



Bias in AI (Supported) Decision Making: Old Problems, New Technologies

ANDREJ KRIŠTOFÍK 

ACADEMIC
ARTICLE



ABSTRACT

Recently different regulations and recommendations for the use of AI-based technology, especially in the judiciary, have become more prevalent. One major concern is addressing bias in such systems. In 2016, ProPublica published a damning report on the use of AI-based technology in making decisions about people's rights and obligations, revealing that such systems tend to replicate and amplify existing biases. The problem arises from the extensive need for training data in machine learning, which is often based on previous data, such as court decisions. For example, the bail system in the US was found to be alarmingly biased against African Americans. Even efforts to avoid mentioning protected characteristics have been shown to be insufficient, as so-called "fairness through unawareness" has been undermined by proxy characteristics. Addressing the bias of AI systems is a crucial issue for the increased involvement of automated means in judicial settings. The following paper examines various biases that might be introduced in AI-based systems, potential solutions and regulations, and compare possible solutions with current approaches of the European Court of Human Rights (ECtHR) towards biases in human judges. The aim is to confront the question of how to approach current bias in judges compared to approaching bias in future AI-based judicial decision-making technology.

CORRESPONDING AUTHOR:

Andrej Krištofík

Institute of Law and
Technology, Masaryk
University, CZ

andrejkristofik@gmail.com

KEYWORDS:

Bias; artificial intelligence;
automated decision-
making; human rights;
fair trial

TO CITE THIS ARTICLE:

Andrej Krištofík, 'Bias in
AI (Supported) Decision
Making: Old Problems, New
Technologies' (2025) 16(1)
*International Journal for
Court Administration* 3. DOI:
[https://doi.org/10.36745/
ijca.598](https://doi.org/10.36745/ijca.598)

The environment of courts and justice has always been rather conservative,¹ as demonstrated by the judicial symbolism that alludes to (ideals) of ancient Greece. This is also seen, in its relationship to adoption of new technologies, that has arguably been sped up by the COVID-19 pandemic.² The uptake of new technologies has slowly surpassed the “simple” virtualization of the procedures via videoconferencing software,³ and we are already witnessing an adoption of various systems based on artificial intelligence (AI). Besides the general uptake of technology,⁴ AI is also gradually taking on more substantive roles and is moving from simpler administrative tasks,⁵ towards decision-making systems, which is, as of now, being employed only in its “safer” mode, as an assistive or supportive decision-making system.⁶ Besides the technical limitation, the so far limited assistive uptake of such technologies merely in their assistive capacity, has several other reasons, such as social, technical or the issue of responsibility. As early as 2000, Tata has speculated that the place of AI in the judiciary has always been viewed through the optics of cognitive process rather than social, which ignores the role of the judiciary in the society.⁷ Its role is often seen as a *human* institution⁸ delivering an essentially human endeavour as justice.⁹ Therefore the gradual introduction of such technologies provides us with the opportunity to

¹ Dory Reiling, *Technology for Justice. How Information Technology can support Judicial Reform* (Amsterdam University Press, 2009), 74.

² See for example the report on *COVID-19 Continuity* of Court Operations During a Public Health Emergency Workgroup. *Post-pandemic recommendations*. <<https://www.americanbar.org/content/dam/aba/administrative/judicial/2021-az-post-pandemic-rec.pdf>> accessed 02.03.2025; or for specific examples see Alicia L. Bannon, Douglas Keith, ‘Remote Court: Principles for Virtual Proceedings During the COVID-19 Pandemic and Beyond’ (2021) 115 *Northwestern University Law Review* 6.

³ Such “simple” transformation is not digitalisation in its true sense, since digitalisation usually requires appropriate changes to relevant processes, not only simple “employment” of any technology, regardless of how advanced it is. A good example is the concept of online courts which is not satisfied by a mere use of Zoom conference to carry out a hearing but is rather defined by more advanced structural and procedural change, usually allowing for the hearings to be carried out asynchronously via a specialised platforms, or not relying on them at all. See Richard Susskind, *Online courts and future of justice* (Oxford University Press, 2012), 60–1.

⁴ See for example this rather comprehensive survey on Portuguese judiciary: Paula Casaleiro et al., *Judicial Perceptions and Use of Technology. Portuguese Report* (University of Coimbra, 2023). <<https://estudogeral.uc.pt/bitstream/10316/108789/1/Judicial%20perceptions%20and%20use%20of%20technology.pdf>> accessed 02.03.2025.

⁵ Reiling notes that not every decision-making action is the same in terms of its complexity and as such some of them are better candidates for the use of AI technology. Dory Reiling, ‘Courts and Artificial Intelligence’ (2020) 11 *International Journal for Court Administration* 2. A good example of that would be a Czech “Regublina” system that is used for building permits by creating computable zoning laws. Official website of the project available at <<http://89.221.214.206/index.php>> Accessed 02.03.2025.

⁶ To that end see CCJE Opinion No. 26 (2023): *Moving forward: the use of assistive technology in the judiciary*, (Council of Europe, 2023). <<https://rm.coe.int/ccje-opinion-no-26-2023-final/1680adade7>> accessed 03.03.2025, that recommends the use of AI-based technology as a support tool rather than a technology that supersedes the judges.

⁷ Cyrus Tata, ‘Introduction: Why Study Judicial Decision Support Systems?’ (2000) 14 *International Review Of Law Computers & Technology* 3, 279–82.

⁸ James Allsop, ‘Technology And The Future of The Courts’ (2019) 38 *University of Queensland Law Journal* 1 <<https://journal.law.uq.edu.au/index.php/uqlj/article/view/1539/1293>> accessed 02.03.2025.

⁹ Oto Weinberger, *Law, Institutions and Legal Politics* (Springer, 1991), 249–250.

continually bridge the perceived trust gap of entrusting such essentially humane process to a non-human actors.¹⁰

However, at the present moment we may draw on some examples of systems working in such assistive capacity.¹¹ Even though the supportive systems are being used merely by way of example – the conclusions of this article should be applicable even to AI-based systems deployed in a full decision-making capacity (in the future). The following article is thus concerned with AI-based decision-making technologies, which will be conceptualized in the following part of the article. Secondly, it is concerned with potential issues to fair trial rights,¹² namely with the problem of algorithmic bias and its significance for fair (automated) trial. The concept of bias will be introduced based on its technological aspects, as well as from a legal point of view, most notably based on its interpretation under the European Convention on Human Rights (hereinafter as “ECHR”). The article holds that (judicial) bias is not a new phenomenon and should not be treated as such, even when it comes to new technologies, as will be further elaborated. Therefore, the existing framework of the European Court of Human Rights (hereinafter as “ECtHR”) on this matter will be utilized for said analysis. Given that this article is concerned with biased decision-making, to which it will present relevant case-law and theory, we will disregard any issues concerning possible predictive capabilities of judicial AI systems.¹³ Similarly, even though the question of accountability is closely related to the technology relevant for this article, dealing with the complexities of said question is beyond its scope.¹⁴

Based on this analysis, that will draw on both the introduction of the relevant technologies as well as the existing frameworks for biased decision making, the article seeks to answer the following issues: a) Is algorithmic bias problematic for supportive or substantive AI-based judicial technology; b) If so, what are the possible remedies; and c) What are the differences between the requirements for non-biased decision-making for humans and machines. Besides these, the last part of this article will introduce and evaluate some already existing legal instruments that are directed at the issue of bias in AI-based systems (in judicial practice). In general, this article seeks to show that the way in which we think about biased decision-making in “human” judges is, *mutatis mutandis*, applicable to algorithmic decision-making, and as such does not constitute a completely new problem in this regard.

2. THE TECHNOLOGY

There is no shortage of Legaltech startups, promising more effective practice to lawyers via the utilization of any combination of the approaches falling under the

10 Dovile Barysė, Roeë Sarel, ‘Algorithms in the court: does it matter which part of the judicial decision-making is automated?’ (2023) 32 *Artificial Intelligence and Law* 1 <<https://doi.org/10.1007/s10506-022-09343-6>> accessed 03.03.2025.

11 As Contini notes ,that is where most of the academic discussion currently is cantered. See Francesco Contini, ‘Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings’ (2020) 2 *Law, Technology and Humans* 1.

12 For the potential impact of the use of AI based technologies on the fair trial see for example Dory Reiling, ‘Courts and Artificial Intelligence’.

13 For such use, as well as other possible roles of AI in the court see *Ibid*.

14 As per this issue see for example Francesco Hanafi, Antonio Cordella, ‘Artificial Intelligence and Decision-Making: the question of Accountability’ *Proceedings of the 54th Hawaii International Conference on System Sciences* (2021).

umbrella of artificial intelligence. As private practices of law firms are more interesting from the business perspective of the developers of such automated tools, we have seen a great deal of developments in the field of lawyer-assisting AI systems,¹⁵ for example for a quick extraction of the relevant pieces of information from large quanta of documents.¹⁶ More advanced examples of AI-based legal technology are different analytical systems, most notably a case prediction software, such as LexMachina.¹⁷ Such “analytical” tools thread a very narrow line, as was shown by France, that has banned case prediction due to the fears of possible interference with the right to a fair trial.¹⁸ Publication of such predictions now carries a hefty fine or even a prison sentence.¹⁹

AI based legal tech is becoming prevalent in the litigants’ sector,²⁰ however the courts are not staying behind. Besides some of the aforementioned technologies, there are other technologies specifically targeted at courts and their decision-making processes. By their nature these systems pose much higher risk for fundamental rights, as they stand to influence²¹ – if not make – the binding decisions about the relevant rights and obligations.

It’s nearly impossible to deal with the issue of algorithmic bias and not mention COMPAS,²² which has since 2016 became almost synonymous with this issue.²³ This system is utilized by courts in several states of the USA for risk assessment. Based on input data that are mostly gathered via a questionnaire, the system based on non-public proprietary algorithm computes a “risk score” ranging from 0 to 9.9, that should

15 Julia Sobowale, ‘How Artificial Intelligence is changing the legal profession’ (2016) *ABA Journal* <https://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession> accessed 02.03.2025.

16 <<https://kirasystems.com/how-kira-works/>> accessed 02.03.2025.

17 <<https://lexmachina.com/>> accessed 02.03.2025.

18 Such ban has been part of the criminal code novelisation and introduces a possible criminal penalty for machine-based profiling of the judges, that can subsequently be used for case prediction. Such is seen as a threat to fair trial namely due to strong French emphasis on deciding cases solely and strictly based on law and facts at hand. For further analysis see Andrej Krištofik, ‘Európska a česká perspektiva francúzskeho zákazu profilovania sudcov’ (2022) 4 *Trestněprávní revue – odborný recenzovaný časopis pro trestní právo*, 239–245.

19 Malcolm Langford, Mikael R. Madsen, ‘France Criminalises Research on Judges’ (2019) *Verfuassunblog* <<https://verfassungsblog.de/france-criminalises-research-on-judges/>> accessed 02.03.2025.

20 For a general overview see Jana Soukupová, ‘AI-based Legal Technology: A Critical Assessment of the Current Use of Artificial Intelligence in Legal Practice’ (2021) 15 *Masaryk University Journal of Law and Technology* 2.

