Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
http://ajee-journal.com

CASE STUDY

Case Study

JUDICIAL AI AND THE IRREPARABLE BIAS PROBLEM

Nataliia Mazaraki* and Dmytro Honcharuk

ABSTRACT

Background: *Courts are increasingly experimenting with large language models (LLMs) for tasks such as legal retrieval, drafting support, anonymisation, and triage. Yet the promise of efficiency collides with a structural problem: bias. Human adjudication already reflects cognitive and institutional biases; LLMs trained on past judgments and legal text inherit and sometimes amplify those biases. This article asks a focused question: If AI belongs in courts at all, what is the safe, lawful, and useful lane—especially with respect to bias? The inquiry is situated within fair-trial guarantees and emerging regulatory expectations.*

Methods: *A staged analysis grounded in legal obligations and informed by relevant technical characteristics is employed. First, sources of human and judicial bias are mapped, along with points at which LLMs introduce or magnify bias. Second, hard- and soft-law guardrails relevant to bias control in the justice sector are synthesised. Third, two instructive case studies—COMPAS/Loomis (U.S.) and Ewert v. Canada—are examined to demonstrate how group-level disparities and model opacity can generate due-process risks and to identify remedies transferable to LLM-assisted workflows. Finally, an operational blueprint is derived and applied to identify low-risk, high-yield assistive uses for Ukraine.*

DOI:
https://doi.org/10.33327/AJEE-18-8.S-c000159

Date of submission: 02 Nov 2025
Date of acceptance: 24 Nov 2025
Online First publication: 11 Dec 2025

Disclaimer:
The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

Results and conclusions: *The analysis shows that fully impartial AI outputs are not attainable in adjudication; bias is ineliminable but can be bounded. For Ukraine, the rational path is to invest first in data curation, secure infrastructure, evaluation capacity, and procurement with audit rights, and to confine AI to retrieval, norm collation, drafting-hygiene checks, and "missed-norms" prompts. The contribution is a governance blueprint that ties specific LLM failure modes to enforceable legal duties and practical safeguards—offering courts a credible, bias-aware lane for AI that improves service while preserving rights.*

1 INTRODUCTION

The legal field is experiencing a cognitive revolution driven by large language models (LLMs). LLMs are changing how courts gather and review case law and write decisions.[1] Research shows that in some jurisdictions, AI technology has begun to provide judgment recommendations in summary proceedings and has, to some extent, improved the efficiency of court document processing (such as China's "Smart Court" project). However, this change may create systemic risks when technological efficiency gains conceal potential cognitive biases.

AI in the judiciary has been explored across multiple contexts: criminal adjudication and sentencing support, bail and parole risk assessment, case-management and triage, judgment drafting and anonymisation, e-discovery, and "smart court" pilots. The research showed measurable efficiency gains in routine tasks but uneven effects on fairness and transparency. Methodologically, the field combines doctrinal analyses of due-process and equality guarantees, case-based audits (e.g., COMPAS/Loomis), technical evaluations of model behaviour, and policy studies of court pilots and procurement.

A consistent theme in this literature is that data-driven tools do not erase human bias; they can mask and reproduce it. O'Neil's early critique framed the problem starkly: data may look objective, but the choices of collection, labelling, modelling and deployment are not[2]—a point echoed in more recent legal analyses.[3] Empirical and technical work converges on three practical bottlenecks for lawful use in courts: data curation (authoritative provenance, representativeness, lawful reuse), document annotation (consistent labelling and redaction at scale), and output validation (citation fidelity, subgroup performance, audit trails).[4] Studies of criminal justice applications report potential gains in speed and consistency, but

1 OECD, *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions* (OECD Publishing 2025) doi:10.1787/795de142-en.

2 Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016).

3 Qingxia Chen, 'Improving the Trial Efficiency of Criminal Cases with the Assistance of Artificial Intelligence' (2025) 5 Discover Artificial Intelligence 110. doi:10.1007/s44163-025-00353-2.

