Freedom, Security & Justice: European Legal Studies

Rivista giuridica di classe A

2025, n. 2

EDITORIALE SCIENTIFICA

Rivista quadrimestrale on line "Freedom, Security & Justice: European Legal Studies" www.fsjeurostudies.eu
Editoriale Scientifica, Via San Biagio dei Librai, 39 - Napoli
CODICE ISSN 2532-2079 - Registrazione presso il Tribunale di Nocera Inferiore n° 3 del 3 marzo 2017

Indice-Sommario
2025, n. 2

Editoriale

Dalla dichiarazione Schuman al Libro bianco sulla prontezza alla difesa europea: verso una revisione del progetto europeo? p. 1
Ugo Villani

Saggi, Articoli, Commenti e Note

Le origini dello Spazio di libertà, sicurezza e giustizia. Pace e conflitti armati (1945-2025) p. 14
Massimo Panebianco

Migrare: un diritto fondamentale? p. 26
Antonio Ruggeri

Il ruolo della Procura europea (EPPO) nella tutela dello Stato di diritto dell'Unione europea p. 42
Serena Crespi

Norme di diritto internazionale e disparità di genere, idee vecchie e nuove. Il caso del *mundio muliebre*, uno stereotipo da rileggere p. 82
Lucia di Cintio

Convenzione delle Nazioni Unite contro il *cybercrime* e tutela dei diritti umani: influenze europee sullo scenario internazionale p. 108
Marco Dimetto

The error in predictive justice systems. Challenges for justice, freedom, and human-centrism under EU law p. 131
Alessandro Ferrara

EU impact on Albanian medical civil liability: a case law approach p. 146
Enkelejda Koka, Denard Veshi, Aisha Morina

La promozione della parità di genere nelle relazioni tra l'Unione europea e i *partner* meridionali p. 162
Claudia Morini

FOCUS

Democracy and the Rule of Law: A New Push for European Values

Il Focus contiene contributi elaborati a seguito della riflessione realizzata nel Seminario conclusivo dello Jean Monnet Module Eu-Draw (2022-2025) "Democracy and the Rule of Law: A New Push for European Values", tenutosi presso l'Università degli Studi di Salerno (1 aprile 2025)

Presentazione del *Focus* p. 192
Angela Di Stasi

Values in the EU external action: mechanisms of implementation and their outcomes p. 194
Stefania Kolarz

Justice and Home Affairs Cooperation (JHAC) in the perspective of enlargement p. 211
Teresa Russo

Brevi riflessioni sulla tutela dei diritti nello "spazio digitale" europeo p. 228
Francesco Buonomenna

Consiglio d'Europa e intelligenza artificiale: un primo tentativo di regolamentazione a tutela dei diritti umani, democrazia e Stato di diritto p. 242
Anna Iermano

La disinformazione *online* come "minaccia ibrida" alla democrazia nell'Unione europea: meccanismi di tutela e strumenti a contrasto per uno Spazio di libertà, sicurezza e giustizia p. 272
Angela Festa

L'"approccio europeo" al contrasto alla disinformazione digitale e alla protezione dei valori democratici: quale contributo dell'*AI Act?* p. 296
Rossana Palladino

THE ERROR IN PREDICTIVE JUSTICE SYSTEMS. CHALLENGES FOR JUSTICE, FREEDOM, AND HUMAN-CENTRISM UNDER EU LAW

Alessandro Ferrara*

SUMMARY: 1. Introduction. – 2. The functioning of predictive justice systems. – 3. The ethical implications of judicial correctness and the claim to justice. – 4. Predictive justice behind the Rawlsian veil of ignorance. – 5. Concerns for human-centrism and freedom. – 6. Conclusions.

1. Introduction

Artificial intelligence has increasingly become one of the central themes of academic discourse over the last decade, not only due to its pervasive material presence across all sectors of contemporary society, but also because of its conceptual impact across disciplines. From law and computer science to philosophy, engineering, mathematics, business administration, and medicine, AI has emerged as a paradigmatic "cross-cutting issue" that compels scholars to reimagine the boundaries of human agency, decision-making, and responsibility[1].

This paper focuses on one of the most delicate and controversial applications of artificial intelligence: its use in the legal field, particularly in supporting – or potentially replacing – judicial decision-making. In recent years, the rapid implementation of digital technologies in justice systems, especially within the European Union, has opened the door to reflections on the legitimacy and desirability of "robot-judges": algorithmic systems capable of autonomously producing legal rulings based on data analysis and probabilistic models[2]. In fact, in order to align with the imperative of safeguarding every individual's right to a fair trial, as enshrined in Article 6(1) of the European Convention on Human Rights (ECHR)[3], many EU Member States have prioritized the digital

Double-blind peer reviewed article.

* Research fellow at the Magna Graecia University of Catanzaro. E-mail: a.ferrara@unicz.it.

[1] L. FLORIDI, *The Ethics of Artificial Intelligence*, Oxford, 2022, p. 15.

[2] T. SOURDIN, *Judge v Robot? Artificial Intelligence and Judicial Decision-Making*, in *University of New South Wales Law Journal*, 2018, no. 4, pp. 1114-1133.

[3] European Convention on Human Rights, Art. 6 ECHR, para. 1. "Everyone is entitled to a fair and public hearing within a reasonable time, before an independent and impartial tribunal constituted by law, for the purpose of determining both his civil rights and obligations and the merits of any criminal charge against him [...]". Available at: https://www.echr.coe.int/documents/convention_ita.pdf (last accessed 11 March 2023).

transformation of their judicial systems. This trend has been accompanied by consistent efforts on the part of the European Commission to evaluate and enhance the quality, independence, and efficiency of national justice systems. Since 2013, the Commission's annual EU Justice Scoreboard has provided comparative data and has functioned as a monitoring tool aimed at standardising and improving judicial performance across Member States[4].

