

# Artificial intelligence and the end of justice

Yulia Razmetaeva\*

**ABSTRACT:** Justice may be nearing its end with the advent of artificial intelligence. The ubiquitous penetration of AI, reinforced by its gaining legitimacy in non-obvious ways, is leading to a shift in the way humans perceive and apply the principles of justice. AI is incapable of truly understanding and interpreting the law, properly justifying decisions, or balancing rights and interests, which escapes public attention as people are excessively focused on its perceived perfection. Difficult to control, AI entails significant dependency of public institutions on private actors. Without undermining artificial intelligence as such, the article is calling to seriously rethink how far we are ready to go along this path.

**KEYWORDS:** Artificial intelligence; algorithms; justice; principles of justice; legitimacy

**SUMMARY:** 1. Principles of justice – 2. Justice concepts for artificial intelligence – 3. Algorithmic logic and individual circumstances – 4. Understanding, interpreting, and justification – 5. Balancing rights and interests – 6. Uncertainty, predictability and a room for debate – 7. (Un)Controlled artificial intelligence and private actors – 8. Legitimacy of algorithms – 9. Conclusions.

## 1. Principles of justice

The growing number of artificial intelligence (AI) technologies and their increasing impact on the public and private spheres of life seem to require an honest answer to the question of whether AI and justice in the human understanding are compatible. Attempts to algorithmise justice as such and the judicial process as the quintessence of justice in particular are becoming more frequent as AI technologies penetrate all levels of human existence. These technologies seem like a logical step forward when considering their accuracy, efficiency, and potential impartiality. Artificial intelligence has what Chris Chambers Goodman calls “an aura of infallibility”, and confidence in the infallibility of AI is growing by the fact that data scientists explain that errors and bias are not the algorithm’s fault.<sup>1</sup>

---

\* *Researcher, Centre for Multidisciplinary Research on Religion and Society, Department of Theology, Uppsala University; Associate Professor at the Department of Human Rights and Legal Methodology, and Head of the Center for Law, Ethics and Digital Technologies, Yaroslav Mudryi National Law University. Mail: [yulia.razmetaeva@crs.uu.se](mailto:yulia.razmetaeva@crs.uu.se). This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation. The author expresses deep gratitude to all colleagues for encouraging this work and to the reviewers whose wise advice and criticism made it possible to significantly improve the quality of this paper. The article was subject to a double-blind peer review process.*

<sup>1</sup>C.C. GOODMAN, *AI/Esq.: Impacts of Artificial Intelligence in Lawyer-Client Relationships*, in *Oklahoma Law Review*, 72, 2019, 149.

The understanding of what AI is has changed in line with the technological development of societies, and today it is rather an umbrella term referring to a range of technologies. Many of those significantly improve people's lives or promise to do so in the foreseeable future, ensuring progress in such spheres as health care<sup>2</sup> or traffic management systems.<sup>3</sup> It may appear, therefore, that the justice system and justice as such could also greatly benefit from AI. However, it seems justifiable to cast a shadow of doubt as to how far we should go implementing artificial intelligence in the justice system? Will the deployment of AI not mean such radical changes in concepts and principles of justice that will be devastating for justice itself? In no way is this article 'algorithmophobic'. Rather, it proposes that we should tread carefully, fully aware of AI's limitations and pitfalls.

If one imagines what justice would be like not in a human world but in a world full of artificial agents, it would most likely be purely rational agents. Rationality itself is not something that endangers justice. Indeed, the principles of justice proposed by John Rawls are the result of reflection in an attempt to figure out what principles would be rational to adopt given the hypothetical initial situation<sup>4</sup> in which people do not know who they are and what potential privileges they have or do not have. However, justice based solely on rationality might become a problem for humanity.

Rawls's concept of justice is but one of many such concepts which form the basis of legal doctrines and judicial practices. At the same time, legal orders grounded on the fundamental principles of human rights, the rule of law, and democracy share the core principles of justice. In particular, such principles include equality in a broad sense or equality before the law and the court as a more specific legal principle. By themselves, the triad of fundamental values also works as principles shared in many legal systems and are also the basis of the European legal order. A more specific European context also includes the values listed in Article 2 of the Treaty on European Union: "The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail".<sup>5</sup>

One needs to adhere to the above principles when it comes to artificial intelligence technologies. Besides, existing well-founded fears about such technologies as artificial intelligence have contributed

<sup>2</sup> See e.g. T. DAVENPORT, R. KALAKOTA, *The potential for artificial intelligence in healthcare*, in *Future Healthcare Journal*, 6(2), 2019, 94-98, <https://doi.org/10.7861/futurehosp.6-2-94>; D. LEE, S.N. YOON, *Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges*, in *International Journal of Environmental Research and Public Health*, 18(1), 2021, 271, <https://doi.org/10.3390/ijerph18010271>; D. M. EL-SHERIF, M. ABOUZID ET AL., *Telehealth and Artificial Intelligence Insights into Healthcare during the COVID-19 Pandemic*, in *Healthcare*, 10(2), 2022, 385, <https://doi.org/10.3390/healthcare10020385>.

<sup>3</sup> See e.g. D. NALLAPERUMA ET AL., *Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management*, in *IEEE Transactions on Intelligent Transportation Systems*, 20(12), 2019, 4679-4690, <https://doi.org/10.1109/TITS.2019.2924883>; A. BOUKERCHE, T. YANJIE, S. PENG, *Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems*, in *Computer Networks*, 182, 2020, 107484, <https://doi.org/10.1016/j.comnet.2020.107484>; A. DEGAS ET AL., *A Survey on Artificial Intelligence (AI) and Explainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory*, in *Applied Sciences*, 12(3), 2022, 1295, <https://doi.org/10.3390/app12031295>.

<sup>4</sup> See J. RAWLS, *A Theory of Justice*, Cambridge, 1971.

<sup>5</sup> Consolidated Version of the Treaty on European Union [2008] OJ C115/13, art 2.

to the emergence of certain value frameworks for them at the level of many legal systems. The numerous guidelines and frameworks regarding AI mentioned justice while the details vary.<sup>6</sup> For instance, the European Union has stressed that AI should be “grounded in our values and fundamental rights such as human dignity and privacy protection”.<sup>7</sup> The value framework for AI based on the principles of justice is proposed in draft acts that are discussed at the levels of the European Union and the Council of Europe, in particular proposals include such principles as democracy and rule of law,<sup>8</sup> as well as respect for human rights and fundamental freedoms, and for rule of law, equality and non-discrimination.<sup>9</sup>

A model of justice that would be universally accepted and uniform for every legal order is hardly ever achievable. However, the core concept of what is just and what is not could be universal. Even people who cannot explain in proper human rights terms the wrongness of how they are treated and how they are subject to suffering and harm can still have vague ideas that this is not right and can rightly demand not to be treated like this.<sup>10</sup> When it comes to details, there will inevitably be essential discrepancies. It is this basic inconsistency that human courts attempt to resolve using a variety of methods, in particular, by balancing rights and interests. Courts, besides, make a significant contribution to forming, developing, and changing some principles of justice in their practice.

In addition, a consensus about principles can be achieved through legal acts issued by legitimate bodies after prior expert discussion and public participation. These mechanisms will potentially lead to the creation of certain elements that can serve as criteria for the implementation of the core principles, such as equality or the rule of law, making them more precise and helping to interpret them. The consensus as to the rule of law may be, in particular, the one proposed by the Venice Commission, which includes such elements as legality, legal certainty, prevention of abuse/misuse of powers, access to justice, as well as equality before the law and non-discrimination.<sup>11</sup>

Not only do the principles of justice develop along with the development of societies, but they also require reflection and rethinking. However, it seems that humans tend to endow the machine with almost divine qualities or even attributes,<sup>12</sup> or at least may somehow divinise such technologies as

<sup>6</sup> M. BRAUN, P. HUMMEL, *Data justice and data solidarity*, in *Patterns*, 3, 2021, <https://doi.org/10.1016/j.patter.2021.100427>.

<sup>7</sup> *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*, COM(2020) 65 final, 2.

<sup>8</sup> Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM(2021)0206 – C9-0146/2021 – 2021/0106(COD). [https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html) (last visited 18/08/2023).

<sup>9</sup> Consolidated working draft of the framework Convention on artificial intelligence, human rights, democracy and the rule of law, Committee on Artificial Intelligence (CAI). Strasbourg, 7 July 2023. CAI(2023)18. <https://rm.coe.int/cai-2023-18-consolidated-working-draft-framework-convention/1680abde66> (last visited 21/09/2023).