21 There reportedly exists a strong tendency to less critically rely on machine recommendations, see for example Cass R. Sunstein, ‘Algorithms, Correcting Biases’ (2019) 86 *Social Research: An International Quarterly* 2 or similarly in context of COMPAS and assistive systems see Gualdi, ‘Artificial Intelligence and Decision-Making: the question of Accountability’.

22 <<https://www.equivant.com/northpointe-risk-need-assessments/>> accessed 02.03.2025.

23 Julia Angwin, et al., ‘Machine Bias’ (2016) *ProPublica* <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, accessed 02.03.2025.

indicate to the judge how likely is the person to re-offend.²⁴ This risk score is only a partial factor, based on which the judge makes her decision, therefore this system *should* play only an assistive role in the whole decision-making process. This becomes problematic when the judge relies solely on the COMPAS score, which ultimately in itself goes against the right to a fair trial, as has been for example stated by the Supreme court of Wisconsin in the case of *State v. Loomis*.²⁵ This sole reliance on the COMPAS score becomes even more problematic when we consider that the logic of this system is not made public due to “trade secret”²⁶ protection. This is especially problematic when we consider its reported bias against Afro-American population. This issue will be described more in the following section, but it represents the main reason for choosing this system as an example.

The example that showcases the whole spectrum of the possible application, and one where such potential threats to the right to a fair trial are more pressing, is the case of full judicial decision-making automation, or the so-called “robo-judges”.²⁷ Full automation of judicial decision-making is still in its infancy, however that is precisely the moment we should start tackling the potential problems it may introduce.²⁸ We can currently find two examples, even though their relevance and actual state is questionable. One example, that has received a lot of attention,²⁹ even though it was gravely misrepresented in many cases – is the case of Estonian robo-judges. Ministry of Justice of Estonia has however objected³⁰ to their system being denoted as such, since in this case it was rather simplistic “process optimalization” as opposed to some advanced judicial decision-making system. The system deployed by Estonia was used to deal with simple cases, where no actual judgment was required, and the claims were “just” rejected for lack of standing etc. However, even this deserves a proper attention and is not to be dismissed nor diminished, since this system has already put the human out of the loop and its decisions were binding.

24 <<https://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf>>.

25 “*State v. Loomis*” (2016) 130 *Harvard Law Review* 1530 <<https://harvardlawreview.org/print/vol-130/state-v-loomis/>> accessed 03.03.2025.

26 Tangential problem is the fact that it leaves us in the dark as to which factors are relevant and whether their consideration are not opposed to the right to a fair trial as well, since this right also constitutes a right to sentence based only on relevant facts and applicable law. As from the ProPublica 2016 study it seems that besides else one of the considered factors was for how long the defendant resided in the state of Wisconsin prior to committing the offence. See Angwin, ‘Machine Bias’.

27 See overview at Tanel Kerikmäe et al., ‘Frontiers in AI Judiciary: A Contribution to Legal Futurology’ (2023) 11 *Acta Baltica Historiae et Philosophiae Scientiarum* 2.

28 Mark D. Fenwick et al., ‘Regulation Tomorrow: What Happens When Technology Is Faster than the Law?’ (2017) 6 *American University Business Law Review* 3. <<https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1028&context=aubl>> Accessed 02.03.2025.

29 Especially after this, misguided at best, Wired article from 2019 Eric Niller, ‘Can AI Be a Fair Judge in Court? Estonia Thinks So’ (2019) *WIRED* <<https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>> accessed 02.03.2025.

30 <<https://www.just.ee/en/news/estonia-does-not-develop-ai-judge>> Accessed 02.03.2025.

The other, supposedly³¹ more advanced, example is the case of the Chinese robo-judges³² and even prosecutors.³³ China is moving more towards heavily digitalized justice system with deployment of “one-stop-shop” judicial apps, that enable the user to file motions, provide testimonies and monitor the progress of their case.³⁴ More importantly, several provinces are using automated systems to analyse relevant data (similarly to COMPAS – but on a more complex scale) and based on this, draft a judgment that is, depending on the province, either binding or subject to a human approval.³⁵ Currently such systems deal mostly with small claims³⁶ and misdemeanours or traffic violations. Some provinces are reportedly attempting to deploy such systems in criminal justice as well.³⁷ This should be however mostly in an assistive capacity and primarily used for sentencing, with the argument that sentencing greatly varies throughout the China and this should help to standardize it.³⁸

These examples describe basic models of relevant implementation of AI-based systems in judicial decision-making, that should to a certain degree reflect also a possible future development. In the following impact analysis of algorithmic bias, the role of AI-based systems could be double. It could either be in an assistive capacity, meaning that this system is not fully autonomous and (substantially)³⁹

31 The word supposedly is here to signify the nature of information we have about these systems. There are almost no relevant insider academic articles on this issue and “popular” articles need to be taken with a grain of salt as we have seen that they tend to be more sensational and misleading even with such countries like Estonia to which we have better “informational access” than China. We can be quite sure China is using such automated judicial systems, their precise capabilities and scopes are however bit more unsure, given the governmental control of the information flow and their interest on “keeping up the appearances” with technological advancements.

32 Straton Papagianneas, Nino Junius, ‘Fairness and justice through automation in China’s smart courts’ (2023) 51 *Computer Law & Security Review* 1 <<https://www.sciencedirect.com/science/article/pii/S0267364923001073>> Accessed 02.03.2025, for some statistical overview see “AI’s best use case may actually be in our judicial system”, *IASToppers* (2023). <<https://www.iastoppers.com/articles/ai-s-best-use-case-may-actually-be-in-our-judicial-system>> [Accessed 02.03.2025].

33 Niu W. Tian, Michael Yuan, ‘Intelligent Justice’: AI Implementations in China’s Legal Systems’ in *Artificial Intelligence and its discontents*, ed. Ariane Haneemayer (Springer, 2022): 197–222.

34 Straton Papagianneas, ‘Automating Intervention in Chinese Justice: Smart Courts and Supervision Reform’ (2023) 10 *Asian Journal of Law and Society* 3. <https://doi.org/10.1017/als.2023.5>.

35 Shi Changqing et. al., ‘The Smart Court – A New Pathway to Justice in China?’ (2021) 12 *International Journal for Court Administration* 1 <https://doi.org/10.36745/ijca.367>.

36 The area of small claims is, due to its low complexity and low stakes, popular test ground for new judicial technology, as was the case for example with pilot programs of online courts, see for example Canadian Civil Resolution Tribunal at <<https://civilresolutionbc.ca/>> accessed 02.03.2025.

37 Jiahui Shi, ‘Artificial Intelligence, Algorithms and Sentencing in Chinese Criminal Justice: Problems and Solutions’ (2022) 33 *Criminal Law Forum*. 10.1007/s10609-022-09437-5.

38 *Ibid.*; Tian, ‘Intelligent Justice’: AI Implementations in China’s Legal Systems’. In general, digitalization in the realm of justice is often deployed with the aim of providing a greater legal certainty, see. Gualdi, ‘Artificial Intelligence and Decision-Making: the question of Accountability’.

39 The word substantially is used here in helpful reference to GDPR and its article 22, that points to an important issue of “rubberstamping” i.e. using automated means but for some reason still keeping the human aspect, which does not provide any relevant consideration or judgement, but stands in between the algorithm and decision to either formally fulfil some requirements or to provide “someone to blame”, such process is not considered substantial.

requires a human-in-the-loop. The other implementation considered by this article is full automation, where the AI-based system, after an input from users,⁴⁰ carries out the legal analysis and then renders an enforceable judgment. This judgment is ideally binding on its own, but may subsequently require a confirmation by a human judge.⁴¹ While consideration of such fully automated systems seems more pressing due to their possibility of greater threat to fundamental rights, it is generally a good idea to consider also assistive systems, for they might as well pose such risk, as we will demonstrate on the example of COMPAS, and they are currently being used to a greater extent. What also can't be underestimated is the fact that fully automated means tend to be deployed firstly at their semi-automated capacity⁴² which was, for example, the case of Chinese Xiao Zhi 3.0⁴³ system that started in 2019 as a mere assistive system before being deployed as fully automated some years later.

Having mapped out a general understanding as to what kind of systems we are discussing, we will move to the next section which will introduce the issue of bias from both technical and legal perspective.

3. THE BIAS

The issue of bias is an area with a great focus from both law⁴⁴ and computer science. So much so that in the area of artificial intelligence it forms almost its own sub-area of study.⁴⁵ As such, this article cannot give a comprehensive overview of the technical aspect of this problem but will describe it to the necessary extent. On the legal side the issue of bias, does not only constitute one of the rights under the umbrella term

⁴⁰ The issue of handling user inputs, that are necessary for the decision-making process is complex and almost separate issue, even in relation to bias, and as such is out of the scope of this article. For the analysis of the impact of user inputs on bias, or more broadly on the need of properly handling this phase of the process see for example Jakob Mokander et al., 'Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations' (2021) 27 *Science and Engineering Ethics* 44. <https://doi.org/10.1007/s11948-021-00319-4>.

⁴¹ Generally, the substantiveness of the role of a human in confirming a machine-made decision is questionable due to a certain cognitive shortcut or bias where humans tend to less critically accept machine-rendered recommendations. For this argumentation, as well as other adjacent reasons see "State v. Loomis" or similarly see Christoph Engel et al., 'Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism' *Artificial Intelligence and Law* (2024). <https://doi.org/10.1007/s10506-024-09389>.

⁴² See for example Simona Demková, 'The Decisional Value of Information in Semi-Automated Decision-Making' (2021) 2 *Review of European Administrative Law* 14 <<https://realaw.blog/2021/11/19/the-decisional-value-of-information-in-semi-automated-decision-making-by-simona-demkova/>> accessed 02.03.2025.

⁴³ Benjamin M. Chen, Zhiu Li, 'How Will Technology Change The Face Of Chinese Justice?' (2021) 34 *Columbia Journal Of Asian Law* 1.

⁴⁴ Especially in the context of the right to a fair trial one may find the term "impartiality" instead of bias. But given that the term bias is more commonly used when describing the relevant issue in algorithms, we will use the term bias, instead of impartiality. The right to a non-biased court is therefor the right to an impartial tribunal.

⁴⁵ Michael Rovatsos, et al., 'Landscape Summary: Bias in Algorithmic Decision-Making' (2019) *Centre for Data Ethics & Innovation* <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf> accessed 02.03.2025.

of the right to a fair trial,⁴⁶ but there exists a considerable amount of case law⁴⁷ and theoretical works, hinting at the importance of this right. The following section introduces this issue from both angles.

3.1. TECHNICAL PERSPECTIVE

Generally, the issue of algorithmic bias may have two causes.⁴⁸ Firstly it is the need for (historical) training data, which may themselves be biased in various ways⁴⁹ and secondly the bias may be introduced by an improper design of the decision making algorithm itself.⁵⁰ This issue is often amplified by the biased data, in a vicious circle, since many machine learning approaches are creating their own algorithms based on the datasets in which they are trying to identify and then recreate patterns.⁵¹ It is then not that the machine learning or algorithms as such would create or introduce a new forms of biases, however the issue is also more serious than a mere recreation of existing “human” biases.⁵² The grave issue here is the effect of bias amplification,⁵³ which is introduced by taking the past biased decisions, recognizing their pattern as the norm and subsequently applying it at much greater, automated, scale than would previously be possible. An example of this is the now scrapped Amazon hiring tool,⁵⁴ that was supposed to review submitted applications and do a preliminary filtering of unfit applicants. Since Amazon did not have many women in the executive roles, the algorithms recognized this pattern and figured out that being a woman disqualifies the applicant from seeking this position. It then went on to throw out all female

46 See for example the analysis of Article 6 by the ECtHR at <https://www.echr.coe.int/documents/d/echr/guide_art_6_criminal_eng> accessed 02.03.2025.