The Covid-19 pandemic has exacerbated pre-existing inefficiencies, resulting in increased case backlog and longer durations of proceedings, as confirmed by the *2022 Rule of Law Report*[5]. In response, several national recovery and resilience plans (NRRPs), such as Italy's, have incorporated ambitious objectives: Italy, for instance, has pledged to reduce the average duration of proceedings in all three levels of jurisdiction by at least 40% by 2026[6]. Within this framework, digital innovation – AI included – is framed as a "broad-based cross-media enabler" for systemic reform[7].

In August 2024, the Artificial Intelligence Act (AI Act) entered into force, introducing a comprehensive regulatory framework for the development and deployment of AI systems within the European Union. Among its foundational provisions, both the Preamble and Article 1 articulate the overarching objective of the Regulation: namely "to promote the uptake of human centric and trustworthy artificial intelligence (AI) while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union"[8].

The Regulation further delineates a distinct set of stricter principles for so-called high-risk AI systems, a category that explicitly includes applications used in the administration of justice. In this regard, Recital 61 of the Preamble stipulates that "Certain AI systems intended for the administration of justice and democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, the rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial. In particular, to address the risks of potential biases, errors and opacity, it is

[4]Communication from the Commission to the European Parliament, the Council, the European Central Bank, the European Economic and Social Committee and the Committee of the Regions, *The EU Justice Scoreboard. A Tool to Promote Effective Justice and Growth*, of 27 March 2013, COM(2013) 160 final, p. 3. Available at: https://commission.europa.eu/system/files/2017-06/justice_scoreboard_2013_en.pdf (last accessed 12 March 2023).

[5] European Commission, *Rule of Law Report 2022. The Situation of the Rule of Law in the European Union*, 2022, p. 12. See the Chapter on the Situation of the Rule of Law in Italy. "The temporary slowdown in judicial activity due to the severe restrictive measures adopted to deal with the COVID-19 pandemic in 2020 has had an impact on both incoming cases and exhausted proceedings, significantly affecting the time to exhaustion", p. 12.

[6] Government of Italy, *National Recovery and Resilience Plan (PNRR). The Reform of Justice*, pp. 55-67. See also: *PNRR Measures Progress Note*, p. 8.

[7] Ibid., p.54.

[8] Regulation (EU) 2024/1689 of the European Parliament and of the Council, *laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828, Artificial Intelligence Act*, of 13 June 2024, art.1, in OJ L, 2024/1689, 12 July 2024, pp. 1-144, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689 (last accessed on 20/05/2025).

appropriate to qualify as high-risk AI systems intended to be used by a judicial authority or on its behalf to assist judicial authorities in researching and interpreting facts and the law and in applying the law to a concrete set of facts. AI systems intended to be used by alternative dispute resolution bodies for those purposes should also be considered to be high-risk when the outcomes of the alternative dispute resolution proceedings produce legal effects for the parties. The use of AI tools can support the decision-making power of judges or judicial independence but should not replace it: the final decision-making must remain a human-driven activity[9]".

While these goals are commendable, the growing use of artificial intelligence raises important philosophical and ethical questions. Hacker and Petkova contend that the legitimacy of high-risk AI applications depends on institutional structures that prioritise human-centred normative reasoning and prevent structural imbalances in digital governance[10]. Indeed, some scholars[11] have suggested that the digitisation of justice may reach a point where the human judge could be fully replaced by predictive algorithms in specific categories of cases. These systems are capable of issuing decisions based on precedent analysis, data mining and machine learning. However, such a shift would fundamentally alter the logic of adjudication, introducing a statistical logic that tolerates a certain—and structurally unavoidable—margin of error. At the same time, even in the more recent scenarios where artificial intelligence systems are designed to support rather than replace human decision-making, concerns remain regarding their potential to undermine human autonomy and freedom. Such dynamics may significantly affect the essential requirement of human-centrism, by limiting the decisional independence of human agents and subtly shifting control towards algorithmic processes.

2. The functioning of predictive justice systems

Artificial Intelligence is usually defined as "the set of scientific methods, theories and techniques aimed at reproducing the cognitive capacities of human beings through machines" [12] that is, machines "capable of performing tasks requiring intelligence when performed by humans"[13]. In 2019, the High-Level Expert Group on AI (AI HLEG), established by the European Commission, offered a broader definition: "Artificial intelligence (AI) systems are systems designed by human beings who, given a complex objective, act within a physical or digital dimension, perceiving their environment

[9] Ibid. preamble para. 61.
[10] P. HACKER, B. PETKOVA, *Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers*, in *Northwestern Journal of Technology and Intellectual Property*, 2017, no. 1, p. 15.
[11] A. LODDER, B. VERHEIJ, *Legal Knowledge and Information Systems: JURIX 2020*, Amsterdam, 2020, pp. 33-45.
[12] European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Justice Systems and Related Areas*, adopted at the 31st Plenary Meeting, Strasbourg, 3-4 December 2018, Glossary.
[13] See R. KURZWEIL, *The Age of Intelligent Machine*, Cambridge, 1990, p. 14; S.J. RUSSELL**,** P. NORVIG, *Artificial Intelligence: A Modern Approach*, Upper Saddle River, 2016, sect. 1.1.

through the acquisition of data, interpreting the collected data, both structured and unstructured, processing the information derived from this data and selecting the best possible actions to achieve (consistent with predefined parameters) the indicated objective. Artificial intelligence systems can be created to learn to adapt their behaviour by analysing how the environment is affected by their past actions[14]."