4 Allison Koenecke and others, 'Tasks and Roles in Legal AI: Data Curation, Annotation, and Verification' (*arXiv preprint*, 2 April 2025) arXiv:2504.01349. doi:10.48550/arXiv.2504.01349.

also underline unresolved issues around privacy, scalability, and algorithmic bias that can translate into group-level disparities if left unmanaged.[5]

Building on this literature, Krištofík offers a court-focused synthesis.[6] He links the ECtHR's subjective and objective tests of judicial impartiality to AI governance, arguing that iterative, life-cycle audits and thorough documentation are the algorithmic counterparts of impartiality guarantees. He reads the CEPEJ Ethical Charter and its assessment tool as the practical route to detect and correct discriminatory patterns. As opacity undermines these safeguards, he calls for outcome monitoring, traceability, and public scrutiny.

Taken together, this record supports a bounded claim: fully impartial AI outputs are not attainable in adjudication, but bias can be constrained when tools are confined to narrow, assistive roles and subjected to evidence-based controls such as representativeness checks, source-grounded retrieval, transparency, logging, subgroup validation, and continuous monitoring. This is the backdrop for this study and motivates the hypothesis tested in the paper.

This paper treats the judiciary as a high-stakes, rights-sensitive environment and asks a focused question: If AI belongs in courts at all, where is the safe, lawful, and useful lane—especially with respect to bias?

To answer this, the study is framed around a few questions that keep bias and fair-trial guarantees at the centre:

1. What makes court AI biased in the first place—in the data, the tools, or the way people use them—and which parts can be reduced versus which are inherent?
2. What do today's European rules and soft-law standards actually allow or forbid for courts, and how do they safeguard reason-giving, equality of arms, and judicial independence?
3. What minimum safeguards are needed?
4. What do practice and cases teach us about the limits of AI in judging and workable remedies?
5. Where is the practical value, especially for Ukraine, and how to keep AI reliable?

5 Lauren E Kois and Preeti Chauhan, 'Criminal Responsibility: Meta-Analysis and Study Space' (2018) 36(3) Behavioral Sciences & the Law 276. doi:10.1002/bsl.2343; Oksana Kaplina and others, 'Application of Artificial Intelligence Systems in criminal Procedure: Key Areas, Basic Legal Principles and Problems of Correlation with Fundamental Human Rights' (2023) 6(3) Access to Justice in Eastern Europe 147. doi:10.33327/AJEE-18-6.3-a000314.

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

2 METHODOLOGY

This study combines legal analysis with technical analysis and targeted case studies, proceeding deliberately from human bias to LLM bias and then to legal guardrails, so the analysis remains anchored in courts' existing fair-trial obligations.

The analysis begins with an examination of human and judicial bias, demonstrating that, to some extent, court judgments themselves reflect these distortions. Bias in LLMs is described as stemming from next-token prediction rather than legal reasoning, with distortions introduced by training data, modelling choices, prompts, context, and interface and automation dynamics.

Subsequently, the relevant hard- and soft-law corpus, together with court-level practice, is elaborated. Empirically, two high-impact case studies (*COMPAS/Loomis and Ewert v. Canada*) are examined as they expose group-level disparities, transparency limits, and judicial remedies that can be generalised to LLM-assisted workflows.

The scope is intentionally normative-operational rather than experimental. Our contribution is a governance blueprint that connects specific LLM failure modes to enforceable legal duties and operational safeguards; empirical validation of any specific tool remains for jurisdiction-level pilots under the outlined controls.

3 BIAS IN THE JUDICIARY: HUMAN, ALGORITHMIC, AND INSTITUTIONAL DIMENSIONS

3.1. A Human Bias

Human judgment is not a neutral measurement device. Under cognitive load, time pressure, or uncertainty, we rely on heuristics[7] that introduce predictable distortions. An early figure (a plea offer, a prosecutor's recommendation, the statutory maximum) anchors subsequent assessments, tugging them toward itself even when arbitrary. Once a working hypothesis forms ("the defendant is likely culpable"), confirmation bias[8] skews attention and memory toward supportive evidence while discounting disconfirming cues, including alternative readings of forensic or eyewitness material. Availability and representativeness replace base-rate reasoning with vivid exemplars and "fitting" crime scripts, inflating perceived likelihoods. Stereotyping means that social cues such as race, age, accent, disability, or socioeconomic status trigger learned associations that subtly shape credibility and

7 Amos Tversky and Daniel Kahneman, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185(4157) Science 1124. doi:10.1126/science.185.4157.1124.