47 Ibid., or see for example ‘Independence and Impartiality of the Judiciary An overview of relevant jurisprudence of the European Court of Human Rights’ (2021) *The Aire Centre*, 112 et seq. <<https://www.rolplatform.org/wp-content/uploads/2021/09/independence-and-impartiality-of-judiciary-eng.pdf>> accessed 02.03.2025.

48 For much broader overview of this issue see Nima Kordzadeh, Maryam Ghasemaghahi, ‘Algorithmic bias: review, synthesis, and future research directions’ (2021) 31 *European Journal of Information Systems* 3.

49 This issue is again too complex for the scope of this article, however the bias can be introduced into data in several ways, for example by reflecting already flawed process, such was the case of COMPAS, as noted for example in Angwin et al, “Machine Bias”, or by selectively including only some portion of the relevant population (i.e. choosing to collect data only in some neighbourhoods or from specific mobile carriers, see A. Alexander Amini et al., ‘The impact of social segregation on human mobility in developing and industrialized regions’ (2014) 3 *EPJ Data Science* 6 <https://doi.org/10.1140/epjds31>), or at the following stage of data preparation and normalisation (Bruijne, Kars de, and Erwin van Veen. ‘Mind the Frame: Bias at the Data-Analysis Stage.’ *Pride and Prejudice: Addressing Bias in the Analysis of Political Violence* (Clingendael Institute, 2017). <<http://www.jstor.org/stable/resrep17336.5>> accessed 02.03.2025). Or similarly see Reiling, ‘Courts and Artificial Intelligence’, that argues that proper data is the absolutely necessary (but not sufficient) condition for the use of judicial AI.

50 Carey K. Morewedge, et al., ‘Human bias in algorithm design’ (2023) 7 *Nature Human Behaviour* 7 <https://doi.org/10.1038/s41562-023-01724-4>.

51 Gareth James, et al., ‘Unsupervised Learning’ in: *An Introduction to Statistical Learning*. Springer Texts in Statistics, ed. Gareth James (Springer, 2023): 503–556.

52 Andrea L. Roth, ‘Trial by Machine’ (2016) 104 *Georgetown Law Journal* 5.

53 Alice Xiang, ‘Mirror, Mirror, on the Wall, Who’s the Fairest of Them All?’ (2024) 153 *Daedalus* 1 https://doi.org/10.1162/daed_a_02058.

54 Elham Albaroudi et al., ‘A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring’ (2024) 5 *AI* 1.

applicant's CVs during the hiring process to such roles.⁵⁵ The bias has already existed in the hiring structures of Amazon, however it was amplified by the automation.⁵⁶

This is a rather basic overview of the bias issue from the technical side,⁵⁷ which does not exhaust the interesting, and even legally important, analysis offered by computer science. Further, it does not offer a mere description of the problem but suggests a solution as well. The seemingly obvious answer to this issue is to simply not consider, thusly not provide, information that could lead to bias (the "protected characteristics"), such as race or gender. It has however been repeatedly demonstrated that this does not work, since machine learning algorithms may utilize "proxy characteristics" as a stand-in for the unavailable protected characteristics.⁵⁸ This can come from such simple proxies like name, where certain names would be more typical for one gender, to more complex proxies such as an address or the age at first arrest as proxies for race, (an actual example from the COMPAS system),⁵⁹ which is a situation that slowly merges both technical and legal perspectives.

3.1.1. Fairness through ignorance and the COMPAS example

Researching the COMPAS system is particularly hard due to the fact that the algorithm itself is protected by Northpointe, the producer of this system, as a business secret.⁶⁰ In 2016 a non-profit organization ProPublica attempted to at least evaluate the reliability of the system by obtaining the risk scores of over 7,000 people in one county, awarded over 2 years.⁶¹ They then compared the risk scores with an actual future criminal record, leaving the timeframe of 3 years for any future offence as per the Northpointe benchmarks, concluding that only 20% of the people that were judged

55 Jeffrey Dustin, 'Amazon scraps secret AI recruiting tool that showed bias against women' (2018) *Reuters* <<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>> accessed 02.03.2025.

56 Such issue is not uncommon in HR algorithms, see Alina Köchling, Marius 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' (2020) 13 *Business Research* <https://doi.org/10.1007/s40685-020-00134-w>.

57 The author sees a potential for bias being introduced into the decision-making algorithms by their designers subconsciously during the process of creation, this would however call for more complex ideological analysis that as of yet does not have much attention in existing research. Proposed remedies should, however, apply to any kind of bias, regardless of how it was introduced. As per the issue of the design see for example "The subject who encodes the legal rules into the software code is making the actual interpretation of the law." In Contini, 'Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings'.

58 Matt J. Kusner et al., 'Counterfactual Fairness' in *Advances in Neural Information Processing Systems*, eds. I. Guyon et al. (NIPS, 2017): 4066–76. <<http://papers.nips.cc/paper/6995-counterfactual-fairness.pdf>> accessed 02.03.2025.

59 Cynthia Rudin et al., 'The Age of Secrecy and Unfairness in Recidivism Prediction' (2020) 2 *Harvard Data Science Review* 1 <https://doi.org/10.1162/99608f92.6ed64b30>.

60 Alyssa M. Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms' (2017) 103 *Iowa Law Review* 1 and Taylor R. Moore, 'Trade Secrets and Algorithms as Barriers to Social Justice' (2017) *Center for Democracy and Technology* <<https://cdt.org/wp-content/uploads/2017/08/2017-07-31-Trade-Secret-Algorithms-as-Barriers-to-Social-Justice.pdf>> accessed 02.03.2025.

61 James Larson, 'How We Analyzed the COMPAS Recidivism Algorithm' (2016) *ProPublica* <<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>> accessed 02.03.2025.

as high-risk reoffenders committed any further crimes.⁶² Besides this unreliability, the study has highlighted a different issue – that the system was prevalently wrong in the high-risk cases. The system could however formally show a higher reliability, since the low-risk offenders were usually low risk as judged by the system.⁶³ Which directly leads to the last point of the ProPublica study and that is the observation that Afro-American population was more likely to be awarded a high-risk score and therefore to be wrongly placed amongst the high-risk reoffenders. ProPublica further points to the questionnaire that is used for input data that contains questions about parents being in jail or illegal drugs being distributed in the neighbourhood, disadvantaging people from low-income areas and households.⁶⁴ Therefore, even without the protected characteristics being disclosed, the system still operated with them via a proxy characteristics. The issue of the questionnaire points also to a different origin of algorithmic bias that is often (mis)used in the analysis of these algorithms. For example, in this case one could argue that the age at first arrest or the fact that one of the parents was in jail points to either personal failings of the defendant or failings in the upbringing that are likely to make him more susceptible to commit (further) crimes. However, such an argument requires more broad critical evaluation of different societal phenomena such as over-policing of certain areas or minorities,⁶⁵ or so called broken windows policing approach.⁶⁶ As Yeung points out this kind of bias is not the “fault” of algorithms as they more or less properly encode the data,⁶⁷ the issue rather lies in historic injustices that were suffered by a certain group which will now be encoded by the algorithm relying on the past data revictimizing the disadvantaged group and deepening the patterns of past oppressions.⁶⁸

As per the solutions to this problem offered by computer science, they, to a certain degree, mirror the current legal recommendations, which is good news, since in the end the solution will require the employment of a certain technical solution based on a legal act. Those solutions usually rely on better monitoring of the presence of bias, via methods such as *counterfactuality* that, simply put, applies the idea that a decision is fair (or non-biased) if the automated decision is the same in a) a real world, as well as b) “counterfactual world” where the defendant belongs to a different demographic, i.e. the potentially discriminating characteristics are different. The other direct yet more tiresome solution is to monitor the training data for bias and

⁶² Julia Angwin, ‘Machine Bias’.

⁶³ See the analysis of this “statisticians fight” in Brian Christian, *Alignment Problem. Machine Learning and Human Values* (W.W. Norton, 2020): 67–79.

⁶⁴ See sample questionnaire used by COMPAS at <<https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE>> accessed 02.03.2025.

⁶⁵ Michael W. Sances, Hye Y. You, ‘Who Pays for Government? Descriptive Representation and Exploitative Revenue Sources’ (2017) 79 *Journal of Politics* 3.

⁶⁶ Joshua C. Hinkle, David Weisburd, ‘The irony of broken windows policing: A micro-place study of the relationship between disorder, focused police crackdowns and fear of crime’ (2008) 36 *Journal of Criminal Justice* 6.

⁶⁷ There currently exists a debate on “mathematical inevitability” of this, as an effect of trade-off between accuracy and non-discrimination. Author believes that this debate, for most parts ignores the socio-historic injustices included in the existing data. See for example Alexandra Chouldechova, ‘A fair prediction with a disparate impact: a study on bias in recidivism prediction instruments’ (2016) 5 *Big Data* 2.

⁶⁸ Karen Yeung, ‘Why Worry about decision-making by machine?’, in *Algorithmic regulation*, ed. Karen Yeung (Wiley, 2017): 32–33.

correct it during the design process.⁶⁹ Close monitoring however remains presently the most viable solution and a first step in the mitigation process. For the purpose of monitoring, it seems to be fitting to utilize the framework established by the ECtHR, which will be introduced in the following section, since it has been established for the purpose of protecting and enabling non-biased judicial decision-making. And as this article argues it should be used for any decision-making, regardless of the way in which it is carried out.

3.2. LEGAL PERSPECTIVE

The right to a non-biased court (decision) is a part of the umbrella right to a fair trial and is almost exclusively interpreted jointly with the right to an independent court.⁷⁰ As much as they are described as an external and internal expression of the same principle, this article focuses solely on the issue of bias.⁷¹ In the existing case law of the ECtHR there are two levels at which bias in judicial decision making can appear, and there are two associated tests for it – internal and external. The internal aspect pertains to any pre-existing convictions or beliefs of the judge that could skew her decision-making in some particular way. The external aspect of it is rather closely tied to the requirement of independence, judging whether the personal composition of tribunal offers sufficient safeguards for non-biased decision-making. Lastly, there also exists the requirement of appearance, meaning that the tribunal must not only be non-biased, it also must appear to be so.⁷² A much simplified, yet rather helpful, description of this requirement is also provided by Bingham, who describes the goal of this requirement (together with the requirement for the independence of tribunal) as the guarantee that the case will be decided (by a judge) based on her best conscious, relevant facts and applicable law only.⁷³

Bias is an internal disposition of the judge and as such it is hard to determine. That is the reason why the ECtHR has focused almost exclusively on the objective test of bias.⁷⁴ As such, one of the basic factors that can determine judges internal disposition is the behaviour towards parties, that can be seen as hostile or biased,⁷⁵ which could be, for example, demonstrated by the language the judge uses.⁷⁶ It is worth noting

69 Such approach, of proper data handling and processing, is set forth by many legal instruments, discussed in section 4, for example by the Canadian Directive on Automated Decision-making or by the European Artificial Intelligence Act, both mentioned below.