To understand how AI systems are applied in the legal field – particularly in the context of predictive justice – it is essential to distinguish between two general categories: knowledge-based systems[15], which operate according to explicit logical rules and symbolic representations, and machine learning systems[16], which learn inductively from data.

It is this second category – machine learning systems[17], and more specifically deep learning – that underpins most of today's predictive justice models. These systems are designed to analyse large volumes of data and, through training processes, learn how to generate outputs that simulate or support human decision-making[18]. The basic structure involves algorithms that improve their performance over time, adapting to the patterns detected in the training datasets. As the amount of data increases, so too does the accuracy and complexity of the model's output[19].

Machine learning models can be broadly classified into three main types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning, in particular, is crucial to the development of predictive justice, and it often incorporates deep neural networks – layered architectures inspired by the structure of the human brain, although far simpler. These networks are composed of input, hidden, and output layers, through which data is processed in stages, refining its meaning at each step. During

[14] High Level Expert Group on Artificial Intelligence (AI HLEG), *A Definition of AI: Main Capabilities and Scientific Disciplines*, 2019, available at: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf (last accessed 12/03/2023).

[15] See, in general, R. BUEST, *Artificial Intelligence Is About Machine Reasoning—or When Machine Learning Is Just a Fancy Plugin*, *CIO*, available at: https://www.cio.com/article/3236030/artificial-intelligence-is-about-machine-reasoning-or-whenmachine-learning-is-just-a-fancy-plugin.html (last accessed 12/03/2023) (explaining the progress of AI and the ability of machines to learn ways of reasoning).

[16] The expression was coined by Arthur L. Samuel in 1959 and later formalised by Tom M. Mitchell to refer to a technology that can teach computer programmes to imitate human cognitive processes including adaptation and experience. In addition to Machine Learning, Artificial Intelligence also includes Speech Recognition, Natural Language Processing and Image Recognition, see E. CALZOLAIO, *Artificial Intelligence and Decision Autonomy: Problems and Challenges*, in E. CALZOLAIO (ed.), *La decisione nel prisma dell'Intelligenza Artificiale*, Milano, 2020.

[17] Ibid., it should be made clear that the concept of learning is not superimposable on the human concept of learning and therefore related to the growth of cognitive capacities but is to be understood in a functional sense as becoming more and more efficient in a given activity.

[18] See also MathWorks, *What Is Machine Learning? 3 Things You Need to Know*, available at: https://www.mathworks.com/discovery/machine-learning.html (last accessed 13/03/2023).

[19] A recurring example is the filtering of *spam* mails. Certain e-mails received in mailboxes are automatically sorted into the *spam* folder by an AI system. This type of system works by analysing the text of a huge number of mails that have already been categorised as *spam* and is able to determine which lexical recurrences are statistically (with a percentage close to 100 per cent) present in the mails categorised as such, and as a result is able to place them in the correct folder. The more data and information the AI analyses, the more the system *learns to* make correct decisions.

training, the model adjusts its internal parameters (weights and biases) in order to minimize the error between predicted and actual outcomes[20].

In the legal domain, such models are trained on vast repositories of judicial decisions. By identifying correlations between certain phrases, legal arguments, fact patterns, and judicial outcomes, the system can generate predictions about the likely decision in a new case. For example, given the factual context of a dispute and references to applicable norms, the algorithm may return a percentage indicating the probability of success of a legal action[21].

However, it is precisely this statistical and correlational nature that reveals the first major limitation of AI in justice. Unlike human reasoning, which includes interpretation, moral evaluation, analogical reasoning, and creativity, the logic of machine learning is purely inductive and non-reflective. The algorithm identifies correlations, but does not understand causation, nor does it possess any concept of fairness, equity, or social context[22].

Even more critically, the outputs of machine learning models always carry a mathematically ineliminable margin of error. This is a direct consequence of the probabilistic nature of the underlying statistical inference. The law of large numbers – a foundational concept in statistics – tells us that, as the number of observations grows, the average of the results obtained from a large number of trials should converge to the expected value. However, this convergence is asymptotic and never absolute: no number of repetitions can guarantee that an individual prediction will be accurate[23].

In this regard, it is important to recall two key definitions of probability used in statistical theory. The empirical definition, elaborated by Richard von Mises, defines probability as the relative frequency of an event in a large series of trials. Specifically, the probability of an event E is approximated by the ratio f(E) = n/N, where *n* is the number of times E occurs, and *N* is the total number of trials. As N increases indefinitely, f(E) tends toward a stable limit. Yet, this limit is ideal and abstract; in real-world applications, we always operate with finite datasets, and thus with residual uncertainty[24].

The classical definition, proposed by Pierre-Simon Laplace, frames probability as the ratio of favourable outcomes to total equally possible outcomes – assuming all are equiprobable. This conceptualization yields a value between 0 and 1, where 0 denotes impossibility and 1 certainty. But even here, the assumption of equal probability rarely holds in complex social systems, especially in the legal domain[25].

Hence, both definitions converge on a fundamental insight: probability is never equivalent to certainty. This has profound implications for AI-based justice. No matter

[20] I. GOODFELLOW**,** Y. BENGIO, A. COURVILLE, *Deep Learning*, Cambridge, 2016, pp. 103-127.
[21] N. ALETRAS, D. TSARAPATSANIS, D. PREOȚIUC-PIETRO, and V. LAMPOS, *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, in *PeerJ Computer Science*, 2016, 2: e93, pp. 1-19, https://peerj.com/articles/cs-93/ (last accessed 5/12/2022).
[22] H. SURDEN, *Machine Learning and Law*, in *Washington Law Review*, 2014, 89, no. 1, pp. 87-115.
[23] B. DE FINETTI, *Theory of Probability*, New York, 1974, vol.1, pp. 38-41.
[24] R. VON MISES, *Probability, Statistics and Truth*, Mineola, New York, 1981, pp. 11-15.
[25] P.-S. LAPLACE, *Théorie analytique des probabilités*, Paris, 1812.

how sophisticated or extensive the dataset, machine learning models cannot eliminate the risk of incorrect predictions – only reduce it to a statistically acceptable margin. But what is "acceptable" when fundamental rights and legal outcomes are at stake?