8 Raymond S Nickerson, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises' (1998) 2(2) Review of General Psychology 175. doi:10.1037/1089-2680.2.2.175.

dangerousness judgments, often without conscious endorsement.[9] Knowing outcomes *ex post* (e.g., that harm occurred) fosters hindsight and outcome bias, making events appear more foreseeable and negligence more apparent than they were prospectively. Finally, affective and situational factors: fatigue, time of day, docket pressure, and framing effects (gain vs. loss) systematically shift decisions.[10] Together, these mechanisms operate reliably enough to be mapped and anticipated, which is precisely why courts must treat them as designable risks rather than random chance.

3.2. Judicial Bias: Sources and Safeguards

Impartiality is both a duty and a presumption of judicial office. Judges must approach each case with "an open mind that is free of prejudgment and prejudice."[11] Public confidence in the courts rests on the belief that judges decide with unwavering impartiality. That confidence, in turn, depends on judges' capacity to anticipate and curb the predictable biases of human decision-making.

Scholars commonly distinguish two forms of judicial bias: actual and apprehended (apparent).[12] Actual bias describes a decision-maker who has prejudged the matter or closed their mind, so that relevant, admissible evidence is unlikely to move them—often due to interests, prior conduct or associations, or exposure to extraneous information. As courts are reluctant to probe a judge's inner thinking, the evidential threshold for actual bias is high. By contrast, apprehended bias applies an objective, observer-focused test: Would a fair-minded, informed observer see a real possibility that the judge is not impartial? This approach protects public confidence by focusing on the circumstances that reasonably threaten neutrality rather than proof of subjective prejudice. On this account, bias is not limited to hostility or animus; it includes structural and informational influences capable of undermining open-minded adjudication.

Equality and non-discrimination sharpen the point. Where bias interacts with protected characteristics (sex, race/ethnicity, disability, age, religion, etc.), it risks unjustified disparate treatment (different outcomes for similar facts) or unjustified disparate impact (systematically higher error or burden rates for a group). Even if no one intends discrimination, the law asks whether a differential is relevant and proportionate. As many biases are predictable, they are foreseeable risks that public authorities have a duty to prevent.

9 Patricia G Devine, 'Stereotypes and Prejudice: Their Automatic and Controlled Components' (1989) 56(1) Journal of Personality and Social Psychology 5. doi:10.1037/0022-3514.56.1.5.

10 Shai Danziger, Jonathan Levav and Liora Avnaim-Pesso, 'Extraneous Factors in Judicial Decisions' (2011) 108(17) Proceedings of the National Academy of Sciences 6889. doi:10.1073/pnas.1018033108.

11 Matthew Groves, 'The Rule Against Bias' (2009) 39 Hong Kong Law Journal 486.

12 Gary Edmond and Kristy A Martire, 'Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making' (2019) 82(4) Modern Law Review 633. doi:10.1111/1468-2230.12424.

Crucially, judges' biases are often reflected in their judgments and sentencing patterns; when those texts and outcomes are reused as training data, court-facing AI can learn and reproduce the same skews, creating feedback loops. Governance must therefore address both human decision-making and the datasets and models that learn from it.

3.3. Automation Bias

Having outlined human and judicial bias, the study turns to the tools increasingly used around courts: large language models (LLMs). Generative AI is software capable of creating, or generating, various media based on data it has observed in the past and influenced by what people consider pleasing and accurate outputs. More broadly, LLMs are machine learning models trained on large amounts of linguistic data.[13] LLMs do not "know" law; they surface patterns from data. LLMs operate on numbers, not words: input text is tokenised into subword units drawn from a fixed vocabulary, and the model learns statistical relationships among these tokens. Most LLMs are autoregressive, which means that given prior tokens, they predict the next, and a transformer's output is a probability distribution over the vocabulary; one token is then selected by sampling, with a tunable degree of randomness that trades creativity against consistency. As training optimises next-token prediction, not legal reasoning *per se*, capabilities track distributional familiarity: performance is strongest on frequent, well-represented patterns in the training corpus and degrades on genuinely novel tasks.[14] This technical profile underlies both the promise and bias risks of court-facing LLMs.