70 David Harris, et al., *Law of the European Convention on Human Rights* (Oxford University Press, 2023): 284, or Ramos Nunes de Carvalho e Sá v. Portugal [GC], ECtHR Judgment 6.11.2018, §§ 150 and 152.

71 Which is not to say that the right to an independent tribunal is not problematic for the concept of algorithmic decision making, quite the opposite, and thus is out of the scope of this article. This right needs to be more closely focused on with regards to independence on other public powers, most notably in the situation of regulating and deploying such systems based on law, or even slightly pushing this concept further and considering the independence on the (private) creator of the system. This last point has been marginally raised in the ProPublica study as well.

72 Or the “justice must not only be done, it must also appear to be done” principle as defined by the ECtHR for example in *Micallef v. Malta* [GC], ECtHR Judgment 15.10.2009, § 98.

73 Tom Bingham, *Rule of Law* (Penguin Books, 2011): 92.

74 *Micallef v. Malta*, § 146.

75 *Buscemi v. Italy*, ECtHR Judgement 16.09.1999, §§ 67–68.

76 *Vardanyan and Nanushyan v. Armenia*, ECtHR Judgement 25.10.2016, § 82.

that there exists a certain threshold for this, since it as has been stated by the ECtHR, justice needs to keep a “human face” and a mere expressions of understanding or sympathy towards the aggrieved party are not sufficient.⁷⁷ Judges conduct is therefore the only way in which we may ascertain his or hers bias in the given case by the subjective test. The objective test offers more options in this regard.

The objective test deals with the available safeguards guaranteeing the impartiality of the tribunal. Further, it is also concerned with the perception of such by the parties, however, this factor plays only a minor role.⁷⁸ The objective test would be, for example, failed by a judge who has made public statements about the matter that were perceived as biased.⁷⁹ Another example is when the judge has already presided over the same case in a different branch or judicial instance, as those are objectively provable instances of their “internal” states (i.e. that the judge has already “made up their mind”).⁸⁰ Objective test may be also failed in case of any financial or personal link between the judge and one of the parties,⁸¹ or if there exists a personal interest of the judge relevant to the case.⁸²

Drawing on the bias-related knowledge from the two relevant fields for algorithmic decision-making, computer science and law, we can conclude two important observations, which will be further dealt with in the following section.

Firstly, borrowing from the computer science, we have observed that mitigating bias requires its identification and monitoring, which can be done by carrying out tests – monitoring the decision-making outcomes for any negative patterns. Secondly, looking at the legal side, we have observed that there exists a framework, established by the ECtHR for identifying biased decision making. It has two parts – subjective and objective test. The subjective test is rather complicated to perform due to its nature and as such not often used and it even may seem irrelevant for algorithmic decision-making, since it is aimed at “the personal conviction and behaviour of a particular judge, that is, whether the judge held any personal prejudice or bias”.⁸³ While it may be argued that it is impossible for an algorithm to possess such anthropic quality as personal conviction and display a certain behaviour, on the other hand this is where the *mutatis mutandis* part of the suitability comes to front. Within computer science, an “artificial intelligence behaviour” field of study is emerging⁸⁴ that is dedicated to investigating “thought-like” inferences on which AI systems are basing their decisions. In this vein then, the “personal conviction and behaviour” could be taken to mean the internal specifics of the algorithm. In the same way that the fault in human judgement

⁷⁷ Karrar v. Belgium, ECtHR Judgement 31.08.2021.

⁷⁸ It is worth noting that subjective and objective aspects of impartiality are essentially connected and often the subjective aspect can be demonstrated via the objective aspect, therefor making “the appearance” a relevant aspect. See for example Denisov v. Ukraine namely § 62, or Micallef v. Malta [GC], 2009, § 98.

⁷⁹ Rustavi 2 Broadcasting Company Ltd and Others v. Georgia, ECtHR Judgement 9.12.2019, §§ 341–342.

⁸⁰ Pasquini v. San Marino, ECtHR Judgement 08.03.2021, § 148.

⁸¹ Pescador Valero v. Spain, ECtHR Judgement 17.06.2003, § 27.

⁸² Gautrin and Others v. France, ECtHR Judgement 25.10.1998, § 59.

⁸³ Meznaric v Croatia, § 29.

⁸⁴ Tore Pedersen, Christian Johansen, ‘Behavioural artificial intelligence: an agenda for systematic empirical studies of artificial inference’ (2020) 35 *AI & Society*. <https://doi.org/10.1007/s00146-019-00928-5>.

arises as a result of a “personal conviction”, the fault in automated decision is arising as a result of “internal specifics of the algorithm”.⁸⁵ Fortunately, it is not necessary for us to attempt to draw any comparisons between the human and AI “thinking”,⁸⁶ since this is where the framework established by the ECtHR becomes useful once again with the second, objective, test. This test implores us to examine external factors that could point to possible (internal) bias,⁸⁷ as those are interconnected.⁸⁸ More precisely, it asks us whether such necessary steps have been taken, to ensure that the decision-making is not biased, particularly whether the composition of tribunal offers sufficient safeguards in this regard.

Similarly to the subjective test, the objective test provides algorithmic justice with a suitable framework for testing the bias in decision-making process, with the necessary adaptations to “non-human decision makers” since considering financial or personal ties makes no sense in this case. What should concern us is the general test of whether the court “offered sufficient guarantees to exclude any legitimate doubt in this respect”.⁸⁹ Given the unstable⁹⁰ nature of some⁹¹ machine learning models, iterative audits performed during the life cycle of such system, which are described in the following section, are to be viewed as *mutatis mutandis* fulfilling the requirements established by the two tests of ECtHR, since such test can discern a presence of potentially negative patterns in the past “actions” of the algorithm. It further fulfils the objective test by offering guarantees in regard to the (un)biased decision, that still respects the evolutive and black box⁹² nature of machine learning based algorithm.⁹³ The deployment of such test, and its fulfilment of the Article 6 requirements, will be examined further in the context of existing legal documents in the following section.

⁸⁵ *Ibid.*

⁸⁶ Hanbo Xie, ‘The promising future of cognitive science and artificial intelligence’ (2023) 202 *Nature Reviews Psychology* 2. <https://doi.org/10.1038/s44159-023-00170-3>.

⁸⁷ *Ugulava v Georgia* no. 2, § 50.

⁸⁸ *Ramos Nunes de Carvalho e Sá v. Portugal* [GC], 2018, § 145.

⁸⁹ *CASE OF KYPRIANOU v. CYPRUS* (Application no. 73797/01) § 118.

⁹⁰ Complex overview of the issue of instability of the machine learning models as it relates to judicial technology can be found in Contini, ‘Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings’.

⁹¹ This is true of most basic Machine Learning models and approaches that are currently popular, after all their dynamicity is often one of the key points for their use. However, there are more recent advances in the field of Machine Learning such as Ensemble Methods or approaches of deploying Machine Learning methods within expert systems that are limiting the scope of the uncertain/developing decision making algorithm. See for example Samyar Janatian, et al., ‘From Text to Structure: Using Large Language Models to Support the Development of Legal Expert Systems’, in *JURIX 2023: The Thirty-sixth Annual Conference*, eds. Giovanni Sielno et al. (IOS, 2023): 379. <<https://ebooks.iospress.nl/volumearticle/65584>> accessed 02.03.2025; Further it is worth keeping in mind that any such machine learning model must be designed as continuous in order to present us with this problem, since standard machine learning models do not change, unless provided with new data and instructed to do so, once they have finished the model training phase. Liyuan Wang et al., ‘A Comprehensive Survey of Continual Learning: Theory, Method and Application’ (2024) 46 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10.1109/TPAMI.2024.3367329.

⁹² Bartosz Brożek, et al., ‘The black box problem revisited. Real and imaginary challenges for automated legal decision making’ (2024) 32 *Artificial Intelligence and Law*.

⁹³ Great example of this would be the Canadian Directive on Automated decision making, analysed below, that asks the deployer to perform the iterative audit over the output data, which can be generated for set of synthetic data but it also does not require us to “open” the black box.

Besides the research questions stipulated in the Introduction part of this article, its overall goal is to demonstrate our “preparedness” for the implementation of automated decision making, at least to the extent to which it concerns bias. This thesis is essentially alluded to in the titles’ “old problems” and it points to the fact that problem of biased decision making is nothing new, unknown nor underdeveloped in the case law. Such has been demonstrated in the previous sections by analysing the issue of bias as it pertains to the right to fair trial, as well as by analysing the general applicability of currently established tools, such as the objective and subjective tests of ECtHR. However, the question of whether our basic understanding requires at least some reinterpretation in the face of the “new technology”, remains.

4.1 CURRENT REGULATORY LANDSCAPE

There are currently several regulatory frameworks, attempting to address this issue,⁹⁴ with varying degree of their legal force. One of the earlier frameworks, focusing on judicial use and problems relevant to this article, is the European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment⁹⁵ published in 2018 by the European Commission for the Efficiency of Justice (the “CEPEJ”). The Charter proposes 5 principles for the ethical use of AI in judiciary, such as transparency and respect for fundamental rights. Further in the Charter, the issue of non-biased decision-making can be found under the general respect for fundamental rights. This in turn should encompass the right to a fair trial which, as has been described in the third section of this article, should in turn encompass the right to an impartial decision. Such an analysis is however not strictly necessary, since one of the five principles expressly refers to the need for ensuring non-discrimination. Tangentially, there is also a principle of quality (of input data) which, if implemented properly, mitigates the potential for bias, insofar as it pertains to training data-emergent bias.⁹⁶ Further, the fourth principle encompasses transparency, *impartiality* and fairness. Thusly, it appears that CEPEJ sees this issue as deserving a particular attention, as it is present in three out of five principles.

Interestingly, CEPEJ does not call for direct ban on use of sensitive data, such as data about race, but emphasizes the need for “particular care” when utilizing them. This issue as such however does not get further, or explicit, attention, despite the fact that the potential use and even availability of the protected characteristics seems to be a rather contentious issue. So much so that a mere lack of explicit ban on the processing of such information should not be taken as an argument for their processing. However, from the remaining principles presented by the CEPEJ, we may draw exactly

⁹⁴ Crofts and Rijswijk argue that the attention of these regulation is focused on particular issues, such as privacy, and not on the structural problems like bias. The author agrees with the general sentiment expressed in their work however the issue is not as much a lack of regulation, but rather lack of understanding of the problem resulting in a proper regulation of this problem. See Penny Crofts, Honni Rijswijk, *Technology: New trajectories in law* (Routledge, 2021): 86.

⁹⁵ See report of the European Commission For The Efficiency Of Justice (CEPEJ) on European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment. Available at <https://www.unodc.org/res/ji/import/regional_standards/ethical_charter/ethical_charter.pdf> accessed 02.03.2025.