Moreover, even the logic of convergence does not imply individual-level guarantees. To quote a simple but illustrative example: there is no finite number M such that, after M rolls of a die, one can be sure of obtaining a six. As M increases, the probability approaches 1 – but the possibility of failure remains[26].

In the legal realm, this means that every AI-generated decision must be understood as hypothetically fallible, even if the probability of error is low. And yet, unlike human judges – who are accountable, must justify their decisions, and can be appealed – AI systems lack transparency and explainability. Their internal processes often remain opaque even to their designers, a phenomenon known as the "black box" problem[27].

Furthermore, AI systems do not create law – they only reproduce existing patterns. Their outputs are conservative, tending to replicate the majority trend in precedent rather than challenge it. In this sense, they risk reinforcing structural biases and inequalities that exist within the data on which they are trained. As such, predictive justice systems are not neutral tools but deeply embedded in the legal culture and jurisprudential tendencies of their training corpus[28].

As Sartor[29] observes, predictive systems do not operate in a vacuum: their training datasets encode specific legal traditions, structural inequalities, and latent biases that may undermine fairness and neutrality in judicial reasoning.

This observation also raises the question of institutional responsibility: if an AI system wrongly predicts a case outcome, who is to blame? The developer? The court that relies on the system? The state that endorses its use? These questions remain largely unresolved and highlight the ethical fragility of automated legal decision-making.

In light of all this, it becomes clear that the statistical sophistication of AI models does not translate into legal or ethical adequacy. The law is not an algorithmic function, and justice is not a matter of probability. The ethical foundation of legal judgment requires more than predictive accuracy; it demands deliberation, moral reasoning, empathy, and public justification – all of which are currently beyond the reach of AI.

Ultimately, the integration of predictive systems into legal practice must be approached with great caution. To replace or even heavily influence judicial decisions with tools that operate on the basis of inherent and opaque margins of error is to institutionalize injustice under the guise of efficiency. The legitimacy of the judiciary rests not only on its capacity to decide quickly or uniformly, but on its capacity to decide justly.

---

[26] A.N. KOLMOGOROV, *Foundations of the Theory of Probability*, New York, 1950, pp. 56-58.
[27] J. BURRELL, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, in *Big Data & Society*, 2016, no. 3, pp. 1-12.
[28] S. BAROCAS, A. D. SELBST, *Big Data's Disparate Impact*, in *California Law Review*, 2016, 104, no. 3, pp. 671-732.
[29] G. SARTOR, *Artificial intelligence and human rights: Between law and ethics*, in *Maastricht Journal of European and Comparative Law*, 2020, 27, no. 6, pp. 705-719.

3. The ethical implications of judicial correctness and the claim to justice

The necessary presence of a certain percentage of error in the use of artificial intelligence technology and systems is generally accepted by the scientific community as well as by modern society. In most, if not all, cases, this percentage of error is accepted because the use of technology or AI results in a qualitative improvement in the outcome of the activity previously performed by the human being.

In the field of law and in particular in the field of judicial activity, the reasoning leading to the ethical acceptance of the use of artificial intelligence systems is probably more complex, since, from a conceptual point of view, the comparison between machine-generated and human-generated output is likely to lead to different conclusions.

The theoretical starting point for the reflections outlined above can be found in a well-established current of Anglo-Saxon and Oxonian jurisprudence, according to which law necessarily raises a claim to validity (*Geltungsanspruch*), and every legal norm is inherently tied to a prescriptive force (*Sollgeltung*) as well as a performative assertion of its normative authority (*Anspruch auf Geltung*) [30] i.e. according to which judgement does not only imply the validity of legal norms but also somehow implies their justice[31]. This tradition emphasizes that the very act of legal judgment cannot be reduced to a mere mechanical application of rules, but implies a commitment to correctness, or at the very least, a claim that the judgment is justified both normatively and ethically.

This perspective invites a crucial ethical question: can one accept, within a system of justice, that a certain proportion of legal decisions – however small – may be known in advance to be wrong, simply because they result from a fallible machine process? Is it morally permissible to institutionalize this margin of error in a domain where every individual case implicates rights, freedoms, and dignity?

To explore this dilemma, we may draw upon the non-positivist theory of law developed by Robert Alexy, who famously argued that *"law necessarily also raises a claim to correctness"* (*Anspruch auf Richtigkeit*)[32]. This claim does not imply that every legal norm is in fact morally right, but that every legal act – especially a judicial decision – must be formulated as if it were correct, and this very form creates a conceptual bridge between law and morality[33].