A court-facing LLM follows a simple life-cycle, and each stage has a characteristic bias risk. It begins with problem framing (what task the tool is meant to help with and which outcomes count as "good"). Choices can pre-tilt results. Next is data building from judgments, legal framework: some parties or case types are barely represented in the record, so the model "learns" mostly from the majority, when anonymisation and redaction is uneven, it may strip out key facts that explain a result but leave clues like postcode, employer, or school and that directs the LLM in a certain way. Finally, case metadata (headnotes, outcome codes) is sometimes incorrect or inconsistently applied. Together, these features teach the system a distorted version of reality and risk perpetuating past inequities. During model adaptation (fine-tuning on legal text), small or skewed in-house datasets may overfit majority writing styles or common fact patterns, sidelining minority languages or rare claims. If the system uses retrieval, ranking that over-weights highly cited or older courts can bias what sources the model sees. In generation, prompt wording and decoding settings favour confident, template-like answers that can normalise stereotypes or default outcomes. Evaluation often misses bias when it reports only average accuracy rather than subgroup

13 Edward Raff, Drew Farris and Stella Biderman, *How Large Language Models Work* (Manning 2025).

14 Dilyan Grigorov, *Introduction to Python and Large Language Models* (Apress 2024). doi:10.1007/979-8-8688-0540-0.

performance or citation fidelity. After deployment, use and interface create an automation bias: time pressure, one-click acceptance, and a lack of disclosure encourage uncritical adoption of AI-generated text. Finally, weak monitoring allows feedback loops, as AI-drafted language re-enters future datasets and hardens the same skew.[15]

4 AI AND LLMS IN COURTS

4.1. Deployment Models and Practices in the Judiciary

Courts are beginning to deploy large language models as assistive tools, not decision-makers. In Portugal, the justice tech agency is piloting AI-based anonymisation of judgments and a "virtual judge assistant" (e.g., STJ's IRIS) for drafting support—uses that require gold-standard tests, human QA, and subgroup error checks. Catalonia has trialled a generative-AI aid for repetitive commercial rulings, framed as assistive only, with disclosure, independent judicial reasons, and "cite-or-abstain" settings to curb hallucinations.[16]

Nationally, Spain permits summarising/drafting support but forbids AI from issuing final decisions. Spain's justice-sector AI policy pivots on two linked duties. First, transparency, impartiality and fairness: systems must be accessible, understandable and auditable. To balance IP with accountability, access to design information should be enabled, along with FAT records (Fairness, Accuracy, Transparency), so that bias can be detected and the interests of justice prevail. Second, prevention of bias and discrimination: algorithms are to undergo periodic evaluations to identify and correct biases arising from training data or model design, with safeguards to protect rights and avoid perpetuating structural injustices. Quality control and auditing scale with risk: where tools might affect the exercise of jurisdiction and judicial independence, oversight lies with the General Council of the Judiciary (CGPJ) under the LOPJ, including "algorithmic surveillance" (collection and analysis of system data/outputs to assess performance, detect bias or errors, and ensure

15 Emily M Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models be Too Big?' (FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency) 610. doi:10.1145/3442188.3445922; Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (*arXiv preprint*, 12 July 2022) arXiv:2108.07258. doi:10.48550/arXiv.2108.07258; Xuezhi Wang and others, 'Self-Consistency Improves Chain of Thought Reasoning in Language Models' (*arXiv preprint*, 7 March 2023) arXiv:2203.11171. doi:10.48550/arXiv.2203.11171; Erik Jones and Jacob Steinhardt, 'Capturing Failures of Large Language Models Via Human Cognitive Biases' (NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems) 11785.

16 'Cataluña Prueba Las Sentencias Judiciales Escritas Con Inteligencia Artificial' (*RTVE*, 21 March 2025) <https://www.rtve.es/noticias/20250321/cataluna-sentencias-judiciales-inteligencia-artificial/16502032.shtml> accessed 1 September2025.

transparency and responsibility). The policy explicitly defines model bias: "GenAI tools incorporate any bias from the data sets used to train them… the model output may make systematic errors or favour certain groups, leading to unfair or discriminatory results."[17]

The Netherlands judiciary's AI programme is deliberately incremental: pilots prioritise low-risk, public-facing tools (e.g., website chatbots and internal knowledge search) while standards for court-facing uses are built out. Experiments are sandboxed, use approved corpora, and keep data inside the judiciary infrastructure; logs are retained, and DPIAs/algorithm registers are prepared in anticipation of the EU AI Act duties. Any generative features are framed as assistive (template completion, style harmonisation), with "cite-or-abstain" settings, role-based access, and explicit prohibitions on AI determining outcomes. Staff training covers automation-bias risks and how to read model/system cards; procurement templates require access to documentation and audit rights.[18]