⁹⁶ Ibid. Or see above the various ways in which bias might be introduced into the decision making algorithm.

this conclusion. More precisely, similarly to scholars such as Cofone, it would seem that by their insistence on the diversity in development as well as rather robust monitoring requirements, they see the utility of these characteristics in their potential for combating and identifying these negative patterns in 'real world' deployment of such systems.⁹⁷ Under the principle of non-discrimination it calls for a subsequent monitoring and mitigation approach. The principle of impartiality should be fulfilled by using diverse datasets and by a regular monitoring of such systems by a public body. The Appendix deals with this issue firstly separately, in relation to assistive technology. Namely technologies of asset dividers,⁹⁸ where it limits itself to a warning of potential bias and establishes that the use of (these) technologies must not deprive a person of access to court. Lastly, the issue of potential bias in the AI-based tools is mentioned in the Appendix as a base for the need to set up ethical frameworks and hold an in-depth public discussion on these issues. The need for public discussion refers to the need to keep "in-the-loop" the potentially discriminated part of the population during the implementation and design phases. The need for the establishment of an ethical framework is explained by problematic adoption, and mostly enforcement of, a legal framework in the digital, thus supranational, environment.

The Charter also presents some basic remedies for algorithmic bias, such as certification schemes or subsequent control and mitigation processes. It also stipulates the need to keep human-in-the-loop at some level, i.e. in the second judicial instance or similar, which is already familiar requirement to us, for example from GDPR.⁹⁹ Such a tool – of human oversight – serves as the possibility of obtaining a human intervention that could and should rectify the possible incorrect decision that has been based on tainted data or improper algorithmic design. As we will see, those recommendations are mostly reiterated in other documents.

There currently does not exist a binding legal document that would deal with the issue of judicial use of artificial intelligence directly. However, we have the now final European Artificial Intelligence Act (hereinafter the "AI Act")¹⁰⁰ AI white paper, as well as the preparatory works produced by the High-Level Expert Group on Artificial

97 Ignacio N. Cofone, 'Algorithmic Discrimination Is an Information Problem' (2019) 70 *Hastings Law Journal* 6. As has been argued by the critics of 'cleaning' the data of such characteristics, their lack makes it nearly impossible to identify the rise of negative decision-making patterns, based on proxy characteristics of them McKane Andrus et al., 'What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness', *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (ACM, 2021)*: 249–260. <<https://dl.acm.org/doi/10.1145/3442188.3445888>> accessed 03.03.2025.

98 Such as Family_Winner or SmartSettleOne, see John Zeleznikow, 'Negotiation, Online Dispute Resolution, and Artificial Intelligence', in *Handbook of Group Decision and Negotiation, Negotiation, Online Dispute Resolution, and Artificial Intelligence*, eds. Marc Kilgour, Colin Eden (2021):1125–1147; Attilio. M. Costabel, 'The Future of Online Justice According to Susskind: From COVID-19 Emergencies to Global Platforms' (2023) 15 *Journal of Multidisciplinary Research* 1.

99 Article 22 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).

100 REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL Of 13 June 2024 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).

Intelligence (“AI HLEG”).¹⁰¹ Most notably the White Paper on Artificial Intelligence¹⁰² and Guidelines for Trustworthy AI¹⁰³ were produced by the AI HLEG. As noted by Crofts and Rijswijk, the White Paper does not deal with the issue of bias directly and instead opts for the issue of discrimination,¹⁰⁴ however the problem as well as the remedies are identical.

The Guidelines do not define the origin or even the term bias, but they mention “inadvertent historic bias” present in the necessary datasets. The main remedy thus focuses on the data, specifically their representativeness and inclusivity, which should be, in a similar fashion to CEPEJs’ recommendation, extended even to the whole design and implementation phase.

The White Paper emphasises an important point for this discussion, and that is the fact that virtually any of our current decision-making suffers from being biased.¹⁰⁵ Besides the repeated mantra of proper data management, the White Paper also suggests that this issue could be mitigated by a proper log keeping, especially of the creation process. These frameworks are not only non-binding but they themselves do not pay much attention to the ways of achieving the recommendations.¹⁰⁶ This is then demonstrated by a certain divide between the “technical” and “legal” where one is not necessarily connected to the other, or more precisely when approach required by one is not possible by the other. This was, for example, a rather common critique of the (earlier versions) of the AI Act¹⁰⁷ that has required technically impossible safeguards, such as precise log keeping that goes against the nature of some of the Machine Learning approaches, such as DeepLearning. A similar discussion is currently ongoing on the requirement of the data management and its utility.¹⁰⁸ However, we hold that while this discussion is pertinent, so is the requirement for data management that has so far found its place in all the existing tools and recommendations. After all, even the argument of Hooker does not directly disregard this safeguard but rather tries to establish a more “efficient” or suitable one – and in general argues for the need for more nuanced discussion on the issue of the (origins of) bias.¹⁰⁹ Hooker further

101 See <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

102 AI HLEG, *White Paper on Artificial Intelligence: a European approach to excellence and trust* (Europa, 2020) <https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en> accessed 02.03.2025.

103 AI HLEG, *Ethics Guidelines for Trustworthy AI* (Europa, 2020). <<https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>> accessed 02.03.2025.

104 Crofts, Rijswijk, *Technology: New trajectories in law*, 87.

105 Ibid., 11.

106 This is not necessarily a critique of these works, since a) those are often produced by experts in different areas than law and governance, that should set the processes and b) one of the, rather popular, approaches to regulating new technologies – the approach of performance-based rules, is based on setting only the goals, i.e. defining the necessary level of protection offered to relevant rights, with the idea that such highly technical processes are best understood by their designers or operators, who are best to judge the way in which to reach these set goals. See for example Roger Brownsword, *Law 3.0* (Routledge, 2020) or Radim Polčák, *Internet a proměny práva*, (Auditorium, 2009).

107 Andrej Krištofik, Pavel Loutocký, ‘ODR and Online courts: What is their future after the AI Act’ (2024) 27 *Jusletter IT. Die Zeitschrift für IT und Recht* 1. <https://dx.doi.org/10.38023/9375d02d-afd2-47fd-9e71-ba679710f15a>.

108 See for example Sara Hooker, ‘Moving beyond “algorithmic bias is a data problem”’ (2021) 2 *Patterns* 4 <<https://www.sciencedirect.com/science/article/pii/S2666389921000611>> accessed 02.03.2025.

109 Ibid.

argues that a specific focus should be paid to the design of the relevant algorithmic models, thusly extending the need for inclusivity even into the design phase, which is a recommendation already known to us from the AI HLEG proposals.

Some of the existing legal tools are, quite reasonably, utilising a certain “middle of the road” approach to this criticism. While the questions of data management cannot be overlooked or disregarded, the inclusivity approach and focus – as well as the subsequent control – should further extend to the design phase as well. In theory, this approach has been explored for example by Cofone, who has argued that after properly exploring the existing datasets we may utilize their ‘faults’, such as the inclusion of protected characteristics, to properly train the algorithms in such controlled fashion, that would better prepare them for encountering the characteristics in the ‘real world’.¹¹⁰

Hooker’s criticism however brings another, perhaps more pressing, issue to the forefront of our discussion. This point seems to be rather neglected in the current discussion on the use of the automated tools by the judiciary, especially by the more sceptical voices. The same way that Hooker argues that the discussion on possibilities of automation (and the bias of such automated systems) should primarily nudge us towards the discussion of the bias present in the judiciary currently – we should keep in mind that the presented framework is viable only if we accept that the current system of judiciary and the way it deals with bias is working properly. If we want to point out the shortcomings of adapting the existing framework to automated decision-making framework we need to also have a discussion on how these shortcomings impact our current decision-making processes. Outside of the possibility of the future automation, this discussion could also serve to the benefit and improvement of the existing system. That is despite the fact that some authors, such as Chatziathanasiou argue that using the flaws of the existing system of judiciary as an argument in favour of the use of automation is dangerous, as it may lead to the erosion of public trust in the institutions.¹¹¹ We hold that such an argument is misleading and actually holding the discussion on the current shortcomings of the system in the face of its possible new developments should have the opposite effect in a democratic liberal state. Kashif and Jianxin have for example argued that the utilisation of AI-based decision-making tools could help us properly identify the existing flaws in the form of structural biases present in the system, which could subsequently be used to not only improve the training data in the data management phase but overall help us identify areas of improvement in the existing judicial decision-making.¹¹²

A certain ‘acceptance’ of the mere adoption model of the framework could be represented by the AI Act. It differs from the previously mentioned documents not only by the fact that it will be binding, but also because it establishes processes that are to be put in place to satisfy its goals. The AI Act explicitly warns against bias in judicial use of AI in the preliminaries,¹¹³ however the articles generally focus on fundamental rights as such. The AI Act uses risk-based classification, and based on

¹¹⁰ Cofone, “Algorithmic Discrimination Is an Information Problem”.

¹¹¹ Konstantin Chatziathanasiou, “Beware the Lure of Narratives: “Hungry Judges” Should Not Motivate the Use of “Artificial Intelligence” in Law”, *German Law Journal* 23, no. 4 (2022): 452–464.

¹¹² Javed Kashif, Li Jianxin, ‘Artificial Intelligence in Judicial Adjudication: Semantic Biasness Classification and Identification in Legal Judgement (SBCILJ)’ (2024) 10 *Heliyon* 9.

¹¹³ Point 40 of the AI Act.

the recommendations of Point 40 as well as classification guidelines set out in the Annex III of the AI Act, it appears that the judicial use of AI should fall into the high-risk category.¹¹⁴ This leads to the proposed remedies, which are unfortunately not applicable to any potentially biased AI-systems, but only to those that are categorized as high-risk.¹¹⁵ The most relevant requirement for these systems is to set up an iterative internal audit process, that would periodically check throughout the whole lifetime of the system, its impact on the fundamental rights, as well as compliance with other requirements. The audit should check for incorrectness of decisions, for example due to the decisions being biased, and whether they impact any fundamental right. The following requirement is to mitigate such possible impact or ideally remove it completely.

A similar approach seems to be currently favoured, at least in the context of AI decision-making, by the Canadian directive automated decision-making¹¹⁶ that is trying to address the issue of bias directly via a similar internal audit process and certification scheme. This prevalence of iterative audits in AI-based, decision-making systems is expected, given their technical nature.¹¹⁷ The iterative nature of the test seems to be a fitting response to the dynamic nature of “learning” systems, since the control mechanism is not fixed in time, the same way the decision-making algorithm is not.¹¹⁸ Similarly, such approach seems to be a one way¹¹⁹ of responding to the issue of “black box”,¹²⁰ that has been utilised for example by the Canadian Directive on Automated decision making,¹²¹ that points out that the iterative nature of such an audit allows the auditor to collect a dataset of the AI decisions and examine them for potential errors and biased decisions. This approach side-steps the issue of black box, since the potentially harmful pattern is being identified over the (large) dataset of outcomes, not the algorithmic process directly.¹²²

¹¹⁴ For more in-depth classification of judicial use of AI under the AI Act see Krištofik, Loutocký, ‘ODR and Online Courts: Their future after the AI Act’.

¹¹⁵ In very broad terms, those are mostly such systems that could potentially impact any fundamental rights.

¹¹⁶ See the text of Directive available at <<https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>> accessed 02.03.2025.

¹¹⁷ Iterative audits are a rather intuitive tool given the issues presented by the technical nature of the black box and as such it is expected that the legislator chooses it. This tool however remains far from ideal.

¹¹⁸ For the reasoning behind iterative audits see Reiling, ‘Courts and Artificial Intelligence’.