According to Alexy, this connection manifests as a performative structure inherent in legal discourse: whenever a court pronounces a judgment, it not only states what the law

[30] J. HABERMAS, *Moral Consciousness and Communicative Action*, Cambridge, 1990, p. 44; see also H. L. A. HART, *The Concept of Law*, Oxford, 1961, pp. 253-258.
[31] J. DERRIDA, *Force of Law: The "Mystical Foundations of Authority*", in D. CORNELL**,** M. ROSENFELD, D.G. CARLSON (eds,), *Deconstruction and the Possibility of Justice*, New York, 1992, pp. 3-67.
[32] L. SARDO, *The Claim of Fairness and the Concept of Normative Necessity*, in *Rivista di Filosofia del Diritto*, 2020, IX(2), p. 344; R. ALEXY, *The Argument from Injustice: A Reply to Legal Positivism*, Oxford, 2002, p. 34.
[33] R. ALEXY, *op.cit.*, 2002, p. 23; see also R. ALEXY, *The Dual Nature of Law*, in *Ratio Juris*, 2010, no. 23, pp. 167-182; R. ALEXY, *Between Positivism and Non-Positivism? A Third Reply to Eugenio Bulygin*, in J. FERRER BELTRÁN**,** J.J. MORESO, D. PAPAYANNIS (eds.), *Neutrality and Theory of Law*, Dordrecht, 2013, pp. 225-238.

is but implicitly asserts that the decision is correct, justifiable, and right. This becomes particularly clear in Alexy's well-known thought experiment: imagine a judge writing in a sentence, *"Titius is sentenced to life imprisonment, which is unjust."* Such an utterance would constitute a performative contradiction (*performativer Widerspruch*), because the very act of judicial pronouncement presupposes a claim to justice. Declaring an unjust sentence would void the act of its legal meaning[34].

This thought experiment reveals what Alexy calls a "necessary truth about the nature of law": namely, that legal decisions are conceptually bound to an aspiration to justice. Law, in this framework, cannot be normatively indifferent; justice is not an accidental quality of law, but part of its conceptual structure[35].

Critics have contested Alexy's theory on various grounds. Some argue that empirical evidence shows that unjust judgments are frequently issued without apparent contradiction. Others challenge the universality of his performative model, especially in legal systems where discretion and compromise play a larger role[36]. Nonetheless, Alexy's claim to correctness has gained widespread recognition as a powerful tool for understanding the normative force of judicial reasoning, even among scholars who reject his broader legal philosophy[37].

Our purpose here is not to adopt Alexy's entire framework, but rather to highlight the specific implication that judicial decisions, in order to be judicial, must at least claim to be just. This is especially relevant when evaluating the ethical viability of predictive justice systems, since such systems inevitably involve outputs that, by statistical construction, include a predictable and non-zero margin of error.

If we accept Alexy's theory, it becomes clear that a decision cannot be simultaneously judicial and unjust, unless it is later absorbed and superseded by a corrective judgment. In this logic, the injustice of a decision only exists retroactively, when a new, overriding decision declares the previous one to have been unjust. In other words, injustice in the judicial domain is not a freestanding ontological state, but a relational outcome, emerging only in comparison to a superior legal pronouncement[38].

This line of reasoning raises a serious concern about the integration of AI into adjudication. If we know in advance that a given system of predictive justice will generate some percentage of decisions that are statistically guaranteed to be incorrect, then we must conclude that it will, by design, produce judicial decisions that violate their own claim to correctness. This violates not only Alexy's principle but also the ethical core of legal rationality.

One might object that human judges also make mistakes, and that AI systems could simply be another imperfect tool. But this analogy misses the mark. Human judicial error is epistemic, discovered through appeals and critique; it is not ontologically embedded in

[34] R. ALEXY, *op. cit.*, 2002, p. 38; see also R. ALEXY, *A Theory of Legal Argumentation*, Oxford, 2009, p. 41.
[35] R. ALEXY, *On the Concept and the Nature of Law*, in *Ratio Juris*, 2008, 21(3), pp. 281-299.
[36] H. KELSEN, *Pure Theory of Law*, Berkeley, 1967, pp. 305-307.
[37] J. HABERMAS, *Between Facts and Norms*, Cambridge, 1996, pp. 199-203.
[38] A. PECZENIK, *The Justification of the Judicial Decision*, in *Ratio Juris*, 1988, 1(1), pp. 43-66.

the system itself. In contrast, AI-based decisions rest on algorithmic structures that cannot be fully justified or explained. They may simulate justification, but do not offer actual normative reasoning[39].

Moreover, AI lacks the reflexive capacity to revise its reasoning in light of broader moral principles. It cannot engage in interpretive debate, weigh competing values, or offer public justification – all of which are essential components of legitimate legal reasoning. Thus, assigning binding legal force to its outputs entails a categorical shift in the structure of judicial authority.

In conclusion, if we accept that every legal decision entails an intrinsic and performative claim to justice, as argued by Alexy, then it follows that no system can ethically produce judicial decisions that are known in advance to be statistically flawed. Doing so would amount to institutionalizing error and thus undermining the very meaning of law as a normative enterprise. The question is not whether predictive systems are useful or efficient, but whether their outputs can sustain the ethical burden of judgment. If they cannot, their integration into the judicial process must be carefully delimited – and perhaps even excluded from certain domains where correctness is indispensable.

4. Predictive justice behind the Rawlsian veil of ignorance

In light of what has been seen in the previous paragraphs, it can be assumed that any artificial intelligence system built to produce a jurisdictional *output* (be it a decision, an outcome probability or the identification of a prevailing orientation) carries with it a certain percentage of error.

From the systems that exist and have been tested to date, it has been possible to verify that, even in optimal cases, the accuracy achieved by algorithms capable of predicting the outcome of a case does not reach satisfactory percentages[40]. A well-known example is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) software, used in the United States to assess the likelihood of recidivism. In the landmark case State v. Loomis (2016), the Supreme Court of Wisconsin acknowledged that the sentencing of the defendant was based in part on COMPAS scores, despite the fact that the software's inner workings were not fully transparent and could not be independently verified. The decision raised serious ethical concerns about the use of opaque, potentially biased algorithms in judicial decision-making[41]. In any case, this argument is not strictly

[39] T. BENCH-CAPON**,** K. ATKINSON**,** H. PRAKKEN, G. SARTOR, *Using Argumentation to Explain and Justify Outcomes in AI & Law*, in *Artificial Intelligence and Law*, 2018, 26, n.4, pp. 403-423.