In Singapore, reported pilots in small-claims contexts use gen-AI for summarisation and user guidance, coupled with retrieval from official sources, edit logging, and on-screen disclosures. Outputs are advisory and must be reviewed; acceptance includes friction (checklists/justification prompts), and periodic sampling by registrars checks citation fidelity, neutrality, and subgroup performance. Interfaces are tuned to reduce anchoring (e.g., withholding suggestions until a human outline exists), and prompts/outputs are retained to create an appeal-ready record.[19]

At Brazil's Superior Labour Court (TST), BEM-TE-VI is an AI-assisted triage and case-management tool (not an LLM) used since 2018 to screen incoming labour appeals. It clusters cases by theme, flags timeliness issues, and supports "virtual triage" and routing to analyst teams, speeding cabinet workflows while leaving adjudication to judges.[20]

17 Ministry of the Presidency, Justice and Parliamentary Relations (Spain), *Policy on the Use of Artificial Intelligence in the Administration of Justice* (Preliminary version, MPJRC 2024).

18 Council for the Judiciary (Netherlands), 'Responsible and Innovative: AI for a fair Dutch Judicial System' (*de Rechtspraak*, 2024) <https://www.rechtspraak.nl/Organisatie-en-contact/innovatie-binnen-de-rechtspraak/Paginas/AI-Decree.aspx> accessed 1 September 2025; Ibrahim Jabri, 'The Use of Artificial Intelligence in the Dutch Courtroom' (Master thesis, TU Delft 2022); 'Rotterdam Court Tests Artificial Intelligence as Writing Aid in Criminal Verdicts' (*NL Times*, 30 March 2025) <https://nltimes.nl/2025/03/30/rotterdam-court-tests-artificial-intelligence-writing-aid-criminal-verdicts> accessed 1 September 2025.

19 Supreme Court of Singapore, 'Guide on the Use of Generative Artificial Intelligence Tools by Court Users' (*Singapore Courts*, 2024) <https://www.judiciary.gov.sg/docs/default-source/news-and-resources-docs/guide-on-the-use-of-generative-ai-tools-by-court-users.pdf> accessed 1 September 2025; Maryam Akhlaghi, 'Navigating AI in the Courts: Lessons from Singapore, South Korea, and Australia' (*Laboratoire de Cyberjustice*, 21 July 2025) <https://www.cyberjustice.ca/en/2025/07/21/navigating-ai-in-the-courts-lessons-from-singapore-south-korea-and-australia/> accessed 1 September 2025.

20 Marcos de Moraes Sousa and Thiago Maia Sayão de Moraes, 'Institutionalization of Innovation: The Perception of Actors in the Brazilian Labor Court with Artificial Intelligence' (2025) 27(142) Revista Jurídica da Presidência 293. doi:10.20499/2236-3645.RJP2025v27e142-3215.

Taken together, courts are experimenting with a range of AI tools (from anonymisation and public-facing chatbots to retrieval-grounded drafting aids and, in some systems, triage/classification) deployed as assistive rather than adjudicative technologies. This diversity of use matters as each tool exposes different bias pathways. The next section traces concrete episodes where such biases surfaced in practice, showing how they produced legal and procedural problems—and what those failures teach about safer design and governance.

4.2. Problem Cases: Bias, Opacity, and Due-Process Risks

The reality of bias in court-facing algorithms is now widely documented, and a fast-growing literature dissects dozens of deployments across jurisdictions. To keep this section focused, two illustrative cases that have shaped the debate and judicial practice are highlighted: the U.S. experience with the COMPAS risk tool (and *State v. Loomis*), and Canada's *Ewert v. Canada*. Together, they span different legal systems and decision points (sentencing; security classification/parole), expose distinct bias mechanisms (disparate false-positive rates; lack of subgroup validity for Indigenous offenders), and show the courts' emerging remedies—assistive-only use, independent judicial reasons, transparency about model limits, and requirements for validation and ongoing monitoring. These paired case studies anchor the abstract concern about "AI bias" in concrete adjudicative settings and supply principles drawn on throughout the paper:

1) COMPAS (Correctional Offender Management Profiling for Alternative Sanctions by Northpointe/Equivant) is a proprietary risk-and-needs assessment used by many U.S. jurisdictions at pretrial, sentencing, and supervision stages. It produces scales such as general and violent recidivism risk from questionnaires and criminal-history data; race is not an input, while sex is used for separate norms. The vendor describes COMPAS as empirically developed and validated across jurisdictions, but provides no details on its model design or training data.[21]

In 2016, ProPublica obtained Broward County COMPAS scores and two-year re-arrest outcomes (18,610 people) and reported higher false-positive rates for Black defendants, prompting a national debate.[22] The algorithm's racial bias caused the parole denial rate of African American defendants to be significantly higher than that of white defendants, revealing the falsity of "technological neutrality." The algorithm not only replicates structural discrimination in human society but also may give rise to more hidden

21 EquiVant Supervision, 'Solutions: Risk & Needs Assessments' (*equiVant Supervision*, 2024) <https://equivant-supervision.com/solutions/risk-needs-assessments/> accessed 1 September 2025.

22 Jeff Larson and others, 'How We Analyzed the Compas Recidivism Algorithm' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> accessed 7 September 2025.

"algorithmic native bias" through its inherent probability model.[23] Northpointe (now Equivant) disputed the analysis, arguing the tool was well-calibrated across groups; subsequent work[24] showed lay predictions can match COMPAS accuracy, sharpening questions about fairness trade-offs and transparency.

In *State v. Loomis*,[25] the Wisconsin Supreme Court allowed the use of COMPAS at sentencing but only with stringent cautions: the score may not be determinative; courts must give independent reasons; and presentence reports must warn of its limitations (proprietary method, group-based design, potential bias). The court acknowledged due-process concerns tied to secrecy but found use permissible under these constraints; the U.S. Supreme Court denied certiorari.

2) In *Ewert v. Canada* (2018 SCC 30),[26] an Indigenous prisoner challenged Correctional Service Canada's use of actuarial risk tools (e.g., PCL-R, VRAG) developed and validated mainly on non-Indigenous populations but applied to inform security classification, programming, and parole. The Supreme Court held CSC breached its statutory duty to ensure information is "as accurate, up to date and complete as possible," as credible evidence of a foreseeable risk of cultural bias existed. CSC had not taken reasonable steps to validate the tools for Indigenous offenders. The Court did not ban actuarial instruments; it required subgroup validation, transparent documentation of the data, populations, and metrics, and ongoing monitoring for disparate error—insisting that any scores be assistive, not determinative, and accompanied by independent reasons. For judicial AI more broadly, Ewert's lesson is straightforward: no black-box deference when fundamental rights are at stake—prove subgroup fairness first, keep humans in control, and continuously audit for bias.

3) Another example comes from the U.S., where two federal district judges were forced to rescind and rewrite rulings after lawyers discovered that the decisions contained non-existent quotations and party descriptions, later traced to the undisclosed use of generative AI by chambers staff. In one case, a New Jersey judge's intern relied on ChatGPT for legal research, producing an order with fabricated case quotations; in another, a Mississippi judge's clerk used the LLM-based tool Perplexity to draft a

23 Li Jialin, 'Exploring Bias Formation Mechanisms in Legal LLMs from a Cognitive Science Perspective' in Xiaofeng Meng and others (eds), *Big Data and Social Computing: BDSC 2025* (Springer 2025) 290. doi:10.1007/978-981-95-0880-8_24.

24 Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4(1) Science Advances eaao5580. doi:10.1126/sciadv.aao5580; Tim Brennan and William Dieterich, 'Correctional Offender Management Profiles for Alternative Sanctions (COMPAS)' in Jay P Singh and others (eds), *Handbook of Recidivism Risk/Needs Assessment Tools* (Wiley-Blackwell 2018) 49. doi:10.1002/9781119184256.ch3.

25 *State v Loomis* 2016 WI 68, 371 Wis 2d 235, 881 NW2d 749 (Wis) <https://case-law.vlex.com/vid/state-v-loomis-no-888404547> accessed 6 September 2025.

