TATuP Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis
Journal for Technology Assessment in Theory and Practice

oekom science

RESEARCH ARTICLE

Artificial intelligence and judicial decision-making: Evaluating the role of AI in debiasing

Giovana Lopes*,[1]

Abstract • As arbiters of law and fact, judges are supposed to decide cases impartially, basing their decisions on authoritative legal sources and not being influenced by irrelevant factors. Empirical evidence, however, shows that judges are often influenced by implicit biases, which can affect the impartiality of their judgment and pose a threat to the right to a fair trial. In recent years, artificial intelligence (AI) has been increasingly used for a variety of applications in the public domain, often with the promise of being more accurate and objective than biased human decision-makers. Given this backdrop, this research article identifies how AI is being deployed by courts, mainly as decision-support tools for judges. It assesses the potential and limitations of these tools, focusing on their use for risk assessment. Further, the article shows how AI can be used as a debiasing tool, i. e., to detect patterns of bias in judicial decisions, allowing for corrective measures to be taken. Finally, it assesses the mechanisms and benefits of such use.

Künstliche Intelligenz und richterliche Entscheidungsfindung: *Nutzung von KI zur Vermeidung kognitiver Verzerrungen*

Zusammenfassung • *Als Schiedsrichter*innen in Rechts- und Tatsachenfragen sollen Richter*innen unparteiisch entscheiden, indem sie ihre Entscheidungen auf der Grundlage maßgeblicher Rechtsquellen treffen und sich nicht von irrelevanten Faktoren beeinflussen lassen. Empirische Untersuchungen zeigen jedoch, dass Richter*innen häufig von kognitiven Verzerrungen (implicit biases) beeinflusst werden, die die Unvoreingenommenheit ihres Urteils beeinträchtigen und eine Gefahr für das Recht auf ein faires Verfahren darstellen können. In den letzten Jahren wurde künstliche Intelligenz (KI) vermehrt für eine Vielzahl von Anwendungen im öffentlichen Bereich eingesetzt, oft mit dem Versprechen, genauer und objektiver zu sein als voreingenommene menschliche Entscheidungsträger*innen. Vor diesem Hintergrund diskutiert dieser Forschungsartikel den Einsatz von KI in Gerichten, insbesondere als Entscheidungshilfe für Richter*innen, und bewertet das Potenzial und die Grenzen dieser Instrumente hinsichtlich deren Einsatz bei der Risikobewertung. Darüber hinaus wird gezeigt, wie KI als Instrument zur Verringerung von Vorurteilen genutzt werden kann, d. h. um Muster der Voreingenommenheit bei gerichtlichen Entscheidungen aufzudecken und ihnen entgegenzuwirken. Abschließend werden die Mechanismen und Vorteile einer solchen Nutzung bewertet.

Keywords • *judicial decision-making, judicial biases, artificial intelligence, risk assessment, debiasing*

This article is part of the Special topic "AI for decision support: What are possible futures, social impacts, regulatory options, ethical conundrums and agency constellations?," *edited by D. Schneider and K. Weber.*
https://doi.org/10.14512/tatup.33.1.08

Introduction

Several cognitive and social psychology studies suggest that judges are susceptible to various implicit biases which, unlike overt prejudice, they tend to be unaware of. These can influence their decisions in ways that are problematic considering the duty of impartiality and the right to a fair trial. The desire to increase objectivity, accuracy, and consistency in judicial decision-making has prompted the adoption of artificial intelligence (AI) to assist with different decision points throughout proceedings. At the same time, recognizing judges' susceptibility to biases raises the issue of how to mitigate them, and it is worth questioning whether AI might have a role to play in it. Hence, the goal of this article is twofold: First, it will describe how AI has been adopted for decision-support in judicial systems, and the challenges aris-