[40] This is for instance the case of an algorithm created by the Univeristy College of London on the basis of the case law (584 decisions) of the European Court of Human Rights in order to predict its decisions, which tried to predict the outcome with 79% accuracy. See N. ALETRAS ET AL., *Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective*, *PeerJ Computer Science*, 2016, 93, no. 2, pp. 1-19, https://peerj.com/articles/cs-93/ (last accessed 5/12/2024).

[41] Supreme Court of Wisconsin, application n. 881 N.W.2d 749, *State v. Loomis*, 2016; for an in-depth study see J. ANGWIN ET AL., *Machine Bias*, in *ProPublica*, 2016, pp. 1-30; D. KEHL**,** P. GUO, S. KESSLER, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, in *Harvard*

related to the future possibilities of developing the accuracy of the outcome but is based on the mere existence of a certain (or indeterminate) percentage of error, no matter how small it is.

In fact, in contrast to what emerges with regard to decisions produced by AI systems, if Robert Alexy's theory of the claim to fairness is taken as true, it can be said that human decisions, at least *ex ante* and at least until they are superseded and absorbed by a subsequent decision (in the context of an appeal, revision or revocation procedure) carry with them, in their entirety, a claim to fairness and, therefore, to justice.

At this point, the question arises as to whether in order to curb the time taken by justice, in accordance with the provision of Article 6 ECHR, it is acceptable to produce a certain percentage of erroneous *outputs*. It is necessary to establish whether Montesquieu was right in stating that justice delayed is justice denied and, above all, what the price of such celerity should be, or whether the mere fact that one speaks of justice does not ethically allow one to abandon the ultimate goal of *iustum*.

From a utilitarian perspective, one might argue that tolerating a limited amount of error is ethically acceptable if it maximizes overall utility and system efficiency. However, this reasoning is precisely what Rawls attempts to overcome by prioritizing the protection of the most disadvantaged members of society, even in probabilistic decision-making systems[42].

In order to test whether it is fair to accept that an AI can produce an *output* that with mathematical certainty determines, in a certain percentage of cases, an incorrect result, we will try to place ourselves behind the Rawlsian veil of ignorance.

According to John Rawls, justice can be identified as all those rules that are necessary for a plurality of people – despite the fact that they are potentially divided by opposing values, goals and personal conceptions – to coexist and cooperate within a society[43] . According to Rawls, therefore, the principles and norms that structure the concept of justice do not derive from the identification of fundamental axioms that ground the correctness of the principles themselves. On the contrary, it is the agreement on certain axioms or principles that makes them universally so[44] . According to Rawls, it is not possible to resolve the issue on the basis of assumptions or axioms based on moral reasons, which in relation to their subjectivity would not see all participants in the

Law School, 2017; E. BINNS, M. VEALE ET AL., *It's Reducing a Human Being to a Percentage: Perceptions of Justice in Algorithmic Decisions*, in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1-14.

[42] J. RAWLS, *A Theory of Justice*, Cambridge, 1971, pp. 75-83.

[43] Ibid., pp. 3-6.

[44] To explain the concept, Brian Bix uses the example of a previously agreed price that is subsequently considered excessive in relation to the value of the good or service purchased. Bix explains, in order to understand Rawls' idea of justice, it is sufficient to imagine a transaction between two persons involving the sale of a good and the payment of a relative price. If the asking price was considered, at the time of the transfer of the good, to be excessively high, the buyer would have a just reason for not concluding the transaction. If, on the other hand, the price had been previously agreed upon, the purchaser would not have a just reason to withhold payment of the price. This example is subject to certain limitations, already pointed out by Bix such as the mental capacity of the subject at the time of consent, the possibility of threats, fraud and so on, B. H. BIX, *Theory of Law, Ideas and Contexts*, in A. PORCIELLO (it. eds.), Torino, 2016, p. 142.

discussion agree on the starting points. On the other hand, it is possible to attempt to come to an agreement through an open discussion process without starting points, so that everyone, once they have come to the identification of the principles they consider just, can only agree with them. The critical point of this reasoning is that individuals do not always come to an agreement on the principles on which to agree as these will often be influenced by their own self-interests and as individuals tend to prefer principles and axioms that generate benefits for their specific position. It is from this assessment that the theory arises that individuals, when discussing the right principles to govern society, should find themselves behind a 'veil of ignorance' that makes them unaware of their position in society, their personal status and their outlook on life[45] . This ultimate situation is called the 'original position'[46] and would be the starting condition capable of rendering legitimate the initial identification of fundamental principles of justice within society. It is precisely because of this reasoning that Rawls defines his theory in terms of "justice as fairness" in the sense of justice as the result of an agreement between persons subject to fair conditions[47]. Rawls' theory of justice as represented in his best-known work, a reference point for modern philosophy, was subject to revision by the same philosopher in a subsequent work[48], following which the concept of justice was again theorised as the set of institutions and practices that allow individuals to coexist and prosper within society, despite their differences regarding the fundamental principles on which it should be founded and above all despite the moral divergences connected to how to conceive the good life. Justice thus becomes connected to a "political" and not a metaphysical idea that clarifies how people with completely different "theories of the good" and opposing ethical convictions manage to coexist within the same social system[49]. In essence, Rawls argues that justice is represented by those fundamental principles that form nothing more than the common substratum of each individual's different theories of the good and as such can be shared.[50]

Well then, taking Rawls' theory as a starting point, it will be possible to ask whether, placed in an original position, in the dark of their own condition, citizens would consider it corresponding to a principle of justice to rely on an AI system for the outcome of a dispute.