<sup>10</sup> M. MAHLMANN, *Mind and Rights: Neuroscience, Philosophy, and the Foundations of Legal Justice*, in M.N.S. SELLERS (ed.), *Law, Reason, and Emotion*, Cambridge, 2017, 96.

<sup>11</sup> *Report on the rule of law*. Adopted by the Venice Commission at its 86th plenary session (Venice, 25-26 March 2011). CDL-AD(2011)003rev-e.

<sup>12</sup> For instance, omniscience or most perfect being can be considered as attributes of divinity. See e. g. about divine attributes, T.J. MAWSON, *The Divine Attributes*, Cambridge, 2018; C. DE FLORIO, A. FRIGERIO, G. GASSER, *Introduction: Divine Attributes*, in *Topoi*, 36, 2017, 561-564, <https://doi.org/10.1007/s11245-016-9414-z>.

general artificial intelligence,<sup>13</sup> and adopt the algorithmic style of thinking,<sup>14</sup> even in matters connected with law and justice, as the perfect one, leaving no room for reflection, changing the perception of justice as such.

Indeed, justice as the core of the justice system is rather an ideal, and perhaps it is the desire to bring the ideal model closer to reality that became the starting point for the deployment of AI. Some AI solutions seem to be able to significantly shorten the length and lower the cost of court proceedings, as well as curb judges' biases. The speed of information processing, including in-depth data analysis according to the given parameters is far beyond human capabilities. A single AI solution can replace many human workers, which can seriously reduce litigation costs. An impartial and tireless algorithm can be a boon for people who have reasonable grounds to suspect that their cases will be unfairly resolved by biased or corrupt judges. However, while trying to offset the shortcomings of the traditional justice system, might we not create even more problems, but of a different nature, the price of which might be too high for humanity?

The cross-cutting methodology of this study is a philosophical analysis and a legal doctrinal approach, specifically focused on the doctrines of justice and the rule of law. The hermeneutic approach is applied in terms of interpretation and understanding of law by human beings and comparing it with the capabilities of certain AI technologies. For the purposes of this article, it should be clarified that although AI is a multiplicity of technologies used in different ways, the research is focused on the consequences for justice of AI as such. Therefore, any distinctions between different AI systems are only relevant insofar as they are important for the concept and principles of justice. In other words the central assumption is that while AI solutions are different, algorithmisation is a general paradigm-shifting problem.

The article is structured as follows: Part 1 is devoted to concepts and principles of justice, as well as the frameworks of the research. Part 2 outlines the problems to choose justice concepts for artificial intelligence, to have a consensus and to overcome the complexity. Part 3 is focused on algorithmic logic and individual circumstances discussing whether AI can provide 'better' justice. Part 4 discusses whether AI will be able to understand, interpret and justify when applying principles of justice. Part 5 sheds light on balancing rights and interests considering whether AI can do this properly and what cases can be entrusted to it in courts. Part 6 addresses uncertainty, predictability and a room for debate considering the implications of algorithmic certainty and forecasting for justice. Part 7 examines the issues of controllability regarding AI and the role of private actors as well as human oversight. Part 8 focuses on the legitimacy of algorithms caused by their imperceptible penetration into all spheres of public and private life and seemingly irreversibility. The article concludes with the main findings of the study in Part 9.

<sup>13</sup> See O. LI, *Artificial General Intelligence and Panentheism*, in *Theology and Science*, 21(2), 2023, 273-287, <https://doi.org/10.1080/14746700.2023.2188373>.

<sup>14</sup> Y. RAZMETAeva, N. SATOKHINA, *AI-Based Decisions and Disappearance of Law*, in *Masaryk University Journal of Law and Technology*, 16(2), 2022, 262.

## 2. Justice concepts for artificial intelligence

There is no denying that AI in courts and public decision-making ought to be developed taking into account the requirements of human rights, transparency and accountability and ought to be built on diverse and protected data. However, when it comes to details and the actual embedding of principles into an algorithm, many questions arise.

First and foremost, prior to programming an algorithm, a definite concept of justice is to be selected. This in itself is not a simple task as there are many concepts of justice and many debates around them.<sup>15</sup> Consensus is possible on the level of common principles of justice, but is it achievable when it comes to the details of concepts?

In addition, one of the questions that will arise is who will choose the principles of justice to be embedded into AI. Choosing between big-tech companies' owners and the authoritative judicial institutions, such as the European Court of Human Rights, the Court of Justice of the European Union or the Supreme Court of the United States, many would choose the courts but in fact, the influence of the former is sometimes greater than that of the latter. As far as technologies are concerned, one big-tech companies' owner might appear to be far more influential than the particular court, no matter how much many would like the opposite. Moreover, the laws of the market make companies launch products fast, without waiting for them to be analysed and approved by lawmakers and courts. As a result, both law-makers and courts are forced to act retrospectively; thus, when the court rules that an algorithm violated certain basic principles of justice, the harm is already caused. Among the negative consequences are, for example, erosion of democratic norms, damages to financial gains, and extending harms to stakeholders.<sup>16</sup>

If the concept is chosen, the complicated judicial concepts are to be turned into lines of code and AI is to be taught to make decisions, which might be a next-to-impossible task given the ultimate complicacy and intricacy of judicial decision-making, – which inherently contains the 'human' part in addition to the purely rational one, – and the peculiarities of machine learning. The idea of codifying justice, or making it computable, has both supporters and opponents. On the one hand, the numerous attempts to build models for measuring fairness and equality reflect the eagerness of many to create justice that relies on AI.<sup>17</sup> On the other hand, there are arguments against, which include the inability of machines

<sup>15</sup> See e.g. T. CAMPBELL, *What Justice is About*, in *Justice. Issues in Political Theory*, London, 1988; R. DWORKIN, *Sovereign Virtue: the theory and practice of equality*, Cambridge, 2000; D.D. RAPHAEL, *Concepts of Justice*, New York, 2001; M. SANDEL, *Justice: What's the Right Thing to Do?*, New York, 2009; I. ENGLARD, *Corrective and Distributive Justice: From Aristotle to Modern Times*, New York, 2009.

<sup>16</sup> S. LU, *Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence*, in *Vanderbilt Journal of Entertainment & Technology Law*, 23(1), 2020, 99.

<sup>17</sup> See L.K. BRANTING ET AL., *Scalable and explainable legal prediction*, in *Artificial Intelligence and Law*, 29, 2021, 213-238, <https://doi.org/10.1007/s10506-020-09273-1>; D. VARONA, J.L. SUÁREZ, *Discrimination, Bias, Fairness, and Trustworthy AI*, in *Applied Sciences*, 12(12), 2022, 5826, <https://doi.org/10.3390/app12125826>; L.F. DE OLIVEIRA ET AL., *Path and future of artificial intelligence in the field of justice: a systematic literature review and a research agenda*, in *SN Social Sciences*, 2, 2022, 180, <https://doi.org/10.1007/s43545-022-00482-w>.

to understand context,<sup>18</sup> tendency to “wrongful generalisations”,<sup>19</sup> and the inclination of many AI models “to cement or even augment existing discriminatory practices and inequalities”,<sup>20</sup> which could make algorithmic justice impossible.

Another premise for the general scepticism towards justice metrics can be the incapability of the machine to reflect on and rethink the decisions taken. This could be called Hannah Arendt’s argument based on her concept of “an inclination and a need to think”.<sup>21</sup> People who apply concepts of justice in courts and public authorities, and people who reflect on them and teach others (lawyers, philosophers, ethicists, etc.) learn and rethink their own approaches all their lives. This includes the transfer of knowledge and experience, and some may say even the formation of a sense of justice and the ability to make moral judgments.

Once the concept is chosen, a lot of collaboration would be necessary among the government, civil society, corporations, NGOs, academicians, etc. However, the stakeholders’ thinking styles are fundamentally different: a judge’s style of thinking is fundamentally different from that of a programmer, so they literally ‘speak different languages’. In addition, as artificial intelligence systems develop, they become increasingly complex and, accordingly, the gap between experts who understand technicalities and people who have to make decisions of public interest deepens. Anna Visvizi calls it the “cognitive divide” between the general audience, the decision makers and the community of experts.<sup>22</sup>

A human judge is a specific person, an algorithm is generalised but it is obscure. A human judge may and will be biased – but one can always talk to a human judge and find out what the decision was based on. A judge is often known for her or his public views, as for example, it was in the cases of Ruth Bader Ginsburg and Antonin Gregory Scalia, judges of the Supreme Court of the United States. An algorithm is impersonal and obscure for a non-programmer. It may and will be biased: but there’s no one to talk to and no easy way to figure out what possible misconceptions and prejudices the decisions were based on.