* *Corresponding author: giovana.figueiredo@unibo.it*
[1] *Department of Legal Studies, University of Bologna, Bologna, IT*

https://doi.org/10.14512/tatup.33.1.28
Received: 20. 08. 2023; revised version accepted: 15. 12. 2023;
published online: 15. 03. 2024 (peer review)

ing from such use; and second, it will evaluate the possibility of using AI for helping to detect and counteract judges' implicit biases, recommending which extralegal factors should be considered when doing so. It is a theoretical and bibliographic research, drawing on direct and indirect sources for a comprehensive analysis of the theme. The article will proceed as follows: I will first provide an overview of how implicit biases affect judicial decision-making, consequently giving rise to arguments favoring the adoption of algorithms to promote more accurate and objective decisions. Subsequently, I will explore their current use in judicial settings, focusing on decision-support tools that are adopted to promote risk assessments. The implications of using automated procedures in the judicial process will be evaluated, with the desirability of their adoption being called into question. I will then offer a different possibility of AI use in the judiciary, namely, to help identify and counteract judicial bias, therefore increasing fairness and legal certainty, before drawing some conclusions.

Given the high stakes involved in judicial decision-making, the issue of how to mitigate judicial bias is important.

Biases in adjudication

There is ample scientific evidence demonstrating how judges – like jurors and laypeople – are prone to both cognitive and social biases. While the former entails some broadly erroneous form of reasoning, the latter entails reasoning based on stereotypes (Zenker 2021). Biases have the potential to reduce the accuracy of a judgment and, throughout different stages of proceedings, can influence judicial decisions. To give some examples of relevant findings:

- Judges' sentencing decisions and compensation awards were found not only to be anchored by the initial demand made by the prosecutor, but also by random and unrelated factors to the decision at hand (Bystranowski et al. 2021).
- In a criminal investigation scenario, irrelevant contextual information affected judges' conviction rate, and confirmation bias led them to prefer incriminating investigations (Rassin 2020). Similarly, the pretrial detention of defendants later influenced judges' assessments of their guilt in criminal cases (Lidén et al. 2019).
- Judges' decisions were biased by the gender of the parties in studies involving hypothetical cases about child custody and relocation, employment discrimination, and criminal sentencing (Miller 2019; Rachlinski and Wistrich 2021).
- Data analysis of judges' bail decisions revealed racial bias against black defendants, even after controlling for variables such as criminal history and past pretrial misconduct (Arnold et al. 2020). In a virtual reality courtroom, minority defendants were treated more harshly by evaluators – including judges – during conviction (Bielen et al. 2021).

The idea that judicial decision-making can be influenced by extralegal factors is problematic considering judges' duty of impartiality and the right to a fair trial. Article 6 of the European Convention on Human Rights (ECHR) establishes the right to a fair trial by an independent and impartial tribunal. Impartiality requires that judicial decisions are based on the objective circumstances of the case, in accordance with the law, and free from external influences. Moreover, it excludes the existence of a prior disposition of the judge's mind that could lead them to favor or harm either party. The European Court of Human Rights (ECtHR) has distinguished between an objective aspect of this requirement, linked with the appearance of impartiality, and a subjective one, linked to "the personal conviction and behavior of a particular judge, that is, whether the judge held any personal prejudice or bias in a given case"[1]. The existence of a subjective approach may lead one to believe that there is an effective remedy to fight against judges' implicit biases, but such remedy is truly limited. The ECtHR has recognized the hardship of establishing a breach of Article 6 on account of subjective partiality, given the difficulty to procure evidence with which to rebut the presumption of impartiality, and has thus justified its common recourse to the objective analysis.[2]

One way to do so is through the implementation of debiasing techniques, which seek to address biases' negative effects by improving either the decision-making process or some relevant characteristics of the decision-maker (Zenker 2021). Another possibility relates to the adoption of artificial intelligence in judicial systems as decision-support or decision-making tools. 'Artificial intelligence' is used as an umbrella term to describe various human-designed technologies that exhibit intelligent behavior, analyzing their environment, and taking actions – with a certain level of autonomy – to achieve specific goals. The use of AI brings with it the promise of more accuracy, objectivity, and consistency, with governments increasingly adopting the technology "to attain greater accuracy when making predictions, replace biased human decisions with 'objective' automated ones, and promote more consistent decision-making" (Green 2022,