¹¹⁹ There are of course many other possible solutions to this problem, such as other technological advances such as the utilisation of back-propagation and other methods of xAI. See Sajid Ali et al., ‘Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence’ (2023) 99 *Information Fusion* <https://doi.org/10.1016/j.inffus.2023.101805> or even an approach of completely disallowing unexplainable models in high stakes decisions, see Cynthia Rudin, ‘Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead’1 *Nature Machine Intelligence*.

¹²⁰ Brozek, ‘The black box problem revisited. Real and imaginary challenges for automated legal decision making’.

¹²¹ Namely its point 6.3.

¹²² And even though important and necessary in the lifecycle of the given system, such data on its “past” decisions can be synthetically created even before the deployment, or within a sandbox in order to spot such negative patterns beforehand. For the approach of regulatory sandboxes for AI systems see Amnon Reichman, Giovanni Sartor, ‘Algorithms and Regulation’ in *Constitutional Challenges in the Algorithmic Society*, eds. Hans-Wolfgang Micklitz et al. (Cambridge University Press, 2021): 131–181.

The impact of the potentially biased decision, as well as the possible recourse, should serve as a guidance to how often such audit should be carried out. The first aspect of the impact is rather obvious and could be taken to essentially mean that the more serious the impact, the more often the system ought to be audited. For more specific guidance this should be taken together with the second aspect, that of the possible recourse. Therefore, the period of the audit should be such that the possible recourse is still available to the aggrieved party in a case that an error or defect is identified. This should inform the period even in such a case where this outcome of audit forms a new appellate reason, with its own timeframe set out in the procedural rules, that reflects the different nature of this appellate reason, as it should present a greater level of legal certainty not only for the aggrieved party, but for the other party as well.

The approach of iterative audits seems to be favoured even outside the public applications of artificial intelligence and automated decision-making, for example in the area of HR management and hiring processes.¹²³ Besides the relative sensitivity and the potential impact on the applicants life of the hiring process decision-making¹²⁴ that makes this area of automation similar to the automation in judicial decision-making, hiring processes were also one of the first to undergo automation with rather undesirable results.¹²⁵ Besides favouring the suggested approach of iterative audit, this area of decision-making automation seems to favour the certification approach as well.¹²⁶

The (public decision-making) certification scheme could then be imagined as a direct follow-up on the iterative audit.¹²⁷ Such understanding is favourable to the user of the system, in our case a court or different public body carrying out the decision-making, as it leaves them with option to either obtain such necessary certification by either outsourcing it to a third party, by obtaining a certificate from some specialized national supervisory authority,¹²⁸ or by performing a self-certification, as is currently widespread approach in other technological areas, such as cybersecurity.¹²⁹ Ajunwa in her article points to the possible issue of regulatory capture in the certification scheme.¹³⁰ This is a pertinent issue that should not be overlooked, nor understated, however it should make us pay greater attention to this issue when designing the certification schemes, rather than see it as an absolute obstacle in their adoption. After all, when it comes to any regulation of new technologies, regulatory capture is always a great risk

¹²³ Ifeoma Ajunwa, 'An auditing imperative for automated hiring systems.' (2020) 34 *Harvard Journal of Law & Technology*, 621.

¹²⁴ Zhisheng Chen, 'Ethics and discrimination in artificial intelligence-enabled recruitment practices' (2023) 10 *Humanities and Social Sciences Communication*. <https://doi.org/10.1057/s41599-023-02079-x>.

¹²⁵ As briefly described in the section 3.1 of this article.

¹²⁶ Ajunwa, 'An auditing imperative for automated hiring systems.'

¹²⁷ See a similar suggestion with complex procedural breakdown at Carlos Galán, 'Certification as a control mechanism of Artificial Intelligence in Europe' (Futurium) <https://ec.europa.eu/futurium/en/system/files/ged/c._galan_phd_-_ai_paper.pdf> accessed 02.03.2025.

¹²⁸ Similar to National Cybersecurity Certification Authorities established by the Cybersecurity Act, see Art. 58 <<https://eur-lex.europa.eu/eli/reg/2019/881/oj>> accessed 02.03.2025.

¹²⁹ See ENISA certification <<https://www.enisa.europa.eu/topics/certification/cybersecurity-certification-framework>> accessed 02.03.2025.

¹³⁰ Ajunwa, 'An auditing imperative for automated hiring systems.'

(as it was for example with AI Act and regulating artificial intelligence) but that should not stop us from regulating emerging technologies the right way.¹³¹

4.2 DE LEGE FERENDA AND OUR CURRENT UNDERSTANDING OF THE ISSUE

New issues often bring new attempts to solve them, which appears to be the case for algorithmic decision making as well. One such attempt is the concept of FAT algorithms (fair, accountable, transparent) with which Barabas argues for tackling the issue of biased decision-making as a part of the broader issue of fairness of these decisions. That should be achieved by a) establishing formal criteria for fairness and b) developing best practices to maintain them.¹³² This leads us to the last question of this article.

We have already established that a) bias is an issue for (semi)automated decision-making, and we have described b) potential remedies, mostly based on monitoring processes, that are currently rather popular within different frameworks¹³³. This leaves us with the last point of c) discussing the need for creating new criteria for bias (or fairness) in decision-making. Which begs the question to what level we should hold algorithmic processes to a higher standard than we do for the current “human-based” ones.

Susskind argues that any technological change must not be done purely for its own sake, but in order to fulfil some broader need in society, to address an existing issue.¹³⁴ Such thesis seems pretty reasonable, the question however is one of proportionality. If we help to resolve the issue of critically overloaded judicial system, do we still need to provide for greater non-bias than is currently the norm? We hold that this is not the case. Not necessarily since the betterment of an overloaded judicial system is enough, but due to the fact that the current interpretation of impartial tribunal sufficiently fulfils requirements of a fair trial. Thus, it should not be required to create a new criterion for the concept of fairness (and subsequently bias) but rather apply, and reinterpret, if necessary, the existing framework, much in the spirit of the “living instrument”.¹³⁵ It follows then, that should we accept that our current judicial system is unbiased, or more precisely that the standard we hold it to provides us with enough confidence in its decisions, then applying the same existing standard, with the necessary alteration,¹³⁶ should result in automated systems that provide us with enough confidence in their decision making, therefore there should not be any reason

¹³¹ Davide Baldini, Kate Francis, ‘AI Regulatory Sandboxes between the AI Act and the GDPR: The role of Data Protection as a Corporate Social Responsibility’ (2024) *CEUR Proceedings*. <<https://cris.maastrichtuniversity.nl/en/publications/ai-regulatory-sandboxes-between-the-ai-act-and-the-gdpr-the-role->> accessed 02.03.2025; Bartolomiej Chomanski, Lode Lauwaert, ‘Digital privacy and the law: the challenge of regulatory capture’ (2024) *AI & Society* <https://doi.org/10.1007/s00146-024-02041-8>.

¹³² Chelsea Barabas, ‘Beyond Bias. Ethical AI in criminal law’ in: *The Oxford Handbook of Ethics of AI*, eds. Markus D. Dubber et al. (2020): 740–741.

¹³³ For example by considering the requirements set forth by the ECtHR.

¹³⁴ Susskind, *Online courts and the future of justice*, 90–91.

¹³⁵ George Letsas, ‘The ECHR as a Living Instrument: Its Meaning and its Legitimacy’ in: *Constituting Europe*, eds. Andreas Follesdal (Cambridge University Press, 2013): 106–141.

¹³⁶ See for example the discussed auditing systems as a necessary trust conditions, or the creation of new transparency obligations.

to create a new standard.¹³⁷ Further, this is supported by the very nature of the right to a fair trial, that is aimed at protection of (fundamental rights of) the citizens, therefore the relevant factor ought to be the impact on citizens' rights, rather than the process by which it is created.

5. CONCLUSION

There is no lack of issues pertaining to judiciary use of AI that still remain to be tackled before we can implement a full judicial automation into judicial systems.¹³⁸ Even then, there is still a great discussion to be had about other, more innovative potentials of these systems, such as predictive justice.¹³⁹ This article was aimed at analysing only one of such problematic aspects, that of machine bias. While this issue presents a rather grave issue for the potential automated judicial decision-making, we have argued that we already have the necessary frameworks in place for us to properly deal with this issue. Further, being able to adopt the existing protection frameworks in such manner that they apply in these novel cases, specifically in the case of public automation, should serve to minimize the possible negative impact of such, as well as it should providing sufficient response to the critics, raising the issue of 'algorithmic tyranny'.¹⁴⁰ This rather important observation stems from the basic premise of this article and that is the fact that the protection we currently have in place is sufficient. As described by the references to the Susskindian maxim, this should however only present the minimal case scenario at which we can accept the utilisation of such systems and should not prevent us from further discussions on how to safeguards these fundamental rights even better.¹⁴¹

The conclusion of the existing sufficient framework follows from the observations made firstly in the introductory parts that have introduced the technical aspects of this problem and subsequently the existing framework, introduced by the ECtHR, for dealing with biased decision-making. Most importantly, the internal and external test for biased decision making were introduced.

The internal test, which requires that the judge should not hold any pre-existing beliefs that could skew her decision making, can be directly tied to the requirement of proper data handling, and be subsequently checked during the iterative audits that should search for emergence of unwanted biased patterns. Similarly to our current interpretation, unless proven otherwise, such system should be considered non-biased. As per the external test, it essentially establishes certain objective checks for the existence of bias. These could be reinterpreted or fulfilled by the required level

¹³⁷ As suggested by some mentioned proposals or for example by Reva Schwartz et al., 'Towards a Standard for Identifying and Managing Bias in Artificial Intelligence' (NIST, 2022). <<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>> accessed 02.03.2025.

¹³⁸ Imam Hanafi et al., 'Legal study on ethical issues in the use of artificial intelligence for legal decisions: critical literature review' (2024) *Inspirasi Futura: Media kajian Keislaman*, 9.

¹³⁹ A brief overview of the predictive justice issue is provided at Contini, 'Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings'.

¹⁴⁰ Bruno Lepri, 'The tyranny of data? The bright and dark sides of data-driven decision-making for social good', in *Transparent Data Mining Big and Small Data*, eds. Tania Cerquitelli et al. (Springer, 2017): 3–24.

¹⁴¹ Hooker, 'Moving beyond "algorithmic bias is a data problem".'

of transparency (and log keeping) of such decision-making systems, not only on the level of individual decisions, but as well on the level of design and implementation process and should be made widely available to public and open research. As per the “process” side of this issue, the first internal aspect should be satisfied precisely by the appropriate *iterative* internal audit and the second external aspect should be satisfied by an appropriate certification scheme.

This was only a very broad introduction to this, legally and technologically complex issue, not only to the problem as such, but to our current understanding as well as currently favoured remedies. Adoption of such AI-based tools into judiciary practice could help us to deal with some prevalent issues, such as the length of proceedings and the overburdened judicial system.¹⁴² On the other, if not deployed properly, such systems could have a grave impact on fundamental rights, such as the right to a fair trial, therefore design and implementation of appropriate tools, such as the suggested iterative controls must not be underestimated.