Attempts to solve the problem of algorithmic bias, which multiplies injustice, often come down to trying to build better algorithms. A more productive AI models are often proposed, for example, one that “can prevent the judgement trap of ‘similar cases with the same judgement’ to the greatest extent”.<sup>23</sup> Moreover, attempts to solve problems created by algorithms using algorithms brings about attempts to decompose fairness into a kind of mathematical components, define algebraically precise criteria of what is and what is not fair, and try “improve fairness” by increasing the computational power of the algorithms, as, for example, in the work whose authors “define the fairness and the corresponding

<sup>18</sup> S. WACHTER, B. MITTELSTADT, C. RUSSELL, *Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI*, in *Computer Law & Security Review*, 41, 2021, 105567, <https://doi.org/10.1016/j.clsr.2021.105567>.

<sup>19</sup> H. COSSETTE-LEFEBVRE, J. MACLURE, *AI’s fairness problem: understanding wrongful discrimination in the context of automated decision-making*, in *AI and Ethics*, 3, 2023, 1265.

<sup>20</sup> I. STRÜMKE, M. SLAVKOVİK, V. I. MADAI, *The social dilemma in artificial intelligence development and why we have to solve it*, in *AI and Ethics*, 2, 2023, 655.

<sup>21</sup> H. ARENDT, *Thinking and Moral Considerations*, in *Social Research*, 38 (3), 1971, 421.

<sup>22</sup> A. VISVIZI, *Artificial Intelligence (AI) and Sustainable Development Goals (SDGs): Exploring the Impact of AI on Politics and Society*, in *Sustainability*, 14 (3), 2022, 1733, <https://doi.org/10.3390/su14031730>.

<sup>23</sup> W. MA, *Artificial Intelligence-Assisted Decision-Making Method for Legal Judgment Based on Deep Neural Network*, in *Mobile Information Systems*, 4636485, 2022, <https://doi.org/10.1155/2022/4636485>.

evaluation metrics” with the aim of “reducing algorithmic bias”.<sup>24</sup> It seems that it will not work in favour of impartiality, since it can be multiplied in AI, nor justice, since it is difficult to assemble the latter into mathematical models and software codes.

Finally, even if people adopted one of the better-developed concepts of justice, like John Rawls’s “justice as fairness”,<sup>25</sup> its complexity is such that it would be impossible to represent it in the form of an algorithm. Even if people somehow managed to algorithmise the extremely complex fairness model, they would then have to apply the algorithm to infinitely complex human relationships. One complexity might be multiplied by the other. Instead of making our lives easier, it could severely confound things. Ultimately, even if humans managed to harness this infinite complexity one day, making the algorithms work the way they should, there would be yet more challenges to overcome in order to create a proper computer model of justice.

### 3. Algorithmic logic and individual circumstances

It is often taken for granted that algorithms can provide ‘better’ justice because they are ‘more intelligent’. Humanity’s entire experience can outweigh algorithmic intelligence, but one-on-one people are already losing the battle. From the well-known experiments involving chess competitions with artificial agents to legal tech<sup>26</sup> solutions and robotic judges,<sup>27</sup> people fall short where the algorithm is able to process an incredible amount of data and calculate a great variety of scenarios. According to Adam Harkens, the “intelligence of an algorithm is judged upon the efficiency it creates, and the economic fluidity it allows”.<sup>28</sup> The focus on efficiency may lead us to a utilitarian understanding of things and a standardised approach replacing an individualised one, contrary to what justice calls for.

The logic of an algorithm requires measuring and weighing various manifestations of human nature, expressing the human part of people with the help of figures and percentages. Things like age, race, gender, and ethnicity are to be assigned numerical attributes, making them the basis of evaluation, which in itself could be considered as discrimination. The application of AI can also bring new dimensions to discrimination in many ways, including appears the algorithmic one. In particular, Ignacio N. Cofone identifies three types of algorithmic discrimination that stem from bias in the process, a bias in

<sup>24</sup> J.-Y. KIM, S.-B. CHO, *An information theoretic approach to reducing algorithmic bias for machine learning*, in *Neurocomputing*, 500, 2022, 36.

<sup>25</sup> See J. RAWLS, *A Theory of Justice*.

<sup>26</sup> See e.g. D.F. ENGSTROM, J.B. GELBACH, *Legal Tech, Civil Procedure, and the Future of Adversarialism*, in *University of Pennsylvania Law Review*, 169, 2021, 1001-1099; G. BUCHHOLTZ, *Artificial Intelligence and Legal Tech: Challenges to the Rule of Law*, in T. WISCHMEYER, T. RADEMACHER (ed.), *Regulating Artificial Intelligence*, Cham, 2020, 175-198, [https://doi.org/10.1007/978-3-030-32361-5\\_8](https://doi.org/10.1007/978-3-030-32361-5_8); K. MANIA, *Legal Technology: Assessment of the Legal Tech Industry’s Potential*, in *Journal of the Knowledge Economy*, 14, 2023, 595-619, <https://doi.org/10.1007/s13132-022-00924-z>.

<sup>27</sup> See e.g. P. RUBIM BORGES FORTES, *Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process*, in *Asian Journal of Law and Society*, 7, 2020, 453-469; J. MORISON, A. HARKENS, *Re-Engineering Justice? Robot Judges, Computerised Courts and (Semi) Automated Legal Decision-Making*, in *Legal Studies*, 39 (4), 2019, 618-635; M.E. KAUFFMAN, M.N. SOARES, *AI in legal services: new trends in AI-enabled legal services*, in *Service Oriented Computing and Applications*, 14, 2020, 223-226.

<sup>28</sup> A. HARKENS, *The Ghost in the Legal Machine: Algorithmic Governmentality, Economy, and the Practice of Law*, in *Journal of Information, Communication and Ethics in Society*, 16(1), 2018, 20.

the input (sample), and the a societal bias captured in representative data.<sup>29</sup> Besides, the opaqueness of AI entails difficulties identifying and assessing whether people evaluated by algorithm were discriminated against on the basis racial or other origin which are protected grounds.<sup>30</sup>

There are several well-known cases where AI built patterns by recreating past discriminatory practices, those which used to exist in societies, or those that still exist but are being fought against. For example, in litigation, AI designed for rating a defendant's risk of committing crimes was prone to significant racial disparities: it is particularly likely to falsely flag black defendants as prospective criminals, while mislabelling white defendants as less likely offenders.<sup>31</sup> Another example is Amazon's hiring algorithm, which filtered off women's resumes and built a pattern based on previous practices of not hiring women for some positions<sup>32</sup> concluding that it is preferable to hire men.

Anti-discrimination solutions that are most often offered include regulating the development of algorithms followed by their examination and experimentation,<sup>33</sup> adding social and political forms of oversight,<sup>34</sup> strengthening the transparency and accountability,<sup>35</sup> improving the quality of data on which algorithms are trained, including large and diverse data,<sup>36</sup> using AI only under the condition of controlling or auditing by humans,<sup>37</sup> as well as using algorithms, especially predictive models, to detect discrimination.<sup>38</sup>

One may argue that measuring and weighing and that the assessment of various traits that makes them the basis of evaluation are also included in what human judges do. Eventually, judges assess the age or gender of both the offenders and the victims of the offence as circumstances that influence the

<sup>29</sup> I.N. COFONE, *Algorithmic Discrimination Is an Information Problem*, in *Hastings Law Journal*, 70, 2019, 1394.

<sup>30</sup> F. ZUIDERVEEN BORGESIU, *Discrimination, artificial intelligence, and algorithmic decision-making*, Council of Europe, Directorate General of Democracy, 2018, 10.

<sup>31</sup> J. ANGWIN ET AL., *Machine Bias. There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, ProPublica, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (last visited 14/03/2024).

<sup>32</sup> See J. DASTIN, *Amazon scraps secret AI recruiting tool that showed bias against women*, Reuters, 2018, <https://www.reuters.com/article/idUSL2N1VB1FQ/> (last visited 15/03/2024); R. IRIONDO, *Amazon Scraps Secret AI Recruiting Engine that Showed Biases Against Women: AI Research scientists at Amazon uncovered biases against women on their recruiting machine learning engine*, DataDrivenInvestor, 2018, <https://medium.datadriveninvestor.com/amazon-scraps-secret-ai-recruiting-engine-that-showed-biases-against-women-995c505f5c6f> (last visited 15/03/2024); L. ANDREWS, H. BUCHER, *Automating Discrimination: AI Hiring Practices and Gender Inequality*, in *Cardozo Law Review*, 44(1), 2022, 145-202.