Assuming that we are in the worst condition, i.e. that of the subject against whom the erroneous output generated by the AI could be produced, it is possible to assume that the members of the public would probably decide to entrust the judicial activity to the human

[45] J. RAWLS, op. cit.,1971, p. 31.
[46] See S. FREEMAN, *Original Position*, in E.N. ZALTA (ed.), *Stanford Encyclopedia of Philosophy*, Substantive Revision Oct. 24, 2023, p. 1 ff., https://plato.stanford.edu/entries/original-position/ (last accessed 20/03/2025).
[47] See also J. RAWLS, *Justice as Fairness*, in *Journal of Philosophy*, 54, 1957, extended in *Philosophical Review*,1958, 67, p. 154.
[48] J. RAWLS, *Justice as Fairness: Political not Metaphysical*, in *Philosophy & Public Affairs*, 1985, 14, p. 223.
[49] J. RAWLS, *Political Liberalism*, New York, 1993, pp. 4-22.
[50] Ibid., pp. 9-11 and 58-66.

being as it meets an expectation of correctness in all cases, excluding the risk of being the recipients of an erroneous *output*.

Moreover, if one considers the judge as a moral agent endowed with professional discretion and deliberative reasoning, then replacing or subordinating such agency to automated reasoning would pose a challenge to the very nature of justice as a moral and dialogical construct[51].

Of course, such a conclusion is only possible if one agrees with the conclusions proposed by Alexy regarding the claim of correctness of judicial decisions.

There is no doubt, in fact, that judicial errors exist and that the judiciary produces a certain percentage of pronouncements that turn out, subsequently, to be erroneous. This error, however, is always 'unmasked' and crystallised in a subsequent decision or pronouncement that judicially affirms its erroneousness, in some way 'healing' the system's assumption of justice. Otherwise, the correctness of any pronouncement or decision could be debated, but from a conceptual point of view it would still remain just only because it comes from the body with the task of establishing what is just and what is not.

5. Concerns for human-centrism and freedom

The arguments just brought forward can be considered useful to start further reasoning on what could be the consequences of using AI in such a way as to replace - albeit with time - human activity in the judicial sphere.

Certainly, the concerns highlighted are more easily shared when one questions the ethical and legal appropriateness of adopting 'strong' predictive justice systems, i.e. those artificial intelligence systems that aim to completely, or almost completely, replace human activity. This type of perplexity - along with the many others raised in recent months concerning the *bias* algorithms, the influence of those who design them, the knowability and transparency of the processes that produce the *outputs*, and the intelligibility of their operation[52] – probably also deserves to be taken into account with regard to 'weak' predictive justice systems; Artificial intelligence systems subject to human control that can be used to assist the judge, but also lawyers, citizens or in any case all stakeholders of the justice system, in identifying prevailing case law precedents, possible outcomes of a judgement or in carrying out alternative dispute resolution procedures.

The issue had already been announced on the basis of the principles stated by the High-Level Expert Group on Artificial Intelligence in the *Ethical Guidelines for Reliable*

[51] L.L. FULLER, *The Morality of Law*, New Haven, 1969, pp. 122-127.
[52] See in order to identify some of the essential principles and recommendations for the proper use of AI, L. FLORIDI ET AL., *AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and Machines*, 2018, 28, pp. 689-707.

Artificial Intelligence of 9 April 2019[53]. It was definitively raised on the basis of the principles enshrined in paragraph 61 of the preamble to the AI Act and in article 1 below.

Therefore, artificial intelligence must be designed and used in a way that respects, protects and promotes human dignity and fundamental rights, the most essential of which is the guarantee of human-centrism when humans retain their autonomy and freedom in the use of AI.

As Hildebrandt[54] argues, preserving human interpretive agency within algorithmic infrastructures is a precondition for ensuring that legal decision-making remains grounded in normative reasoning rather than technical calculation.

In other words, machines must not replace humans in ethically relevant decision-making roles, nor must they threaten human autonomy or responsibility.

The Council of Europe, in its *Recommendation to Member States on the human rights impacts of algorithmic systems*[55], and the European Commission, in its *White Paper on Artificial Intelligence*[56], have essentially stated that AI must be 'human-centric', ensuring that humans have complete control over the outputs produced by artificial intelligence, and the competence and freedom to validate and deviate from them respectively. This requires transparency, traceability, human supervision, and the prevention of discriminatory or dignity-damaging risks.

In fact, some recent studies[57] have demonstrated a tendency in human beings, known as 'algorithmic appreciation', to prefer the advice, guidance and, more generally, the outputs produced by AI systems to those of other human beings or derived from their own convictions and skills. This tendency has also been demonstrated in medical diagnosis, where the percentage of correct diagnoses increased when the results of AI systems were excluded, as doctors, both young and experienced, found it more difficult to deviate from them, giving rise to over-appreciation[58].

However, the problem highlighted by these studies seems to contradict one of the cornerstones of the AI Act and all regulation in the field of artificial intelligence. The influence that AI systems risk having on self-awareness, trust, and the ability to generate

---

[53] European Commission: Directorate-General for Communications Networks, Content and Technology and Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, *Ethics guidelines for trustworthy AI*, Publications Office, 2019, https://data.europa.eu/doi/10.2759/346720 (last accessed on 20/05/2025).

[54] M. HILDEBRANDT, *Law for Computer Scientists and Other Folk*, Oxford, 2020.