1 ECHR, Micallef v. Malta, Judgment of 15 October 2009, Application No. 17056/06, p. 22.

2 ECHR, Kyprianou v. Cyprus, Judgment of 15 December 2005, Application No. 73797/01.

p. 3). However, and at least for the time being, not only do these systems also have several limitations that can further deepen the problem of bias in adjudication, but there is also a risk-magnifying potential associated with AI that is not present with human decision-making (Dietterich 2019). In the following session, I will examine how AI has been adopted in judicial systems, specifically as decision-aid tools for assessing risk, and the challenges posed by this use.

AI in judicial systems

To assist adjudication, several countries are experimenting with and integrating digital technologies, particularly AI, in their judicial systems. Applications like advanced case-law search engines, online dispute resolution, or document categorization and screening can potentially lower the cost of dispute resolution and help courts address their backlog of cases, many of which are low-volume, low-value, and low-complexity matters (Steponenaite and Valcke 2020). Furthermore, some evidence suggests that algorithms are better at making policy-relevant predictions than public servants (Kleinberg et al. 2018). This makes the prospects of adopting digital technologies in judicial systems, particularly supportive and advisory AI-based tools, quite promising. On the other side, the use of algorithms to make consequential decisions about the application of public policy to individuals in street-level bureaucracies like courts, police departments, and welfare agencies has been highly controversial (Angwin et al. 2016; Heaven 2020; Allhutter et al. 2020).

Considering this scenario, institutions such as the European Union (EU) and the Council of Europe (CoE) are working towards ensuring that the development, implementation, and use of AI is done in an ethical and lawful way, especially in contexts where there is a high impact on individuals' fundamental rights. While the EU is in the final stages of approving a regulation creating standardized rules for AI (European Commission 2021), the CoE's Commission for the Efficiency of Justice adopted the first European Ethical Charter on the use of AI in judicial systems (CEPEJ 2018). In it, the Commission, which is responsible for evaluating European judicial systems and defining concrete ways to improve their performance, provides an ethical framework to guide private and public stakeholders throughout the development and implementation of AI in the judiciary. Continuing this work, in April 2023 CEPEJ also launched a Resource Centre on Cyberjustice and AI, which aims to serve as a publicly accessible focal point for reliable information on AI systems applied in the transformation of judicial systems (CEPEJ 2023). One of its first endeavors was to obtain an overview of these systems, providing a starting point for further examination of their risks and benefits for professionals and end-users. A total of 58 systems were identified in CoE member states, and then classified according to their main field of application. Categories include, e.g., 'anonymization tools', which are used for removing identifying information of court users, and 'natural language processing tools', used for speech recognition and the automatic transcription of court procedures. My analysis here will focus, however, on the category of 'decision-support and decision-making', which encompasses tools meant to facilitate or fully automate decision-making processes in justice systems, considering that some think that highly accurate AI systems could improve the performance of judges, or even come to replace them (Chatziathanasiou 2022).

First, it is worth highlighting that the use of the tools mapped by CEPEJ by judges themselves is still quite limited, and the initiative for their development remains primarily within the private sector, focusing on insurance companies, lawyers, and legal services wishing to reduce the uncertainty and unpredictability of judicial decisions. The French application Predictice, for example, is a predictive justice tool developed to calculate the chances of success of a legal action according to different variables, using jurisprudence analysis algorithms. More recently, it has launched a generative AI tool called Assistant, developed to answer legal professionals' questions by citing reliable and up-to-date sources (Larret-Chahine 2023). Even though most applications of this kind have their use currently restricted to private agents, "public decision-makers are beginning to be increasingly solicited by a private sector wishing to see these tools […] integrated into public policies" (CEPEJ 2018, p. 14).