<sup>33</sup> J. KLEINBERG, J. LUDWIG, S. MULLAINATHAN, C.R. SUNSTEIN, *Discrimination in the Age of Algorithms*, in *Journal of Legal Analysis*, 10, 2018, 114.

<sup>34</sup> M. MANN, T. MATZNER, *Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination*, in *Big Data & Society*, 6(2), 2019, <https://doi.org/10.1177/2053951719895805>.

<sup>35</sup> B. LEPRI, N. OLIVER, E. LETOUZÉ ET AL., *Fair, Transparent, and Accountable Algorithmic Decision-making Processes*, in *Philosophy & Technology*, 31, 2018, 611-627, <https://doi.org/10.1007/s13347-017-0279-x>.

<sup>36</sup> G. RE FERRÉ, *Data donation and data altruism to face algorithmic bias for an inclusive digital healthcare*, in *BioLaw Journal – Rivista Di BioDiritto*, 1, 2023, 116.

<sup>37</sup> A. KÖCHLING, M.C. WEHNER, *Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development*, in *Business Research*, 13, 2020, 837.

<sup>38</sup> I. ŽLIOBAITÉ, *Measuring discrimination in algorithmic decision making*, in *Data Mining and Knowledge Discovery*, 31, 2017, 1060.



decision of the case, for example to justify that specific crimes should be qualified as gender-based violence, domestic violence or violence against minors. In certain categories of cases, the evaluation of human traits is the basis of legal reasoning, for example, when it comes to violations of standards prohibiting discrimination.

However, what distinguishes algorithms from the evaluation that can be given by human judges is that quantitative values are assigned to human traits in advance. In order to create or teach AI, it is necessary to give clear quantitative attribution to these characteristics. It should be noted that this may not apply to self-learning AI technologies. At the same time, self-learning algorithms, in particular artificial neural networks, may have other grounds to limitate them being used in justice system, the main of which is their general obscurity<sup>39</sup> and a more specific problem with explainability that fuels the argument that it would be reckless to use AI tools that decision makers could not understand.<sup>40</sup>

If one can conduct a thought experiment and imagine that human judges weigh race, age, or gender in their minds, assigning them 10%, 12%, and 8% influence on the decision of the case, respectively one may ask: doesn't that way of making decisions by human judges seem strange? This is certainly a simplified view, and AI is capable of processing data from millions of human characteristics and taking into account millions of their combinations. However, the fact that quantification will be expressed in thousandths of a percent and that the accuracy of algorithmic models is increasing every day does not make it less of a problem that people who develop or teach certain AI should take the path of pure utilitarianism by assigning computational values to human characteristics.

On the other hand, there is a limit to the complexity of an algorithmic definition, which inevitably leaves out details, at times – essential details. Algorithms tend to be standardised, justice must be individualised. Everything that lies beyond the scope of an algorithm – unique circumstances, outstanding human manifestations, real intentions and emotional states – will be disregarded by a machine (but would be taken into account by a human judge). This distorts the very idea of justice. While judges, when considering a case and making a decision, can take into account individual circumstances – within the framework of the principles of law, the requirements of substantive and procedural norms – the logic of the algorithm is mainly limited to the evaluation of previously input data. Despite their immense capacity for self-learning, the ability of AI to take into account an infinite number of circumstances and individual traits is questionable.

Indeed, only the most radical proponents of using AI for decision-making in courts and public authorities are currently ready to completely substitute machines for humans. The proposed AI models are designed mainly to support decision-makers rather than replace them. For example, Joe Collenette,

---

<sup>39</sup> See Y. BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, 31(2), 2018, 889–938; S. STRAUß, *From Big Data to Deep Learning: A Leap Towards Strong AI or 'Intelligentia Obscura'?*, in *Big Data and Cognitive Computing*, 2(3), 2018, 16, <https://doi.org/10.3390/bdcc2030016>; R. SCHAEFFER, M. KHONA, I.R. FIETE, *No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit*, in *Advances in Neural Information Processing Systems*, 35, 2022, 16052–16067; J.S. BOWERS ET AL., *On the importance of severely testing deep learning models of cognition*, in *Cognitive Systems Research*, 82, 2023, 101158, <https://doi.org/10.1016/j.cogsys.2023.101158>.

<sup>40</sup> J. ADAMS, *Defending explicability as a principle for the ethics of artificial intelligence in medicine*, in *Medicine, Health Care and Philosophy*, 6(2), 2023, <https://doi.org/10.1007/s11019-023-10175-7>.

Katie Atkinson and Trevor Bench-Capon offer a legally-grounded symbolic model of a complex real world legal domain using methods of computational models of argument and claiming that it is intended only to support legal decision making.<sup>41</sup> Besides, when AI is deployed as supportive it seems to be perceived with less concern than if it were technologies intended to completely replace human decision-makers and apply principles of justice. In particular, Henry Swofford and Christophe Champod found that “people tend to be more receptive to algorithms if they are integrated as a factor that supplements as opposed to supplants human decision making and the human retains some amount of influence on the ultimate outcome”.<sup>42</sup>

The accumulation of legal data and the development of new models and methods of artificial intelligence may provide more reliable and trustworthy technologies that will give the reasons to sceptics to be less concerned with the fact that court decisions are likely to be made with the support of algorithms. In particular, such new models and methods might include those based on machine learning approaches, especially those based on natural language<sup>43</sup> as well as deep learning frameworks.<sup>44</sup> At the same time supporting decisions might turn into making decisions especially if the model looks successful. In addition, the process of relying on the recommendations of artificial intelligence systems can be gradual and inconspicuous, but inevitable in the sense of forming the habit of trusting algorithms more and more. A habit seems to be formed to rely on algorithmic protocols, rather than own reasoning, even in the case of less important, auxiliary algorithms. Moreover, there might be a great temptation to share responsibility for decisions with a high-precision machine.

#### 4. Understanding, interpreting, and justification

Is AI able to understand, interpret and justify when applying principles of justice? In order to analyse this question, one needs to turn to a broader perspective that includes the fundamental problems of agency and thinking.

Isn't it audacious that without fully understanding how the human brain works, humanity seeks to create an artificial one? These attempts to create an alternative to human intelligence include different types of AI, such as, for instance, analytical, human-Inspired, and humanised AI.<sup>45</sup> Such attempts have

<sup>41</sup> J. COLLENETTE, K. ATKINSON, T. BENCH-CAPON, *Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights*, in *Artificial Intelligence*, 317, 2023, 103861, <https://doi.org/10.1016/j.artint.2023.103861>.

<sup>42</sup> H. SWOFFORD, C. CHAMPOD, *Implementation of algorithms in pattern & impression evidence: A responsible and practical roadmap*, in *Forensic Science International: Synergy*, 3, 2021, 100142, <https://doi.org/10.1016/j.fsisyn.2021.100142>.

<sup>43</sup> S. VILLATA ET AL. *Thirty years of artificial intelligence and law: the third decade*, in *Artificial Intelligence and Law*, 30, 2022, 561.

<sup>44</sup> A. TAGARELLI, A. SIMERI, *Unsupervised law article mining based on deep pre-trained language representation models with application to the Italian civil code*, in *Artificial Intelligence and Law*, 30, 2022, 417.

<sup>45</sup> A. KAPLAN, M. HAENLEIN, *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*, in *Business Horizons*, 62(1), 2019, 18.

intensified around designing artificial general intelligence, which would include “consciousness and universal cognitive abilities which at least are on the same level as in humans”.<sup>46</sup>

AI is often seen as perfect, in some inhuman sense of perfection, when it performs tasks impressively efficiently, quickly, and accurately. At the same time, what people see as the result of AI action is not the result of a thinking process, but rather an imitation. Human beings may be inclined to trust algorithms because the latter demonstrate a successful imitation of thinking or something very similar to thinking. Impressions of communication with ChatGPT, for example, can be quite strong to forget that this AI simulates communication and does not really do it as if the user had a human (or almost human) participant of the dialogue on the other side of the screen. That is, this imitation of thinking demonstrated by AI is so impressive that people cannot help endowing the algorithm with subjectivity and even agency.

Bearing in mind the differences between human and non-human thinking, there are no definitive guarantees that any artificial intelligence involved when applying justice will understand the law and interpret it properly. Moreover, there are no reliable guarantees that AI will be able to convince individuals of the fairness of the decision. Undoubtedly algorithms researchers and developers have made significant progress as to AI’s explainability, from specific methods such as, for example, post-hoc explainability methods,<sup>47</sup> to approaches such as, for example, pragmatic tools to increase the transparency of models.<sup>48</sup> Yet it should be emphasised that explanation<sup>49</sup> is not the same as justification.

