorithm in the specific case, thereby they do not allow parties in a trial to have real control over its operations.

Secondly, even if the algorithm were transparent, the problem of ensuring the judge's decision-making autonomy vis-à-vis the machine would remain, beyond mere verbal formulas[49]. It can be seriously doubted that this goal can be achieved solely through the introduction of procedural rules claiming for independent evaluation and explicit motivational burden. It is foreseeable that, when judges will be allowed to use AI, judicial motivations will be affected in most cases by AI results. Instead, it is necessary to prevent judge's conclusions to become a mere *ex post* justification, a decision based on the adherence to the outcome offered by AI. Because of this, it is essential, first and foremost, that in criminal proceedings judge and parties have adequate tools to completely understand the functioning of the algorithm used case by case.

4. The European Union's approach: the "opacity" of the legal solution suggested by the European Commission in the proposal for the adoption of the "AI Act". The problem of algorithmic transparency is specifically addressed in the proposal for the adoption of the "AI Act" presented in the spring of 2021 by the European Commission. In *considerandum* 38 of the version resulting from the European Parliament's general position of May 2023, it is explicitly stated that «the use of AI tools by law enforcement and judicial authorities should not become a factor of inequality, social fracture or exclusion. The impact of the use of AI tools on the defense rights of suspects should not be ignored, notably the difficulty in obtaining meaningful information on their functioning and the consequent difficulty in challenging their results in court, in particular by individuals under investigation». The European institutions are therefore fully aware of the risks associated with the use of non-transparent AI systems – meaning the inability to verify and "falsify" the acquired knowledge – for law enforcement purposes. *Considerandum* 47 of the proposal also states that «to address the opacity that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems».

[48] See Esposito, *Does Explainability Require Transparency?*, in *Sociologica*, 2022, 16(3), 23.
[49] See Lasagni, *Difendersi dall'Intelligenza artificiale o difendersi con l'intelligenza artificialle?*, cit., 1554.

This statement, unlike the previous one, is not specifically related to the use of AI for law enforcement purposes but applies to any high-risk artificial intelligence system. At the same way, the regulatory solution suggested in the proposal concerning the “opacity” of the algorithm does not specifically address the algorithms used for investigation and prosecution of crimes.

Article 13 of the proposal states that «High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured, with a view to achieving compliance with the relevant obligations of the user and of the provider set out in Chapter 3 of this Title».

To determine whether the provision proposes an adequate solution to the opacity of the algorithm in criminal proceedings, we need to take a step back. As mentioned in the previous paragraph, the issue of algorithmic transparency should be addressed (with suitable regulatory solutions) from two perspectives. Firstly, it is necessary to ensure the full intelligibility of the AI system’s operation to allow parties (and judges) to contest its results. Secondly, it is crucial to ensure that all users of the AI system can comprehend the results and rationalize them using familiar human categories. This is an essential prerequisite to prevent the decision based on an automatically generated result from becoming a mere human “ratification” of that result.

The solution proposed by the European Commission is, in a way, a middle ground between these two needs. As seen, it states that the AI tool’s functioning must be «sufficiently transparent to enable users to interpret the system’s output and use it appropriately». It is not easy to understand how to interpret the locution “sufficiently transparent”. At first glance, it can be excluded that the provision imposes a high level of technical transparency of the algorithm, as suggested by the adverb “sufficiently”. Based on this provision, it can be inferred that full intelligibility of the algorithm’s functioning is not a necessary requirement for its circulation in the internal market. Instead, the proposal seems to refer to a concept of transparency to be intended as the capacity of a human being to understand the result and assess its reasonableness using familiar human categories, even if the full technical transparency of the algorithm is not fulfilled. This conclusion is supported not only by the reference to the need to ensure that the user can “interpret the system’s output” but also by the following provisions on information obligations established in the second part of Article 13 of the proposal. It

states that «high-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct, and clear information that is relevant, accessible, and comprehensible to users». The information and data to be provided are further specified in paragraph 3 and include system's capabilities and performance limits, accuracy level, factors that may affect the algorithm's operation and the reliability of its results, and so on.

As mentioned before, this solution can be considered a middle ground, at least when applied to the field of criminal proceedings. On the one hand, it solves in a non-rigorous manner the problem of technical transparency, that does not seem to be an actual prerequisite for the AI's regular circulation in the internal market. On the other hand, it addresses the issue of the permanent evaluative autonomy of human beings concerning the machine's result on a general level, without any specific solution for criminal proceedings, and through an informative obligation that appears inadequate and insufficient. A judge required to apply pretrial detention to an individual for whom an artificial intelligence has calculated a high risk of recidivism is unlikely to deviate from the machine's result simply because an instruction manual informs him of what "could go wrong" in the tool's functioning[50]. It is evident that this level of preparation is ineffective and merely formal.

In other words, it is necessary to introduce regulatory solutions that pursue a dual objective: imposing a level of technical transparency sufficient to allow all procedural actors to challenge the result of the algorithm; on the other hand, reducing the risk of excessive reliance of the judge on the outcome offered by artificial intelligence, ensuring the autonomy of his evaluations even when supported by an AI system. It is evident that these needs require an intervention that specifically takes into account the levels of protection of rights and procedural guarantees recognized in the field of criminal procedure.

[50] It is worth, in this regard, quoting the words of Gascón Inchausti, *Desafíos para el proceso penal en la era digital*, cit., 204, according to whom «en términos generales, es comprensible la tendencia humana a tratar de delegar en un tercero las decisiones complejas o, al menos, todas o parte de las bases de esas decisiones —la sumisión pericial es manifestación de la anterior—. Esta tendencia se acentúa en tiempos como los actuales, en que la presión mediática y las críticas precipitadas —y no jurídicas— a las resoluciones judiciales minan la independencia judicial. Y si el tercero en quien se delega, en todo o en parte, la toma de la decisión no es otra persona —que podría tal vez rechazar la asunción de responsabilidad— sino una «máquina», las consecuencias pueden ser evidentes: las predicciones efectuadas por sistemas de inteligencia artificial se acabarán integrando en el proceso de toma de decisión en asuntos complejos, como lo son siempre los vinculados al riesgo de reiteración delictiva. Y no lo harán necesariamente de forma ecuánime o neutra, sino gozando de una cierta «apariencia de mejor condición», justamente por esa pretensión de objetivar lo que intrínsecamente no es objetivable».

Therefore, it can be said that the solution offered by the proposal for the AI Act establishes minimum requirements and guarantees that must be met for all high-risk uses of artificial intelligence. Another issue is how to implement these solutions in the specific field of criminal procedural law, where the interests to be protected are not only of the highest importance but also subject-specific. It will be up to national legislators to decide how to safeguard these interests and principles without sacrificing the benefits that can derive from the use of artificial intelligence in criminal procedure.