Second, not all the applications listed at the Resource Centre can technically be categorized as AI, as is the case for many of the risk assessment tools, mainly used for assessing the risk of recidivism. These make up for the majority of 'decision-support and decision-making' tools that have been listed by CEPEJ as being currently used in the public sector, namely by judges. Examples include OASys, the Offender Assessment System used by the prison and probation services in England and Wales (Justice Data Lab 2016), RITA, the Finish Risk and Needs Assessment Form (Salo et al. 2019), or RISC, the Recidivism Assessment Scale adopted in the Netherlands (van Essen et al. n.d.). Risk assessments are perceived and often marketed as an objective means of overcoming human bias in decision-making and have been adopted to assist with several decision points throughout the criminal justice system, from pretrial release to post-conviction sentencing, probation, and parole. These tools do not use new statistical methods commonly associated with AI, such as machine learning (ML), but are rather overwhelmingly based on regression models (Barabas et al. 2018). The main goal of regression is to identify a set of variables (e.g., prior arrest) that are predictive of a given outcome variable (e.g., risk of reoffending). This process can be automatized and improved using ML methods (Ghasemi et al. 2021), but still constitute an incremental innovation in the way risk assessments have historically worked, instead of being truly transformational.

One example of a risk assessment tool that incorporates a machine learning approach is the Correctional Offender Management Profiling for Alternative Sanctions, or COMPAS, used

 https://doi.org/10.14512/tatup.33.1.28

by some United States' courts to assess the likelihood of recidivism (van Dijck 2022). Ever since an exposé by the news outlet ProPublica revealed that the software was biased against blacks (Angwin et al. 2016), it has become the primary example of the risks posed by algorithmic crime prediction overall. The controversy surrounding its use revolves around what the models measure and intend to measure, the accuracy of the predictions, and whether they might increase inequality and discrimination or otherwise compromise fairness (Mayson 2019; Rudin et al. 2020). The question of (un)fairness is of particular concern, given that risk assessment tools can lead to discriminatory outcomes based on race or ethnicity (Jordan and Bowman 2022). Furthermore, since the software is proprietary, the data and algorithms are not transparent, neither for the suspect nor for the judge (van Dijck 2022), a problem that was addressed in the case of Loomis v. Wisconsin[3] (2016). In this case, while admitting the system's flaws, the Court claimed that it is up to judges to exercise discretion when assessing a risk score.

In high-stakes decisions such as criminal justice risk assessments, it is common to place emphasis on the decision-makers' discretion in incorporating algorithmic advice into their decisions, to make their use acceptable even in light of flaws. However, human discretion does not necessarily improve outcomes. Decision-makers are susceptible to automation bias, a tendency to defer to automated systems, reducing the amount of independent scrutiny exhibited when deciding (Parasuraman and Manzey 2010). Similar issues arise when humans collaborate with predictive algorithms. Recent research has found that people are bad at judging the quality of algorithmic outputs and determining whether and how to override those outputs (Green 2022). In simulated pretrial and sentencing decisions, for instance, risk assessments made participants – including judges – place a greater emphasis on its results than on other relevant factors (Green and Chen 2021). It is thus likely that, instead of improving the issue of bias by promoting an 'objective' score, the incorporation of results into a decision may nonetheless result in biased outcomes.

The challenges discussed in this section do not necessarily entail a categorical rejection of employing AI for assisting in judicial decisions, but it does raise the question of which uses might be advantageous without presenting a risk to the fairness of a trial. Here, one possibility relates to its use for triaging, allocation, and workflow automation, facilitating some activities during the lifecycle of proceedings and minimizing the need for human input. For instance, one Higher Regional Court in Germany began using AI to help process the large volume of lawsuits relating to a scandal involving vehicle manufacturers. With more than 13,000 cases pending and around 600 entries added monthly, the AI is used to analyze the files and organize them according to the facts, but the judges remain responsible for processing the content, reviewing it, and making decisions (SWR 2022).