Algorithms are becoming more and more advanced in recognizing and extracting information from texts. This also applies to legal texts: laws and court decisions, draft acts, essays by jurists, etc. Undoubtedly, the speed of text processing and the accuracy of data extraction with the subsequent forming of coherent and meaningful responses are overwhelming. The information that is extracted by AI is also suitable for further efficient processing since it is often structured data. For instance, such structured data<sup>50</sup> is producing as output of extraction even in natural language processing, which are designed to get closer to understanding information the way humans do. However, even though algorithms are perfect at extraction processing, interpretation and piecing together chunks of data, outputting responses hardly distinguishable from human responses or even outstanding them, we still cannot say that machines ‘understand’ what they ‘read’.

---

<sup>46</sup> O. LI, *Re-creating the world – On necessary features for the creation of AGI*, in *New Techno Humanities*, 3(1), 2023, 57.

<sup>47</sup> D. VALE, A. EL-SHARIF, M. ALI, *Explainable artificial intelligence (XAI) post-hoc explainability methods: risks and limitations in non-discrimination law*, in *AI Ethics*, 2, 2022, 815-826, <https://doi.org/10.1007/s43681-022-00142-y>.

<sup>48</sup> L. LONGO, R. GOEBEL, F. LECUE, P. KIESEBERG, A. HOLZINGER, *Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions*, in A. HOLZINGER, P. KIESEBERG, A. TJOA, E. WEIPPL (eds.), *Machine Learning and Knowledge Extraction. CD-MAKE 2020. Lecture Notes in Computer Science*, 12279, Cham, 2020, [https://doi.org/10.1007/978-3-030-57321-8\\_1](https://doi.org/10.1007/978-3-030-57321-8_1).

<sup>49</sup> This research did not intend to dwell on the complex issue of AI explainability, as its main point is that it that is not enough to understand the mathematical algorithms behind decisions: to justify the justice is the primary focus of this section.

<sup>50</sup> K. KREIMEYER ET AL., *Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review*, in *Journal of Biomedical Informatics*, 73, 2017, 19.

Automatic extraction of information from machine-readable legal texts has advantages over ‘manual’ extraction. It’s faster and cheaper, can be more error-free, and ultimately more efficient. In particular, the joint end-to-end model in which extraction is coupled with classification for evidence information in legal texts (court record documents) shows experimental results as 72.36% F1 score, outperforming baseline systems.<sup>51</sup> Moreover authors of this study presuppose that this model “can be applied for better analysing and understanding legal texts, avoiding a lot of manual labour by experts and professionals in the legal field”.<sup>52</sup> Such models, under certain conditions, can be excellent supporting tools for people to reduce the time they have to spend to analyse the facts of the case, legislation and legal practice before making a decision.

If we give the algorithm certain criteria, it will successfully cope with finding, recognizing and ‘understanding’ that, for example, 100 judgments in the database are decisions to enforce the collection of child support from parents. A smart algorithm that is given a specific task can help us establish patterns in the sample. For example, it may show that 49% of the court cases mentioned above involve parents permanently residing in countries other than the applicant’s country of residence, and 17% of these 49% have not contacted a parent, who takes care of children on a full-time basis, for more than 2 years. This information, sorted into a convenient format, can help us understand some patterns, predict trends, and build models regarding child support cases.

Perhaps AI is capable of analysis and synthesis and therefore can claim at least partially to be an agent that thinks even if purely rationally. However is the algorithm really capable of understanding legal texts? Can algorithms extract meanings from texts? It is reasonable to assume that at least some of them can understand the text’s meaning in terms of capability to find and to analyse the facts. At the same time it can be argued that any algorithm can not understand the texts as it cannot interpret the facts in the light of an infinite number of contexts. In contrast, in order to determine what a fair decision will be in a particular case, human judges apply not only legislation and precedents, but also infinitely complex considerations based on their understanding of the law and the experience of the law. Thus, it can be assumed that any judge is capable of understanding the text’s meaning, including the ability to both find and analyse the facts and interpret them in the light of an infinite number of contexts. This undoubtedly refers to any judge who has passed a qualifications-based selection for the position, provided there is an independent judicial system in the society that demonstrates a certain level of public trust and where there is a legal system based on the rule of law. In addition, the judges must justify this decision in such a way as to convince others of its fairness. Understanding the law, that is the human understanding of it, is an important part of the interpretation of law. The proper interpretation and convincing explanation is also part of the rule of law. This will probably forever remain beyond the scope of the machine.

<sup>51</sup> D. JI, P. TAO, H. FEI, Y. REN, *An end-to-end joint model for evidence information extraction from court record document*, in *Information Processing and Management*, 57, 2020, 102305, <https://doi.org/10.1016/j.ipm.2020.102305>.

<sup>52</sup> D. JI, P. TAO, H. FEI, Y. REN, *An end-to-end joint model for evidence information extraction from court record document*, cit.

## 5. Balancing rights and interests

In many cases, the implementation of the concept of justice calls for balancing rights and interests as well as accounting for the ever-changing society. This is a well-trodden path for human beings – but it seems impossible for AI. As Luciano Floridi rightly noted when writing about the apparent neutrality of AI: “The design of any technology is a moral act”.<sup>53</sup> Applying the principles of justice is a moral act. Moreover, it is a moral act that must be done against the background of the development of society and, accordingly, concepts of justice. Then why does one seem to assume that it is possible to develop a perfect technology, trust that technology to apply principles of justice and even make decisions, and stop there?

Human societies evolve, and concepts of justice evolve along with it. Despite this, we are still trying to preserve the fundamental principles of our coexistence. In particular, such fundamental foundations as human rights do not exist in a vacuum, they intersect and conflict, being implemented by people in society, and are balanced by the courts if there is a need to protect them. Legal doctrines, such as the doctrine of dynamic<sup>54</sup> or evolutive<sup>55</sup> interpretation of the Convention for the Protection of Human Rights and Fundamental Freedoms applied by the European Court of Human Rights, concentrate but do not conserve legal experience. Thus, courts preserve the line of justice, but take into account the development of society. Can algorithms properly address these considerations and maintain a delicate and dynamic balance?

The ability to properly balance rights and interests also depends on proper understanding and justification. The latter, in turn, seems to include an understanding of the discourse as such or, in a narrower sense, of the context of a specific case. In particular, to balance freedom of expression with privacy, prohibition of discrimination or protection against hate speech, it might be necessary to understand the discourse of freedom of expression, as well the political, social, religious, etc. context of the cases. Therefore, judges can, for example, distinguish protest forms of expression from public order violations, and interpret hate speech as contradicting the values of the Convention. For example, the ECtHR found that in the case of *Shvydka v. Ukraine*, there was a non-verbal expression protected by Article 10.<sup>56</sup> In this case, the applicant being a representative of an opposition political party broke a ribbon on the wreath during an official ceremony. She then ripped the ribbon with the name of the Ukrainian President Yanukovich thereby asserting her protest against the fact that he is the president. In the case of *Lenis v. Greece* ECtHR established that the complaint regarding the application of Article 10 is inadmissible.<sup>57</sup> In this case, a church official posted a homophobic article on his personal blog, when the Greek Parliament had been about to debate proposed legislation introducing civil unions for same-sex

<sup>53</sup> L. FLORIDI, *On Good and Evil, the Mistaken Idea That Technology Is Ever Neutral, and the Importance of the Double-Charge Thesis*, in *Philosophy & Technology*, 36 (60), 2023, <https://doi.org/10.1007/s13347-023-00661-4>.

<sup>54</sup> See P. ŁĄCKI, *Consensus as a Basis for Dynamic Interpretation of the ECHR – A Critical Assessment Get access Arrow*, in *Human Rights Law Review*, 21(1), 2021, 186-202, <https://doi.org/10.1093/hrlr/ngaa042>.

<sup>55</sup> See K. DZEHTSIAROU, *European Consensus and the Evolutive Interpretation of the European Convention on Human Rights*, in *German Law Journal (Special Issue Legitimacy and the Future of the European Court of Human Rights)*, 11(10), 2011, 1730-1745, <https://doi.org/10.1017/S2071832200017533>.

<sup>56</sup> See *Shvydka v. Ukraine* (Application No. 17888/12), European Court of Human Rights, 30 October 2014.

<sup>57</sup> See *Lenis v. Greece* (Application No. 47833/20), European Court of Human Rights, 30 August 2023.

couples. The court ruled that the applicant was attempting to deflect the freedom of expression of the Convention from its real purpose by using it for ends which were clearly contrary to the values which the Convention sought to promote. Both discourse and context are so far something human beings can be immersed in, but not algorithms.