FUNDING INFORMATION

No funding has been received by the author in lieu of this research.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Andrej Krištofik  orcid.org/0000-0001-8150-0362
Institute of Law and Technology, Masaryk University, CZ

REFERENCES

- “AI’s best use case may actually be in our judicial system”, *IASToppers* (2023).
<https://www.iastoppers.com/articles/ai-s-best-use-case-may-actually-be-in-our-judicial-system> [Accessed 02.03.2025]
- AI HLEG, Ethics Guidelines for Trustworthy AI (Europa, 2020). <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html> [Accessed 02.03.2025]
- AI HLEG, White Paper on Artificial Intelligence: a European approach to excellence and trust (Europa, 2020) https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en [Accessed 02.03.2025]
- Ajunwa Ifeoma, “An auditing imperative for automated hiring systems.” *Harvard Journal of Law & Technology* 34 (2020).
- Albaroudi Elham, et al., “A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring” *AI* 5, no. 1 (2024): 383–404. DOI: <https://doi.org/10.3390/ai5010019>
- Ali Sajid, et al., “Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence”, *Information Fusion* 99 (2023). DOI: <https://doi.org/10.1016/j.inffus.2023.101805>

¹⁴² A different great legal discussion that firstly needs to happen is also the choice of proceedings that are automated as not all legal decision making processes are equal. This is however outside of the scope of this article but for a similar article see for example Reiling, “Courts and Artificial Intelligence”.

- Allsop James, "Technology And The Future Of The Courts", *University of Queensland Law Journal* 38, no. 1 (2019): 1–14. <https://journal.law.uq.edu.au/index.php/uqlj/article/view/1539/1293> [Accessed 02.03.2025]
- Amini Alexander, et al., "The impact of social segregation on human mobility in developing and industrialized regions" *EPJ Data Science* 3, no. 6 (2014). DOI: <https://doi.org/10.1140/epjds31>
- Andrus McKane, et al., "What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness", *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM, 2021): 249–260. DOI: <https://doi.org/10.1145/3442188.3445888>
- Angwin Julia, et al., "Machine Bias", *ProPublica* (2016). <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [Accessed 02.03.2025]
- Baldini Davide, Kate Francis, "AI Regulatory Sandboxes between the AI Act and the GDPR: The role of Data Protection as a Corporate Social Responsibility", *CEUR Proceedings* (2024). <https://cris.maastrichtuniversity.nl/en/publications/ai-regulatory-sandboxes-between-the-ai-act-and-the-gdpr-the-role-> [Accessed 02.03.2025]
- Bannon Alicia, Douglas Keith, "Remote Court: Principles for Virtual Proceedings During the COVID-19 Pandemic and Beyond" *Northwestern University Law Review* 115, no. 6 (2021): 1875–920.
- Barabas Chelsear, "Beyond Bias. Ethical AI in criminal law" in: *The Oxford Handbook of Ethics of AI*, eds. Markus D. Dubber et al. (2020). DOI: <https://doi.org/10.1093/oxfordhb/9780190067397.013.47>
- Barysé Dovile, Roeë Sarel, "Algorithms in the court: does it matter which part of the judicial decision-making is automated?", *Artificial Intelligence and Law* 32, no. 1 (2023): 117–146. DOI: <https://doi.org/10.1007/s10506-022-09343-6>
- Bingham Tom, *Rule of Law* (Penguin Books, 2011).
- Brożek Bartosz, et al., "The black box problem revisited. Real and imaginary challenges for automated legal decision making", *Artificial Intelligence and Law* 32 (2024): 427–440. DOI: <https://doi.org/10.1007/s10506-023-09356-9>
- Brownsword Roger, *Law 3.0* (Routledge, 2020). DOI: <https://doi.org/10.4324/9781003053835>
- Buijine Kars de, Erwin van Veen. "Mind the Frame: Bias at the Data-Analysis Stage." *Pride and Prejudice: Addressing Bias in the Analysis of Political Violence* (Clingendael Institute, 2017). <http://www.jstor.org/stable/resrep17336.5> [Accessed 02.03.2025]
- Buscemi v. Italy, ECtHR Judgement 16.09.1999.
- Carlson Alyssa M, "The Need for Transparency in the Age of Predictive Sentencing Algorithms", *Iowa Law Review* 103, no. 1 (2017): 303–329.
- Casaleiro Paula, et al., *Judicial Perceptions and Use of Technology. Portuguese Report* (University of Coimbra, 2023). <https://estudogeral.uc.pt/bitstream/10316/108789/1/Judicial%20perceptions%20and%20use%20of%20technology.pdf> [Accessed 02.03.2025]
- CCJE Opinion No. 26 (2023): *Moving forward: the use of assistive technology in the judiciary*, (Council of Europe, 2023). <https://rm.coe.int/ccje-opinion-no-26-2023-final/1680adade7> [Accessed 02.03.2025]
- Changqing Shi, et. al., "The Smart Court – A New Pathway to Justice in China?", *International Journal for Court Administration* 12, no. 1 (2021). DOI: <https://doi.org/10.36745/ijca.367>

- Chatziathanasiou Konstantin, "Beware the Lure of Narratives: "Hungry Judges" Should Not Motivate the Use of "Artificial Intelligence" in Law", *German Law Journal* 23, no. 4 (2022): 452–464. DOI: <https://doi.org/10.1017/glj.2022.32>
- Chen Benjamin M., Zhiu Li, "How Will Technology Change The Face Of Chinese Justice?" *Columbia Journal Of Asian Law* 34, no. 1 (2021): 1–58.
- Chen Zhisheng, "Ethics and discrimination in artificial intelligence-enabled recruitment practices", *Humanities and Social Sciences Communication* 10 (2023). DOI: <https://doi.org/10.1057/s41599-023-02079-x>
- Chomanski Bartolomiej, Lode Lauwaert, "Digital privacy and the law: the challenge of regulatory capture", *AI & Society* (2024). DOI: <https://doi.org/10.1007/s00146-024-02041-8>
- Chouldechova Alexandra, "A fair prediction with a disparate impact: a study on bias in recidivism prediction instruments", *Big Data* 5, no. 2 (2016). DOI: <https://doi.org/10.1089/big.2016.0047>
- Christian Brian, *Alignment Problem. Machine Learning and Human Values* (W.W. Norton, 2020).
- Cofone Ignatio, "Algorithmic Discrimination Is an Information Problem", *Hastings Law Journal* 70, no. 6 (2019): 1389–1444.
- Contini Francesco, "Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings", *Law, Technology and Humans* 2, no. 1 (2020): 4–18. DOI: <https://doi.org/10.5204/lthj.v2i1.1478>
- Costabel Attilio M, "The Future of Online Justice According to Susskind: From COVID-19 Emergencies to Global Platforms", *Journal of Multidisciplinary Research* 15, no. 1 (2023): 79–86.
- COVID-19 Continuity of Court Operations During a Public Health Emergency Workgroup. Post-pandemic recommendations. <https://www.americanbar.org/content/dam/aba/administrative/judicial/2021-az-post-pandemic-rec.pdf> [Accessed 02.03.2025]
- Crofts Penny, Honni Rijswijk, *Technology: New trajectories in law* (Routledge, 2021): 86. DOI: <https://doi.org/10.4324/9780429288685-5>
- Demková Simona, "The Decisional Value of Information in Semi-Automated Decision-Making", *Review of European Administrative Law* 2, no. 14 (2021): 29–50 <https://realaw.blog/2021/11/19/the-decisional-value-of-information-in-semi-automated-decision-making-by-simona-demkova/> [Accessed 02.03.2025]. DOI: <https://doi.org/10.7590/187479821X16254887670874>
- Dustin Jeffrey, Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters* (2018). <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> [Accessed 02.03.2025]
- Engel Christoph, et al., "Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism" *Artificial Intelligence and Law* (2024). DOI: <https://doi.org/10.1007/s10506-024-09389-8>
- Fenwick Mark, et al., "Regulation Tomorrow: What Happens When Technology Is Faster than the Law?", *American University Business Law Review* 6, no. 3 (2017). <https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1028&context=aubl> [Accessed 02.03.2025]. DOI: <https://doi.org/10.2139/ssrn.2834531>
- Galán Carlos, "Certification as a control mechanism of Artificial Intelligence in Europe" (Futurium) https://ec.europa.eu/futurium/en/system/files/ged/c_galan_phd_-_ai_paper.pdf [Accessed 02.03.2025]
- Gautrin and Others v. France, ECtHR Judgement 25.10.1998.

- Gualdi Francesco, Antonio Cordella, "Artificial Intelligence and Decision-Making: the question of Accountability" *Proceedings of the 54th Hawaii International Conference on System Sciences*, (2021): 2297–2306. DOI: <https://doi.org/10.24251/HICSS.2021.281>
- Hanafi Imam, et al., "Legal study on ethical issues in the use of artificial intelligence for legal decisions: critical literature review". *Inspirasi Futura: Media kajian Keislaman* (2024). DOI: <https://doi.org/10.31941/pj.v23i2.4659>
- Harris David, et al., *Law of the European Convention on Human Rights* (Oxford University Press, 2023). DOI: <https://doi.org/10.1093/he/9780198862000.001.0001>
- Hinkle Joshua C, David Weisburd, "The irony of broken windows policing: A micro-place study of the relationship between disorder, focused police crackdowns and fear of crime", *Journal of Criminal Justice* 36, no. 6 (2008): 503–12. DOI: <https://doi.org/10.1016/j.jcrimjus.2008.09.010>
- Hooker Sararh, "Moving beyond "algorithmic bias is a data problem", *Patterns* 2, no. 4 (2021) <https://www.sciencedirect.com/science/article/pii/S2666389921000611> [Accessed 02.03.2025]. DOI: <https://doi.org/10.1016/j.patter.2021.100241>
- "Independence and Impartiality of the Judiciary An overview of relevant jurisprudence of the European Court of Human Rights", *The Aire Centre* (2021) <https://www.rolplatform.org/wp-content/uploads/2021/09/independence-and-impartiality-of-judiciary-eng.pdf> [Accessed 02.03.2025]
- James Gareth, et al., "Unsupervised Learning" in: *An Introduction to Statistical Learning. Springer Texts in Statistics*, ed. Gareth James (Springer, 2023): 503–556. DOI: https://doi.org/10.1007/978-3-031-38747-0_12
- Janatian Samyar, et al., "From Text to Structure: Using Large Language Models to Support the Development of Legal Expert Systems", in *JURIX 2023: The Thirty-sixth Annual Conference*, eds. Giovanni Sielno et al. (IOS, 2023): 379. <https://ebooks.iospress.nl/volumearticle/65584>. [Accessed 02.03.2025]. DOI: <https://doi.org/10.3233/FAIA230962>
- Karrar v. Belgium, ECtHR Judgement 31.08.2021.
- Kashif Javed, Li Jianxin, (2024) "Artificial Intelligence in Judicial Adjudication: Semantic Biasness Classification and Identification in Legal Judgement (SBCILJ)", *Heliyon* 10, no. 9 (2024). DOI: <https://doi.org/10.1016/j.heliyon.2024.e30184>
- Kerikmäe Tanel, et al., "Frontiers in AI Judiciary: A Contribution to Legal Futurology", *Acta Baltica Historiae et Philosophiae Scientiarum* 11, no. 2 (2023): 55–75. DOI: <https://doi.org/10.11590/abhps.2023.2.03>
- Köchling Alina, Marius 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", *Business Research* 13 (2020): 795–848. [Accessed 02.03.2025]. DOI: <https://doi.org/10.1007/s40685-020-00134-w>
- Kordzadeh Nima, Maryam Ghasemaghaj, "Algorithmic bias: review, synthesis, and future research directions", *European Journal of Information Systems* 31, no. 3 (2021): 388–409. DOI: <https://doi.org/10.1080/0960085X.2021.1927212>
- Krištofik Andrej, "Európska a česká perspektíva francúzskeho zákazu profilovania sudcov" *Trestnôprávní revue – odborný recenzovaný časopis pro trestní pravo*, no. 4 (2022): 239–245.
- Krištofik Andrej, Pavel Loutocký, "ODR and Online courts: What is their future after the AI Act" *Jusletter IT. Die Zeitschrift für IT und Recht* 27, no. 1 (2024): 197–204. DOI: <https://doi.org/10.38023/9375d02d-afd2-47fd-9e71-ba679710f15a>
- Kusner Matt J, et al., "Counterfactual Fairness" in *Advances in Neural Information Processing Systems*, eds. I. Guyon et al. (NIPS, 2017): 4066–76. <http://papers.nips.cc/paper/6995-counterfactual-fairness.pdf> [Accessed 02.03.2025]