3 Wisconsin Supreme Court, Wisconsin v. Loomis 2016 WI 68.

AI as a debiasing tool

After examining how AI has been used for decision-support in judicial settings, I will now explore the possibility of applying AI methods for helping detect and counteract judges' implicit biases. One way of doing so is through predictive judicial analytics in the form of machine learning. As Chen (2019 a) explains, biases are most likely to manifest in situations where judges are closer to indifference between options. These contexts of 'judicial indifference' are also ones where the highest levels of disparities in inter-judge accuracy are present – essentially conditions where judges are unmoved by legally relevant circumstances. A judge could be said to have strong preferences over legally relevant circumstances (which are covariates) when it is costly to depart from the legally optimal outcome, defined as the outcome that would be generated through consideration of legal facts alone. But a judge may also have weak preferences over legally relevant circumstances, meaning that there is a relatively low cost in departing from the legally optimal outcome. In such cases of legal indifference, a different set of covariates that are legally irrelevant (and thus should not predict a legal outcome) can be expected to have greater influence. In other words:

> "If a judge can be predicted prior to observing the case facts, one might worry about the use of snap or pre-determined judgements, or judicial indifference. To put it differently, the preferences of judges over the legally relevant covariates may affect the influence of irrelevant features. A judge could be said to have weak preferences, meaning that there was a relatively low cost in departing from the legally optimal outcome. In such cases of legal indifference, irrelevant factors can be expected to have greater influence. Behavioral bias reveals when decision-makers are indifferent" (Chen 2019 b, p. 16).

The accuracy of predictions depends on the number of judicial attributes or characteristics of the case that are analyzed by the system. In a study conducted by Dunn et al. (2017) on asylum courts, using only data available at the case opening (i. e., information on the judge's identity and the nationality of the asylum seeker), researchers were able to predict case outcomes with an accuracy of 78 %. Through this notion of 'early predictability', ML could be used to automatically detect judicial indifference, alerting to situations where extralegal factors are more likely to influence a decision. This raises the question of which other sets of data could be incorporated into the legally irrelevant covariates to improve predictive accuracy, requiring an analysis of which are the main extralegal factors that influence judges when deciding. Based on an examination of the literature on social biases, initial contenders include race, gender, and ethnicity of the defendant and of the judge, the latter for the assessment of ingroup favoritism. And based on findings on cognitive biases, the number of (un)favorable previous decisions by the court, the comparison of caseloads between judges, and whether it is a spe-

https://doi.org/10.14512/tatup.33.1.28

cialized court (all being indicative of contrast effects) are variables likely to influence the way judges decide. Other factors besides biases include the time of day at which a decision is made (Shroff and Vamvourellis 2022; Danziger et al. 2011), and temperature (Chen and Loecher 2019; Heyes and Saberian 2019). These are of course only initial suggestions and do not fully encompass all the irrelevant covariates that can affect a decision. But the advantage of using machine learning techniques for this purpose is precisely that any sort of data can be used to feed the model, enabling patterns and trends to emerge without necessarily requiring a theoretical explanation for such.

In the in-depth study on the use of AI in judicial systems that accompanies CEPEJ's ethical charter (2018), the commission analyzes the benefits and risks of different applications, encouraging their use to various degrees. While specific judge profiling is highly discouraged, among uses to be considered following additional studies is offering judges an assessment of their activities with an informative aim of assisting in decision-making. Indeed, offering judges' feedback regarding their decisions is a fundamental step in debiasing, alongside other interventions such as the promotion of general bias awareness, training in rules and representations, exposure to stereotype-incongruent models, and the adoption of scripts and checklists (Wistrich and Rachlinski 2017) – all of which can be directed once bias is identified. Furthermore, AI offers a mechanism of detecting bias in real time, and could hence alert judges to situations where biases are likely to occur (e.g., after a string of positive decisions), allowing them to intervene before a biased decision takes place.