It therefore seems that there is a very limited list of types of cases that can be entrusted to AI. The application of artificial intelligence is possible in cases that are routine and simple and do not require in-depth analysis but habitual procedures. In particular, these may be cases related to family law, uncomplicated, repetitive and related to the formal recognition of actual status. For example, in many legal systems, going to court is required in cases where both parties want to divorce and have no claims against each other but have common minor children. AI application is possible where the complexity of the case is extremely low due to the fact that the circumstances are very simple and there is no dispute in the sense of a conflict as such, that is, there is no real conflict of law and legitimate interests that should be determined or balanced. In most cases, the outcome of such cases is quite predictable. Cases that are a rather formal part of an administrative or civil process can also be algorithmised, in particular, those related to the issuance of judicial prescriptions or confirmation of established facts. The use of AI should be determined by the way cases are handled. In turn, the method of consideration, as Dory Reiling writes, depends on such factors as the complexity of the information in a case as well as the degree of predictability of the outcome.<sup>58</sup> To develop this idea it should be emphasised that the predictability of the outcome here will rather mean not the availability of the most accurate forecast with the help of AI, but the absence of doubts about the essence and circumstances of the case, which allows human judge to reasonably assume what the outcome will be as law and justice demand.

In those legal systems that have actively implemented predictive analytics, the perception of AI in the courts may be more lenient or based on the belief that this, on the contrary, will contribute to compliance with the requirements of justice. For instance, such predictions, as Daniel L. Chen mentioned regarding to the legal systems of the United States, “would not be used to suggest a decision, but used as inputs to increase efficiency and fairness of law”.<sup>59</sup> However, the question of foreseeability should be considered from a broader perspective in order to assess how the latter is compatible with justice.

## 6. Uncertainty, predictability and a room for debate

Algorithms are designed to help cope with uncertainty, allowing people to calculate and predict future decisions. As Kathrin Hartmann and Georg Wenzelburger write, “it seems that the main impetus for the use of algorithmic evidence indeed is the perceived reduction in uncertainty”.<sup>60</sup> Nevertheless, uncertainty and the capacity for doubt are also important parts of human nature.

<sup>58</sup> A. D. (DORY) REILING, *Courts and Artificial Intelligence*, in *International Journal for Court Administration*, 11(2) (8), 2020, <https://doi.org/10.36745/ijca.343>.

<sup>59</sup> D.L. CHEN, *Judicial analytics and the great transformation of American Law*, in *Artificial Intelligence and Law*, 27, 2019, 16.

<sup>60</sup> K. HARTMANN, G. WENZELBURGER, *Uncertainty, Risk and the Use of Algorithms in Policy Decisions: A Case Study on Criminal Justice in the USA*, in *Policy Sciences*, 54, 2021, 284.

The ‘uncertainty’, one of the definitions of which is “unpredictability of events”,<sup>61</sup> might be a problem in some contexts, but it is not something humans can – and should – overcome once and for all. AI, however, is aimed at certainty, at the final resolution of some problem. It also seeks to increase predictability to the greatest extent possible. Justice, on the other hand, is not barely about certainty. It is very often about uncertainty and doubt. That is why people need independent and fair courts to resolve doubts to a certain extent and to individualise justice depending on the legal basis and circumstances of the case.

What’s more, it seems that a lack of room for doubts and debates stemming from these doubts means a lack of justice. The long-term consequences of handling complex cases and making informed decisions often involve extensive public debate. Such a deliberation is capable of provoking changes in legislation and legal practice, or in a broader sense, changes in legal approaches and serious social shifts. The absence of discussion and the replacement of debate with a percentage ratio of calculatedness in decisions impoverishes human societies.

One might claim that the output of algorithms does not express as 100% certainty, but rather represents a probability scale. Therefore it is the judge who decides, and AI only suggests what the decision should be with a certain percentage of probability. A similar logic was applied by the Wisconsin Supreme Court in *Loomis v. Wisconsin* indicating that the risk assessment algorithm was only a recommendation for the judge, was not the sole basis for a decision, and did not affect independent decision-making, which coincided with the algorithm’s assessment.<sup>62</sup>

AI probabilistics, especially machine learning probabilistic, is extremely useful when it helps to define things that do not allow double interpretation, such as, for instance, when determining sequences in viruses or wind speed. On the other hand, probabilistics can be controversial when it somehow determines the outcome of a judicial decision. It is true that a causal relation between an algorithmic recommendation and a particular coincident decision could be dubious, making it hard to prove that it was causal even if it actually was. In light of this, it is worth paying particular attention to the outcome of the *SCHUFA Holding (Scoring)* case,<sup>63</sup> in which the Court of Justice of the European Union had to determine whether the automated establishment of a credit score should be interpreted as a decision made and based solely on automated processing or profiling.<sup>64</sup> Recently the Court ruled that, under certain conditions, the automated establishment of a probability value based on personal data related to a person must be interpreted as “automated individual decision-making” when a third party strongly

---

<sup>61</sup> G. MERLHIOT, M. MERMILLOD, J.-L. LE PENNEC, F. DUTHEIL, L. MONDILLON, *Influence of uncertainty on framed decision-making with moral dilemma*, in *PLoS ONE*, 13(5), 2018, e0197923, <https://doi.org/10.1371/journal.pone.0197923>.

<sup>62</sup> See *State v. Loomis*, 881 N.W.2d 749, 2016.

<sup>63</sup> See *Request for a preliminary ruling from the Verwaltungsgericht Wiesbaden (Germany)* lodged on 15 October 2021 — *OQ v Land Hesse*, Case C-634/21; Opinion, *SCHUFA Holding and Others (Scoring)*, Case C-634/21, ECLI:EU:C:2023:220.

<sup>64</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016, OJ L 119/1, Article 22.

relies on this probability to establish, implement or terminate a contractual relationship with that person.<sup>65</sup> Determinative or not, the recommendation provided by an algorithm can be perceived as more accurate than a human one.

On the basis of those AI models that demonstrate high accuracy, even more accurate models are developed. This was the case with AlexNet,<sup>66</sup> which, as a model of image classification, demonstrated the potential of deep neural networks and launched the new era of them. The high accuracy of some algorithms can be perceived as part of their perfection and ultimately infallibility in the proposed outputs. Indeed, AI produces results in terms of a probability but this is an outstanding probability. The decision-makers who see the output with a probability of 98%, or even more than 98% classification accuracy on some datasets,<sup>67</sup> may be inclined to over-rely on algorithmic recommendations, especially if they know that the 2% inaccuracy can be explained by differences in how to classify certain characteristics and not by the inability to classify them. This might happen, for instance, in the case of AI that recognises images, because people who add images to datasets do not agree among themselves how to categorise<sup>68</sup> about 2% of them. On the other hand, to test the assumption of a tendency to follow algorithmic recommendations, it would be extremely useful to compare the statistical data of the decisions made by human judges with the suggestions of AI in those jurisdictions where they are used as a part of legal proceedings.

We are seeing an explosion of attempts to build AI-based prognostic algorithms to forecast decision outcomes with a high level of ‘certainty’. A search of the term ‘algorithmisation’ in Sciencedirect base showed 1,910,621 results as of May 28, 2022, and it turned into 1,000,000+ results as of September 18, 2023. A significant part of the articles from the first 50 in this sample (suggested by the algorithm when sorted by relevance) reflect efforts to come up with better algorithms and improve their performance. In particular, some articles suggest an adaptive search algorithm for different types of optimization problems, which shortens the time to find solutions,<sup>69</sup> “an intelligent switching mechanism that can adapt the algorithm selection based on the observed landscape during the optimization”,<sup>70</sup> and

<sup>65</sup> See *Judgment of the Court (First Chamber) of 7 December 2023 SCHUFA Holding (Scoring)*, Case C-634/21, ECLI:EU:C:2023:220.

<sup>66</sup> A. KRIZHEVSKY, I. SUTSKEVER, G.E. HINTON, *ImageNet Classification with Deep Convolutional Neural Networks*, in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, 1097-1105.

<sup>67</sup> See e.g. E. ELYAN AT AL., *Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward*, in *Artificial Intelligence Surgery*, 2(1), 2022, 24-45, <http://dx.doi.org/10.20517/ais.2021.15>; Y. LI, X. FENG, Y. LIU, X. HAN, *Apple quality identification and classification by image processing based on convolutional neural networks*, in *Scientific Reports*, 11, 2021, 16618, <https://doi.org/10.1038/s41598-021-96103-2>; M. BLAIVAS, R. ARNTFIELD, M. WHITE, *DIY AI, deep learning network development for automated image classification in a point-of-care ultrasound quality assurance program*, in *Journal of the American College of Emergency Physicians Open*, 1(2), 2020, 124-131, <https://doi.org/10.1002/emp2.12018>.