[55] Council of Europe, *Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*, of 8 April 2020, CM/Rec(2020)1 https://search.coe.int/cm#{%22CoEIdentifier%22:[%2209000016809e1154%22],%22sort%22:[%22CoEValidationDate%20Descending%22]} (last accessed on 20/05/2025).

[56] European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust*, Brussels, of 19 February 2020, COM(2020), https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (last accessed on 20/05/2025).

[57] J.M. LOGG, J.A. MINSON, D.A. MOORE, *Algorithm appreciation: People prefer algorithmic to human judgment*, in *Organizational Behaviour and Human Decision Processes*, 151, 2019, pp. 90-103; S.A. JESSUP, G.M. ALARCON, S.M. WILLIS, M.A. LEE, *A closer look at how experience, task domain, and self-confidence influence reliance towards algorithms*, in *Applied ergonomics*, 121, 2024, pp. 104363.

[58] E. JUSSUPOW, K. SPOHRER, A. HEINZL, J. GAWLITZA, *Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence*, in *Information Systems Research*, 2021, 32, no. 3, pp. 713-735.

autonomous outputs considered reliable is essential to guaranteeing both human-centrism and fundamental human rights founded on freely exercised human knowledge and experience.

6. Conclusions

What emerges clearly is that the progressive integration of AI into the judicial system and in the administration of justice requires a balance between efficiency and ethical legitimacy. In other words, improving the speed and uniformity of judicial decisions must not come at the expense of fundamental principles such as individual dignity, equality before the law, and the preservation of human discretion in the decision-making process[59].

In fact, while the problem of the error inherent in AI systems is most evident in cases where there is no subsequent control by the human being over the output produced, it might be otherwise when it is the human being who makes the formal decision or finally forms the *output* with the help or guidance of an artificial intelligence; be it a judgement, a defensive act or the simple decision to undertake a trial.

However, even in these scenarios, there remains the risk that the technological instrument exerts a form of epistemic authority that is difficult to contradict, especially when it is perceived as objective or neutral by those who use it. This phenomenon, known in the literature as automation bias, can lead judges and legal practitioners to accept algorithmic suggestions even when they contradict their own legal reasoning or ethical intuition[60].

This aligns with Binns's findings, which suggest that algorithmic systems, when embedded in decision-making loops, may displace human judgment by fostering epistemic deference rather than deliberative reasoning[61].

In the latter case, as much as one may consider the issue of curbing the time taken by justice to be meritoriously pursued, a further issue emerges that could be the subject of reflection: the issue of freedom.

Even in cases where the human being has the last word following the production of the *output*, one wonders what the consequences might be in terms of freedom of self-determination on the part of the judge, but also of the lawyer and/or the individual; The question also arises as to what professional responsibilities those who decide to depart from the *output* generated by the AI system incur, or may incur, and whether the production of prognostic indications - which, as we have seen, may at any time prove to be fallacious - on the outcome of a case, on a majority orientation or on the dangerousness

[59] M. HILDEBRANDT, *Law for Computer Scientists and Other Folk*, Oxford, 2020, pp. 153-156.
[60] D. KAHNEMAN, O. SIBONY, C. R. SUNSTEIN, *Noise: A Flaw in Human Judgment*, New York, 2021, pp. 117-119.
[61] R. BINNS, *Human Judgment in algorithmic loops: Individual justice and automated decision-making*, in *Regulation & Governance*, 2022, 16, pp. 197-211.

of a defendant, may represent a direct or indirect limitation on the freedom and autonomy of individuals, ultimately compromising the human-centric foundation of the system.

Moreover, one should not underestimate the potential for AI to reinforce structural biases already present in the legal system. Many datasets used to train predictive models reflect existing disparities in sentencing or in the exercise of prosecutorial discretion. If these data are used uncritically, the resulting models may perpetuate or even exacerbate such inequalities[62]. This aspect is particularly relevant from a bioethical and legal-philosophical perspective, as it concerns the moral responsibility of institutions in ensuring non-discriminatory treatment and respect for human dignity.

In light of the perplexities raised in this paper, it might be appropriate – even if one considers implementing the use of artificial intelligence to assist the operators of the justice system in order to shorten their time – to identify a path through which the principles and safeguards enshrined in the AI Act do not remain mere "dead letter," but rather become capable of finding direct application within the legal systems of the EU Member States ensuring the freedom to decide independently and to deviate from the *output* produced by artificial intelligence without any consequences.

The aim is not only to prevent the technological delegation of human decision-making from leading to technological dependence – generating the loss of worker control over machines, the loss of human and professional qualities, the loss of autonomy, and the dilution of responsibility – but also to ensure the survival of the claim of fairness and justice of the judicial *output* and the lordship of the human being over artificial intelligence.

ABSTRACT: The increasing integration of artificial intelligence into the administration of justice – although promoted as a means to enhance efficiency, in line with Article 6 of the European Convention on Human Rights – raises significant legal and ethical concerns. Drawing on Robert Alexy's theory of the claim to correctness and John Rawls's concept of justice, the article questions the ethical legitimacy of such systems – characterized by an inherent margin of error – in relation to the judicial function, as well as their compatibility with the fundamental principles affirmed in the AI Act. Indeed, even when AI is used to support rather than replace human decision-making, the influence of algorithmic recommendations may lead to cognitive biases and epistemic deference, thereby conflicting with human freedom, autonomy, and the principle of human centrism. The article concludes by calling for the adoption of robust safeguards to ensure that the use of AI in the justice system does not compromise the fundamental values of fairness, freedom, and human dignity.

KEYWORDS: Predictive Justice System – Epistemic Deference – Ethics of AI – AI Act – Human-centrism and AI.

[62] V. EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York, 2018, p. 5.