Concluding remarks

The adoption of digital technologies like artificial intelligence in judicial settings often comes from a desire to increase objectivity, accuracy, and consistency in decision-making, improving the quality of decisions traditionally made by humans. However, we have seen how the use of AI for decision-support in adjudication, albeit still not prevalent in CoE member states, can worsen issues already identified in the risk assessment tools that they seek to automate. These include the accuracy (or lack thereof) of their predictions, the reproduction of existing patterns of prejudice and bias, the lack of transparency and opportunities for defendants to challenge their outcomes, and the difficulty of decision-makers to properly evaluate assessments. Thus, instead of using AI to make decisions that traditionally pertain to judges (e.g., assessing the risk of recidivism), a different possibility was offered, namely, to employ the technology for debiasing purposes. AI can help identify the situations where judicial bias is likely to take place, based on the analysis of covariates that, despite being legally irrelevant, have been shown to influence judicial decisions, some of which were listed here. By identifying the instances in which bias commonly arises, it is possible not only to alert judges but also to target debiasing interventions, such as educating judges on the subject and offering feedback on their work, with the ultimate goal of ensuring objectivity and impartiality in their decisions.

Funding • This work received no external funding.
Competing interests • The author declares no competing interests.

References

Allhutter, Doris; Cech, Florian; Fischer, Fabian; Grill, Gabriel; Mager, Astrid (2020): Algorithmic profiling of job seekers in Austria. How austerity politics are made effective. In: Frontiers in Big Data 3 (5), pp. 1–17. https://doi.org/10.3389/fdata.2020.00005

Angwin, Julia; Larson, Jeff; Mattu, Surya; Kirchner, Lauren (2016): Machine bias. There's software used across the country to predict future criminals. And it's biased against blacks. In: ProPublica, 23. 05. 2016. Available online at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, last accessed on 22. 01. 2024.

Arnold, David; Dobbie, Will; Hull, Peter (2020): Measuring racial discrimination in bail decisions. In: NBER Working Paper Series, pp. 1–84. https://doi.org/10.3386/w26999

Barabas, Chelsea; Virza, Madars; Dinakar, Karthik; Ito, Joichi; Zittrain, Jonathan (2018): Interventions over predictions. Reframing the ethical debate for actuarial risk assessment. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81, pp. 62–76. Available online at https://proceedings.mlr.press/v81/barabas18a.html, last accessed on 22. 01. 2024

Bielen, Samantha; Marneffe, Wim; Mocan, Naci (2021): Racial bias and in-group bias in virtual reality courtrooms. In: The Journal of Law and Economics 64 (2), pp. 269–300. https://doi.org/10.1086/712421

Bystranowski, Piotr; Janik, Bartosz; Próchnicki, Maciej; Skórska, Paulina (2021): Anchoring effect in legal decision-making. A meta-analysis. In: Law and Human Behavior 45 (1), pp. 1–23. https://doi.org/10.1037/lhb0000438

Chatziathanasiou, Konstantin (2022): Beware the lure of narratives. 'Hungry Judges' should not motivate the use of "Artificial Intelligence" in law. In: German Law Journal 23 (4), pp. 452–464. https://doi.org/10.1017/glj.2022.32

Chen, Daniel (2019 a): Machine learning and the rule of law. In: Michael Livermore and Daniel Rockmore (eds.): Law as Data. Santa Fe, NM: SFI Press, pp. 433–441.