- Langford Mikael, Mikael R. Madsen, “France Criminalises Research on Judges” *Verfassungsblog* (2019). <https://verfassungsblog.de/france-criminalises-research-on-judges/> [Accessed 02.03.2025]
- Larson James, “How We Analyzed the COMPAS Recidivism Algorithm”, *ProPublica*, (2016). <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> [Accessed 02.03.2025]
- Lepri Bruno, “The tyranny of data? The bright and dark sides of data-driven decision-making for social good”, in *Transparent Data Mining Big and Small Data*, eds. Tania Cerquitelli et al. (Springer, 2017): 3–24. DOI: https://doi.org/10.1007/978-3-319-54024-5_1
- Letsas George, “The ECHR as a Living Instrument: Its Meaning and its Legitimacy” in: *Constituting Europe*, eds. Andreas Follesdal (Cambridge University Press, 2013): 106–141. DOI: <https://doi.org/10.1017/CBO9781139169295.005>
- Micallef v. Malta [GC], ECtHR Judgment 15.10.2009. DOI: [https://doi.org/10.1016/S1350-4789\(09\)70022-4](https://doi.org/10.1016/S1350-4789(09)70022-4)
- Mokander Jakob, et al., “Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations”, *Science and Engineering Ethics* 27, no. 44 (2021). DOI: <https://doi.org/10.1007/s11948-021-00319-4>
- Moore Taylor R, “Trade Secrets and Algorithms as Barriers to Social Justice”, *Center for Democracy and Technology* (2017). <https://cdt.org/wp-content/uploads/2017/08/2017-07-31-Trade-Secret-Algorithms-as-Barriers-to-Social-Justice.pdf> [Accessed 02.03.2025]
- Morewedge Carey K, et al., “Human bias in algorithm design” *Nature Human Behaviour* 7 (2023): 1822–24. DOI: <https://doi.org/10.1038/s41562-023-01724-4>
- Niller Eric, “Can AI Be a Fair Judge in Court? Estonia Thinks So”, *WIRED* (2019). <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/> [Accessed 02.03.2025]
- Pasquini v. San Marino, ECtHR Judgement 08.03.2021.
- Pedersen Tore, Christian Johansen, “Behavioural artificial intelligence: an agenda for systematic empirical studies of artificial inference”, *AI & Society* 35 (2020): 519–532. DOI: <https://doi.org/10.1007/s00146-019-00928-5>
- Pescador Valero v. Spain, ECtHR Judgement 17.06.2003.
- Polčák Radim, *Internet a proměny práva*, (Auditorium, 2009).
- Reichman Amnon, Giovanni Sartor, “Algorithms and Regulation” in *Constitutional Challenges in the Algorithmic Society*, eds. Hans-Wolfgang Micklitz et al. (Cambridge University Press, 2021): 131–181. DOI: <https://doi.org/10.1017/9781108914857.009>
- Reiling Dory, “Courts and Artificial Intelligence”, *International Journal for Court Administration* 11, no. 2 (2020). DOI: <https://doi.org/10.36745/ijca.343>
- Reiling Dory, *Technology for Justice. How Information Technology can support Judicial Reform* (Amsterdam University Press, 2009). DOI: <https://doi.org/10.5117/9789087280710>
- Roth Andrea L, “Trial by Machine”, *Georgetown Law Journal* 104, no. 5 (2016): 1245–1306.
- Rovatsos Michael, et al., “Landscape Summary: Bias in Algorithmic Decision-Making”, *Centre for Data Ethics & Innovation* (2019) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf [Accessed 02.03.2025].
- Rudin Cynthia, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”, *Nature Machine Intelligence* 1 (2019): 206–215. DOI: <https://doi.org/10.1038/s42256-019-0048-x>

- Rudin Cynthia, et al., "The Age of Secrecy and Unfairness in Recidivism Prediction", *Harvard Data Science Review* 2, no. 1 (2020). DOI: <https://doi.org/10.1162/99608f92.6ed64b30>
- Rustavi 2 Broadcasting Company Ltd and Others v. Georgia, ECtHR Judgement 9.12.2019.
- Sances Michael, Hye Y. You, "Who Pays for Government? Descriptive Representation and Exploitative Revenue Sources", *Journal of Politics* 79, no. 3 (2017): 1090–4. DOI: <https://doi.org/10.1086/691354>
- Schwartz Reva, et al., "Towards a Standard for Identifying and Managing Bias in Artificial Intelligence" (NIST, 2022). <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf> [Accessed 02.03.2025]. DOI: <https://doi.org/10.6028/NIST.SP.1270>
- Shi Jiahui, "Artificial Intelligence, Algorithms and Sentencing in Chinese Criminal Justice: Problems and Solutions" *Criminal Law Forum* 33 (2022):121–148. DOI: <https://doi.org/10.1007/s10609-022-09437-5>
- Sobowale Julia, "How Artificial Intelligence is changing the legal profession", *ABA Journal* (2016) https://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession [Accessed 02.03.2025]
- Soukupová Jana, "AI-based Legal Technology: A Critical Assessment of the Current Use of Artificial Intelligence in Legal Practice", *Masaryk University Journal of Law and Technology* 15, no. 2 (2021): 279–300. DOI: <https://doi.org/10.5817/MUJLT2021-2-6>
- "State v. Loomis", *Harvard Law Review* 130, no. 1530 (2016). <https://harvardlawreview.org/print/vol-130/state-v-loomis/> [Accessed 02.03.2025]
- Straton Papagianneas, "Automating Intervention in Chinese Justice: Smart Courts and Supervision Reform" *Asian Journal of Law and Society* 10, no. 3 (2023): 463–89. DOI: <https://doi.org/10.1017/als.2023.5>
- Straton Papagianneas, Nino Junius, "Fairness and justice through automation in China's smart courts", *Computer Law & Security Review* 51, no. 1 (2023). <https://www.sciencedirect.com/science/article/pii/S0267364923001073> [Accessed 02.03.2025]. DOI: <https://doi.org/10.1016/j.clsr.2023.105897>
- Sunstein Cass, "Algorithms, Correcting Biases", *Social Research: An International Quarterly* 86, no. 2 (2019): 499–511. DOI: <https://doi.org/10.1353/sor.2019.0024>
- Susskind Richard, *Online courts and future of justice* (Oxford University Press, 2012), 60–1.
- Tata Cyrus, "Introduction: Why Study Judicial Decision Support Systems?" *International Review Of Law Computers & Technology* 14, no. 3 (2000): 279–82. DOI: <https://doi.org/10.1080/713673365>
- Tian Niu W, Michael Yuan, "Intelligent Justice: AI Implementations in China's Legal Systems" in *Artificial Intelligence and its discontents*, ed. Ariane Haneemayer (Springer, 2022): 197–222. DOI: https://doi.org/10.1007/978-3-030-88615-8_10
- Vardanyan and Nanushyan v. Armenia, ECtHR Judgement 25.10.2016.
- Wang Liyuan, et al., "A Comprehensive Survey of Continual Learning: Theory, Method and Application" *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46 (2024). DOI: <https://doi.org/10.1109/TPAMI.2024.3367329>
- Weinberger Ota, *Law, Institutions and Legal Politics* (Springer, 1991). DOI: <https://doi.org/10.1007/978-94-011-3458-3>
- Xiang Alice, "Mirror, Mirror, on the Wall, Who's the Fairest of Them All?", *Daedalus* 153, no. 1 (2024): 250–267. DOI: https://doi.org/10.1162/daed_a_02058
- Xie Hanbo, "The promising future of cognitive science and artificial intelligence", *Nature Reviews Psychology* 202, no. 2 (2023). DOI: <https://doi.org/10.1038/s44159-023-00170-3>

- Yeung Karen, "Why Worry about decision-making by machine?", in *Algorithmic regulation*, ed. Karen Yeung (Wiley, 2017): 32–33.
- Zelevnikow John, "Negotiation, Online Dispute Resolution, and Artificial Intelligence", in *Handbook of Group Decision and Negotiation, Negotiation, Online Dispute Resolution, and Artificial Intelligence*, eds. Marc Kilgour, Colin Eden (2021):1125–1147. DOI: https://doi.org/10.1007/978-3-030-49629-6_38

Krištofik
*International Journal
 for Court Administration*
 DOI: 10.36745/ijca.598

TO CITE THIS ARTICLE:

Andrej Krištofik, 'Bias in AI (Supported) Decision Making: Old Problems, New Technologies' (2025) 16(1) *International Journal for Court Administration* 3. DOI: <https://doi.org/10.36745/ijca.598>

Published: 28 April 2025

COPYRIGHT:

© 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

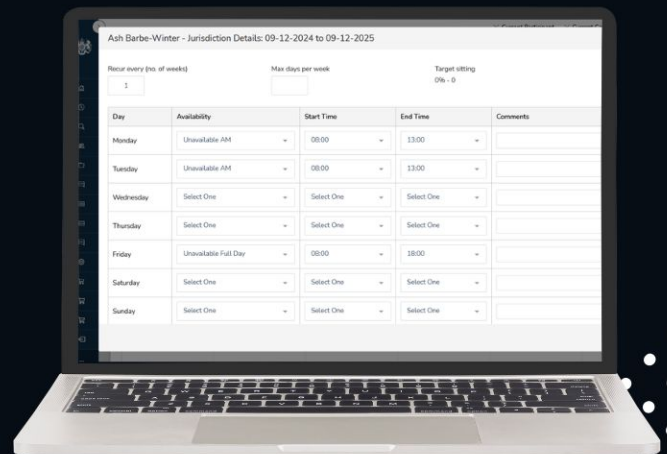
International Journal for Court Administration is a peer-reviewed open access journal published by International Association for Court Administration.



Streamline scheduling and listing processes with smarter session and resource planning tools

Offering greater flexibility to schedule events and manage listings, Scheduling & Resource Management allows you to eliminate manual processes.

Part of Ready Case, a modular cloud-based platform that brings speed and precision to judicial and non-judicial processes, Scheduling & Resource Management ensures more effective use of your judicial resources and physical spaces.



Key benefits of Scheduling & Resource Management



- 
- ✓ Increase the speed of scheduling in person or remote sessions
 - ✓ Cut through backlogs and scheduling conflicts
 - ✓ Lock in the right resources for sessions with less effort
 - ✓ Modular solution, allowing you take what you need, when you need it



Want to learn more?
Visit readytech.io/ready-case