<sup>68</sup> For example, the ImageNet database contains more than 14 million images that were manually annotated, categorised, and used then to teach AI. When labelling an image, people did not always agree on which category it should be assigned to.

<sup>69</sup> G.Z. OZTAS, S. ERDEM, *Random search with adaptive boundaries algorithm for obtaining better initial solutions*, in *Advances in Engineering Software*, 169, 2022, 103141, <https://doi.org/10.1016/j.advengsoft.2022.103141>.

<sup>70</sup> K. MEIDANI, S. MIRJALILI, A.B. FARIMANI, *Online metaheuristic algorithm selection*, in *Expert Systems With Applications*, 201, 2022, 117058, <https://doi.org/10.1016/j.eswa.2022.117058>.



the gradient algorithm for independent component analysis for complex valued data, which showed very good results in terms of speed of operation and separation quality and could be useful in on-line applications.<sup>71</sup> It appears that the amount of work with algorithms conducted at the moment is mind-blowing, the bulk of this work being devoted to their optimisation, raising their effectiveness and efficiency.

Another search of the terms ‘algorithm’ plus ‘court’ with 17,468 results as of May 28, 2022 and 19,216 results as of September 18, 2023 showed that a large part of these articles, including those offered first at the top of the search from the first 50 in this sample were devoted to algorithmic prediction of court decisions in different jurisdictions. Some of these studies are aimed at indirect forecasting, for example, at understanding the motivation behind and strategies of the the Supreme Court of the United States judges’ votes – in particular, the ways the justices group themselves into a majority coalition and a minority coalition.<sup>72</sup> Other studies are aimed at direct prediction of verdicts, for example, predicting the outcomes of the courts of the Turkish legal system with high accuracy using deep learning-based methods.<sup>73</sup> Moreover, some of the studies mentioned above directly present the ability to apply the proposed algorithms as an advantage that allows to “assist lawyers and judges to verify trends of judgments” and “help lawyers predict the judicial outcomes for very specific situations”.<sup>74</sup> In addition, an open repository of judgement documents called the European Court of Human Rights Open Data (ECHR-OD) were suggested<sup>75</sup> which can, among other things, contribute to a better prediction of the European Court of Human Rights judgements. This could undermine the very concept of justice since if we already know the outcome of a case, there’s little if any reason to seek justice at all. Striving for predictability appears to be outstanding and seems to outshadow striving for justice and fairness. However, predictability – and certainty as such – in law is not nearly the same as algorithmic certainty.

Indeed, legal certainty is an element of the rule of law and includes, among other things, the requirements of predictability. These requirements in the most general sense mean that people have the right to understand the consequences of their actions and the content of laws. However, the predictability of the consequences of legal or illegal actions does not mean that we can have one hundred percent certainty in the outcome of a lawsuit. Rather, it might work only in certain simple cases, where the right and the wrong, the fair and the unfair are obvious. In complex court cases, on the contrary, those in which there is a conflict of rights, or a clash of rights and legitimate interests, predictability does not work in favour of the rule of law. The predictability of the outcome of complex cases cannot but be

<sup>71</sup> D. MIKA, *Fast gradient algorithm with toral decomposition for complex ICA*, in *Mechanical Systems and Signal Processing*, 178, 2022, 109266, <https://doi.org/10.1016/j.ymsp.2022.109266>.

<sup>72</sup> N. GIANIRACUSA, C. RICCIARDI, *Computational geometry and the U.S. Supreme Court*, in *Mathematical Social Sciences*, 98, 2019, 98, 1-9.

<sup>73</sup> E. MUMCUOĞLU, C.E. ÖZTÜRK, H.M. OZAKTAS, A. KOÇ, *Natural language processing in law: Prediction of outcomes in the higher courts of Turkey*, in *Information Processing and Management*, 58, 2021, 102684, <https://doi.org/10.1016/j.ipm.2021.102684>.

<sup>74</sup> W.P.D. FERNANDES, I.Z. FRAJHOF ET AL., *Extracting value from Brazilian Court decisions*, in *Information Systems*, 106(10196), 2022, <https://doi.org/10.1016/j.is.2021.101965>.

<sup>75</sup> A. QUEMY, R. WREMBEL, N. ŁOPUSZYŃSKA ET AL., *A large reproducible benchmark on text classification for the legal domain based on the ECHR-OD repository*, in *Information Systems*, 119, 2023, 102258, <https://doi.org/10.1016/j.is.2023.102258>.

low due to complexity and the dynamic nature of the certain balance of rights coupled with a lack of consensus as to this balance.

AI works the reasons based on the previous experience which was input in the form of data. Similarly, judges, in most cases, take the bulk of the previous experience and precedents into account. That said, human judges are potentially ready and able to deal with entirely new cases, passing fair decisions regardless of the cases' novelty. While interpersonal conflicts do fall under a limited number of scenarios, the particular details and circumstances are always different, hardly predictable and probably impossible to anticipate.

## 7. (Un)Controlled artificial intelligence and private actors

Entrusting AI with applying the principles of justice in courts and public decision-making will likely lead to their uncontrollability, which is a tri-fold problem. First and foremost, it might be difficult to control algorithms due to the speed of their development as new technologies get deployed before it is clear what hazardous long-term consequences they will entail. Due to this overwhelming speed, attempting to harness AI is similar to attempting to harness a wild horse at full gallop. Never before have such profoundly and imperceptibly world-changing phenomena grain into the lives of individuals and societies with such speed and irreversibility.

Secondly, as it was discussed above, AI is obscure. Its technicalities are incomprehensible for the general public, and its complexity stands in the way of understanding its hidden mechanisms. This renders AI intangible, impersonal and even awe-inspiring. It's no wonder, then, that it has already been suggested by many specialists that AI is to be subject to human oversight.<sup>76</sup>

While such oversight seems to be a good thing, it shouldn't be a 'sedative pill'. As Ben Green rightly maintains, "human oversight policies create a regulatory loophole: it provides a false sense of security in adopting algorithms and enables vendors and agencies to foist accountability for algorithmic harms onto lower-level human operators".<sup>77</sup> He proposed a shift from human oversight to institutional oversight, obliging agencies to justify that it is appropriate to adopt an algorithm through evidence-based, public and transparent process.<sup>78</sup> Be it human or institutional control, a false sense of security appears; people are led to think that they exercise a full control over algorithms, while they are not. It cannot possibly be the case as humans create algorithms to process enormous amounts of data – after all, that is their primary function – which means that controlling them would take an enormous amount

<sup>76</sup> See M. TADDEO, L. FLORIDI, *How AI can be a force for good*, in *Science*, 361, 2018, 751–752; R. KOULU, *Proceduralizing control and discretion: Human oversight in artificial intelligence policy*, in *Maastricht Journal of European and Comparative Law*, 27(6), 2020, 720-735; L. ENQVIST, 'Human oversight' in the EU artificial intelligence act: what, when and by whom?, in *Law, Innovation and Technology*, 2023, <https://doi.org/10.1080/17579961.2023.2245683>; B. MESKÓ, E.J. TOPOL, *The imperative for regulatory oversight of large language models (or generative AI) in healthcare*, in *npj Digital Medicine*, 6, 2023, 120.

<sup>77</sup> B. GREEN, *The flaws of policies requiring human oversight of government algorithms*, in *Computer law & security review*, 45, 2022, 105681, <https://doi.org/10.1016/j.clsr.2022.105681>.

<sup>78</sup> B. GREEN, *The flaws of policies requiring human oversight of government algorithms*, cit.



of human time and effort, and this is either impossible or offsets the very reason people create algorithms. Despite numerous attempts to overcome the lack of control over AI systems<sup>79</sup> and a growing understanding that this requires serious joint efforts and possibly new institutions in various jurisdictions, the problem of impossibility may not be solved in the near future. In other words, human beings are not able to follow what algorithms do since the impressive speed and volume of data processing that AI technologies demonstrate in the ways in which they operate.

Thirdly, AI has always been and will inevitably remain less controllable than other official entities responsible for implementing justice due to the private ownership of algorithms, their parts or entire AI systems. By and large, companies refuse to reveal details of their algorithms' functioning, referring to the fact that such information is part of their trade secrets. In some cases, for example with algorithms intended for medical care, this creates additional risks<sup>80</sup> from the deployment of AI. In a broader sense, private ownership of chunks of code, which are part of an algorithm designed to decide a person's destiny, contradicts the essential principles of justice.