Chen, Daniel (2019 b): Judicial analytics and the great transformation of American law. In: Artificial Intelligence and Law 27 (1), pp. 15–42. https://doi.org/10.1007/s10506-018-9237-x

Chen, Daniel; Loecher, Markus (2019): Mood and the malleability of moral reasoning. In: SSRN Electronic Journal, pp. 1–62. https://dx.doi.org/10.2139/ssrn.2740485

Danziger, Shai; Levav, Jonathan; Avnaim-Pesso, Liora (2011): Extraneous factors in judicial decisions. In: Proceedings of the National Academy of Sciences 108 (17), pp. 6889–6892. https://doi.org/10.1073/pnas.1018033108

Dietterich, Thomas (2019): Robust artificial intelligence and robust human organizations. In: Frontiers of Computer Science 13 (1), pp. 1–3. https://doi.org/10.1007/s11704-018-8900-4

Dunn, Matt; Sagun, Levent; Şirin, Hale; Chen, Daniel (2017 a): Early predictability of asylum court decisions. In: ICAIL '17. Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law. New York, NY: Association for Computing Machinery, pp. 233–236. https://doi.org/10.1145/3086512.3086537

European Commission (2021): Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on artificial

intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Brussels: European Commission. Available online at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206, last accessed on 22. 01. 2024.

CEPEJ – European Commission for the Efficiency of Justice (2018): European ethical charter on the use of artificial intelligence in judicial systems and their environment. Strasbourg: Council of Europe. Available online at https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c, last accessed on 22. 01. 2024.

CEPEJ (2023): Resource centre on cyberjustice and AI. Available online at https://public.tableau.com/app/profile/cepej/viz/ResourceCentreCyberjusticeandAI/AITOOLSINITIATIVESREPORT, last accessed on 22. 01. 2024.

Ghasemi, Mehdi; Anvari, Daniel; Atapour; Mahshid; Wormith, Stephen; Stockdale, Keira; Spiteri, Raymond (2021): The application of machine learning to a general risk-need assessment instrument in the prediction of criminal recidivism. In: Criminal Justice and Behavior 48 (4), pp. 518–538. https://doi.org/10.1177/0093854820969753

Green, Ben; Chen, Yiling (2021): Algorithmic risk assessments can alter human decision-making processes in high-stakes government contexts. In: Proceedings of the ACM on Human-Computer Interaction 5 (CSCW2). New York, NY: Association for Computing Machinery, pp. 1–33. https://doi.org/10.1145/3479562

Green, Ben (2022): The flaws of policies requiring human oversight of government algorithms. In: Computer Law & Security Review 45, pp. 1–22. https://doi.org/10.1016/j.clsr.2022.105681

Heaven, Will (2020): Predictive policing algorithms are racist. In: MIT Technology Review, 17. 07. 2020. Available online at https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/, last accessed on 10. 01. 2024.

Heyes, Anthony; Saberian, Soodeh (2019): Temperature and decisions. In: American Economic Journal 11 (2), pp. 238–265. https://doi.org/10.1257/app.20170223

Kleinberg, Jon; Lakkaraju, Himabindu; Leskovec, Jure; Ludwig, Jens; Mullainathan, Sendhil (2018): Human decisions and machine predictions. In: The Quarterly Journal of Economics 133 (1), pp. 237–293. https://doi.org/10.1093/qje/qjx032

Larret-Chahine, Louis (2023): Predictice lance assistant, une IA générative pour les professionnels du droit. In: Predictice Blog, 26. 05. 2023. Available online at https://blog.predictice.com/assistant-ia-pour-les-professionnels-du-droit, last accessed on 10. 01. 2024.

Jordan, Kareem; Bowman, Rachel (2022): Interacting race/ethnicity and legal factors on sentencing decisions. A test of the liberation hypothesis. In: Corrections 7 (2), pp. 87–106. https://doi.org/10.1080/23774657.2020.1726839

Lidén, Moa; Gräns, Minna; Juslin, Peter (2019): 'Guilty, no doubt'. Detention provoking confirmation bias in judges' guilt assessments and debiasing techniques. In: Psychology, Crime & Law 25 (3), pp. 219–247. https://doi.org/10.1080/1068316X.2018.1511790

Mayson, Sandra (2019): Bias in, bias out. In: The Yale Law Journal 128 (8), pp. 2218–2300. Available online athttps://www.yalelawjournal.org/article/bias-in-bias-out, last accessed on 10. 01. 2024.

Miller, Andrea (2019): Expertise fails to attenuate gendered biases in judicial decision making. In: Social Psychological and Personality Science 10 (2), pp. 227–234. https://doi.org/10.1177/1948550617741181

Parasuraman, Raja; Manzey, Dietrich (2010): Complacency and bias in human use of automation. In: Human Factors 52 (3), pp. 381–410. https://doi.org/10.1177/0018720810376055

Rachlinski, Jeffrey; Wistrich, Andrew (2021): Benevolent sexism in judges. In: San Diego Law Review 58 (1), pp. 101–142. Available online at https://digital.sandiego.edu/sdlr/vol58/iss1/3, last accessed on 22. 01. 2024.

Rassin, Eric (2020): Context effect and confirmation bias in criminal fact finding. In: Legal and Criminological Psychology 25 (2), pp. 80–89. https://doi.org/10.1111/lcrp.12172

Salo, Benny; Laaksonen, Toni; Santtila, Pekka (2019): Predictive power of dynamic (vs. static) risk factors in the Finnish risk and needs assessment form. In: Criminal Justice and Behavior 46 (7), pp. 939–960. https://doi.org/10.1177/0093854819848793

Steponenaite, Vilte; Valcke, Peggy (2020): Judicial analytics on trial. An assessment of legal analytics in judicial systems in light of the right to a fair trial. In: Maastricht Journal of European and Comparative Law 27 (6), pp. 759–773. https://doi.org/10.1177/1023263X20981472

Shroff, Ravi; Vamvourellis, Konstantinos (2022): Pretrial release judgments and decision fatigue. In: Judgment and Decision Making 17 (6), pp. 1176–120. https://doi.org/10.1017/S1930297500009384

Rudin, Cynthia; Wang, Caroline; Coker, Beau (2020): The age of secrecy and unfairness in recidivism prediction. In: Harvard Data Science Review 2 (1), pp. 1–53 https://doi.10.1162/99608f92.6ed64b30

SWR (2022): OLG Stuttgart setzt KI bei Diesel-Klagen ein. In: SWR Aktuell, 24. 10. 2022. Available online at https://www.swr.de/swraktuell/baden-wuerttemberg/stuttgart/olg-stuttgart-mit-ki-gegen-flut-von-dieselklagen-100.html, last accessed on 10. 01. 2024.

Justice Data Lab (2016): Incorporating offender assessment data to the justice data lab process. London: Ministry of Justice. Available online at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491688/oasys-methodology.pdf, last accessed on 22. 01. 2024.

Van Essen, Laurus; Van Alphen, Huib; Van Tuinen, Jan-Maarten (n.d.): Risk assessment the Dutch way. A scalable, easy to use tool for probation reports. In: Confederation of European Probation News. Available online at https://www.cep-probation.org/risk-assessment-the-dutch-way-a-scalable-easy-to-use-tool-for-probation-reports/, last accessed on 22. 01. 2024.

Van Dijck, Gijs (2022): Predicting recidivism risk meets AI act. In: European Journal on Criminal Policy and Research 28 (3), pp. 407–423. https://doi.org/10.1007/s10610-022-09516-8

Wistrich, Andrew; Rachlinski, Jeffrey (2017): Implicit bias in judicial decision making. How it affects judgement and what judges can do about it. In: Sarah Redfield (ed.): Enhancing justice: Reducing bias, pp. 87–130. https://doi.org/10.31228/osf.io/sz5ma

Zenker, Frank (2021): De-biasing legal factfinders. In: Christian Dahlman, Alex Stein and Giovanni Tuzet (eds.): Philosophical foundations of evidence law. Oxford: Oxford University Press, pp. 395–410. https://doi.org/10.1093/oso/9780198859307.003.0027

GIOVANA LOPES
is a doctoral candidate in "Law, Science and Technology" at the University of Bologna and KU Leuven and an affiliated research fellow at the Centre for IT & IP Law (CiTiP).

https://doi.org/10.14512/tatup.33.1.28