In addition, private actors who develop and sell AI technologies "will have incentives to encourage public disillusionment with the legal system by accentuating the inconsistency and bias of human judgement – and pitching their own services as superior".<sup>81</sup> Therefore the dependence of public authorities on the private AI market will perhaps only deepen. Among other things, this could contribute to the further inevitable penetration of algorithms into the public sphere of peoples' lives and societies functioning.

## 8. Legitimacy of algorithms

'Wait – I did not vote for that algorithm – and it's already telling me what is just and what not!' – this reaction is expectable – and explicable. The gradual, imperceptible and profound penetration of algorithms into public and private spheres of lives has suddenly reached a point where they are literally acquiring legitimacy.

According to Adamantia Rachovitsa and Niclas Johann, algorithmic systems of different types may "create and sustain its own powerful claim to self-referential exceptional legitimacy and authority"<sup>82</sup>. The usefulness and necessity of algorithms seems to no longer need justification, and because of this usefulness, people do not question their legitimacy. A continuing intrusion of AI into both everyday tasks

<sup>79</sup> See D.R. AMARILES, P.M. BAQUERO, *Promises and limits of law for a human-centric artificial intelligence*, in *Computer Law & Security Review*, 48, 2023, 105795, <https://doi.org/10.1016/j.clsr.2023.105795>; L. FLORIDI ET AL., *CapAI – A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act*, 2022, <https://papers.ssrn.com/abstract=4064091> (last visited 03/12/2023); J. DAVIDOVIC, *On the purpose of meaningful human control of AI*, in *Frontiers in Big Data*, 5, 2022, 1017677, <https://doi.org/10.3389/fdata.2022.1017677>.

<sup>80</sup> See A. DURKIN, P.A. STA MARIA, B. WILLMORE, A. KAPCZYNSKI, *Addressing the Risks That Trade Secret Protections Pose for Health and Rights*, in *Health and Human Rights*, 23(1), 2021, 129-144.

<sup>81</sup> R.M. RE, A. SOLOW-NIEDERMAN, *Developing Artificially Intelligent Justice*, in *Stanford Technology Law Review*, 22, 2019, 247.

<sup>82</sup> A. RACHOVITSA, N. JOHANN, *The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case*, in *Human Rights Law Review*, 22(2), 2022, ngac010, <https://doi.org/10.1093/hrlr/ngac010>.



and the justice system becomes irreversible, because individuals already perceive them as part of the familiar landscape and an integral part of their lives as well as the society as a whole. It now appears next to impossible to separate the technology from humans and abandon the benefits that people have got accustomed to relying on. 'Indispensable' appears to begin to mean 'legitimate'.

The formation of human systems based on the principles of justice took, without exaggeration, centuries and is still ongoing. Most elements of these systems (legislative, executive and judicial power) include certain requirements of legitimacy. AI seems to have bypassed these requirements. For instance, it seems applicable to Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), an algorithm designed to assess a defendant's risk of recidivism and are using by US courts in several states<sup>83</sup> or to System Risk Indication (SyRI), an algorithm designed to identify potential social welfare fraud and used by Dutch government's.<sup>84</sup>

The use of algorithms in public decision making can be seen as growing. This may reflect in a direct use of AI solutions, as, for example, in the case of migration and asylum management<sup>85</sup> or indirect use of them, as, for example, in the case of public agencies make decisions based on recommendations provided by institutions that used credit scoring models<sup>86</sup> for giving them. The standards of democratic legitimacy include a certain level of access to information for people, as well as opportunities available to everyone to participate in the discussion and resolve issues of public interest. It also assumes that people can understand and approve (or disapprove) of decisions and actions that are of public interest or significantly affect social groups or society as a whole. According to Pascal D. König and Georg Wenzelburger certain form of decision-making is also part of the requirements of democratic standards, since "liberal democracy is mainly about how outputs are produced".<sup>87</sup> Changing the form of government or granting broad autonomy to regions within the state requires long preparation, special procedures and, most likely, the expression of the will of the people in a referendum or several other

<sup>83</sup> See A.L. WASHINGTON, *How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate*, in *Colorado Technology Law Journal*, 17(1), 2018, 131-160; C. THOMAS, A. PONTÓN-NÚÑEZ, *Automating Judicial Discretion: How Algorithmic Risk Assessments in Pretrial Adjudications Violate Equal Protection Rights on the Basis of Race*, in *Minnesota Journal of Law & Inequality*, 40(2), 2022, 371-407, <https://doi.org/10.24926/25730037.649>; J. DRESSEL, H. FARID, *The accuracy, fairness, and limits of predicting recidivism*, in *Science Advances*, 4, 2018, eao5580, <https://doi.org/10.1126/sciadv.aao5580>.

<sup>84</sup> See A. RACHOVITSA, N. JOHANN, *The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case*, cit.; M. VAN BEKKUM, F.Z. BORGESIU, *Digital welfare fraud detection and the Dutch SyRI judgment*, in *European Journal of Social Security*, 23(4), 2021, 323-340, <https://doi.org/10.1177/13882627211031257>; N. APPELMAN, R.Ó. FATHAIGH, J. VAN HOBOKEN, *Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands*, in *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 12(4), 2021, <https://www.jipitec.eu/issues/jipitec-12-4-2021/5407> (last visited 24/09/2023).

<sup>85</sup> L. NALBANDIAN, *An eye for an 'I': a critical assessment of artificial intelligence tools in migration and asylum management*, in *Comparative Migration Studies*, 10, 2022, 32, <https://doi.org/10.1186/s40878-022-00305-0>.

<sup>86</sup> See R.Y. GOH, L.S. LEE, *Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches*, in *Advances in Operations Research*, 2018, 1974794, <https://doi.org/10.1155/2019/1974794>; K. LANGENBUCHER, *Responsible A.I.-based Credit Scoring – A Legal Framework*, in *European Business Law Review*, 31(4), 2020, 527-572, <https://doi.org/10.54648/eulr2020022>.

<sup>87</sup> P.D. KÖNIG, G. WENZELBURGER, *Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics*, in *Government Information Quarterly*, 37(3), 2020, 101489, <https://doi.org/10.1016/j.giq.2020.101489>.

democratic procedures. The deployment of AI rather resembles an attempt to receive tacit recognition that AI has imperceptibly and irreversibly established in society including public decision-making and the justice system. In other words, it looks like an attempt to grant that since AI is already here, human beings must accept its inevitable intervention.

The requirements for institutions that people consider legitimate have been crystallising over centuries. Despite strong checks and balances these institutions are still far from the ideal model, and regularly faced with problems of justice in various jurisdictions. Adding AI, which is already gaining legitimacy and is uncontrollable in many ways, only may make it worse. It is doubtful that algorithmic justice will successfully overcome both old and new challenges. The already existing problems with justice might be multiplied by, not offset with, emerging trials and tribulations.

## 9. Conclusions

This article certainly does not aim to urge humanity to ditch artificial intelligence. Nevertheless, the debate around AI and justice is far from hypothetical or far-fetched. Discussions as to the coexistence of human agents and artificial ones are already being led, – AI is ‘a thing of the present’, rather than some theoretical threat looming in the distant future. AI already occupies a significant part of the current life of people and societies. Is entrusting justice to AI something we really want? After all, justice is one of the core values of human lives and one of society’s foundations – while artificial intelligence will never be truly human. Wouldn’t it look like attempts of human beings to get rid of what actually makes them different from machines?

It seems that individuals are gradually forgetting that it is humans who created the machine, not the other way around. There are a number of good reasons not to entrust the issue of justice to AI technologies, some of which have been presented in this study. Artificial intelligence, as it seems today at least, cannot and will never be able to understand and interpret the complexity and intricacy of law; nor will it be able to balance rights and legitimate interests or provide proper justification for the decisions taken. Artificial intelligence is probably going to remain difficult to control, in particular – but not exclusively – due to the role of private actors in its development and promotion. That said, AI is set to increasingly acquire legitimacy, which will inevitably pose a threat to public institutions and, in essence, the entire structure of society. All of the above might change the perception of justice in societies and ultimately bring about the end of justice as we know it.

Yes, perhaps this article should have been titled more precisely Artificial Intelligence and the End of Justice as We Know It. It appears that attempts to supplement human justice with algorithmical one will inevitably lead to the replacement of the former by the latter. Considering the fact that the foundations of human existence are anthropocentric and axiological, the whole concept of justice will change irreversibly, dramatically and arguably for the worse. This may seem rather dystopian, in a sense even Orwellian, a picture – but this is what the author would like to prove wrong about.

*W. S. J. van*