26 *Ewert v Canada* 2018 SCC 30 [2018] 2 SCR 165 (SCC) <https://decisions.scc-csc.ca/scc-csc/scc-csc/en/item/17133/index.do> accessed 6 September 2025.

temporary restraining order that referenced parties and allegations unrelated to the actual dispute. Only after these errors were called out did the judges acknowledge AI involvement in letters to the Administrative Office of the U.S. Courts, prompting criticism from scholars and a Senate inquiry into the judiciary's AI practices.[27]

5 BIAS CONTROLS IN JUDICIAL AI: HARD- AND SOFT-LAW APPROACHES

Fundamental fair-trial guarantees constrain the use of AI in courts:[28] the right to a fair hearing before an independent and impartial tribunal (ECHR, Art. 6) and the right to an effective remedy and a fair trial (EU Charter, Art. 47). Any AI-assisted workflow must therefore preserve reason-giving, equality of arms, and judicial independence. Where AI tools risk obscuring legal reasoning, skewing access to information, or shifting decision-making authority away from the judge, they jeopardise these guarantees. The next subsection traces how these guarantees are operationalised across key regulatory and soft-law instruments governing AI in the judiciary, showing how fair-trial rights have been translated into concrete requirements and limits for court-facing AI.

Under *the EU Artificial Intelligence Act*,[29] AI systems used by or on behalf of judicial authorities to assist in researching/interpreting facts and law or in applying law to facts are explicitly classified as high-risk (Annexe III, pt. 8(a)). High-risk systems must satisfy the Act's controls: risk management (Art. 9), high-quality data and data-governance (Art. 10), technical documentation (Art. 11), logging/record-keeping (Art. 12), transparency to deployers (Art. 13), human oversight (Art. 14), and accuracy/robustness/cybersecurity (Art. 15), alongside post-market monitoring by providers (Art. 72). Separately, certain practices are banned outright—most notably social scoring that leads to detrimental or unfavourable treatment (Art. 5(1)(c)).

The EU High-Level Expert Group's Trustworthy AI framework[30] tackles bias by hard-wiring it across all seven requirements and operationalising it through the ALTAI self-assessment.[31]

27 Justin Henry, 'Judges Admit to Using AI After Made-Up Rulings Called Out (1)' (*Bloomberg Law*, 23 October 2025) <https://news.bloomberglaw.com/business-and-practice/judges-called-out-for-nonfactual-rulings-admit-to-use-of-ai> accessed 23 October 2025.

28 Iryna Izarova and others, 'Advancing Sustainable Justice Through AI-Based Case-Law Analysis' (2024) 7(1) Access to Justice in Eastern Europe 127. doi:10.33327/AJEE-18-7.1-a000123.

29 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) <http://data.europa.eu/eli/reg/2024/1689/oj> accessed 1 September 2025.

30 EC High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (EU Publications Office 2019). doi:10.2759/346720.

31 EC High-Level Expert Group on Artificial Intelligence, *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment* (EU Publications Office 2020). doi:10.2759/002360.

In practice: diversity, non-discrimination and fairness demand representative datasets, scrutiny of sensitive/proxy features, explicit fairness goals, and subgroup error analysis; human agency and oversight is used to counter automation bias with real escalation paths and decision rights for judges; technical robustness and safety requires stress-testing for distribution shift and disparate error rates; privacy and data governance links lawful bases with data-minimisation that avoids proxy discrimination and documents provenance; transparency asks for traceability, explanation and disclosure of limitations so biased patterns can be detected and contested; societal and environmental well-being pushes impact analysis on vulnerable groups; and accountability requires roles, logs, auditability and redress. ALTAI translates this into concrete checks for courts and vendors—e.g., whether training/validation sets are representative, which fairness metrics and trade-offs were chosen (and why), how explanations support bias discovery, what post-deployment monitoring and complaint channels exist, and whether a system should not be deployed if bias cannot be meaningfully mitigated. As procurement and deployment criteria, these items allow judiciaries to require bias audits, subgroup performance reports, mitigation plans, oversight triggers, and ongoing monitoring before any tool is admitted into courtroom workflows.

The Court of Justice of the European Union's AI Strategy treats bias as a first-order risk and organises its entire approach around preventing it without sacrificing judicial independence.[32] It couples strong human control with rigorous risk management that front-loads bias analysis through dataset provenance checks, representativeness testing, and documentation of known limitations. Explainability and traceability are mandated to detect, audit, and correct skewed patterns rather than have them silently propagate into legal reasoning. New systems are introduced only via cautious pilots under close user supervision, which limits automation bias and enables empirical monitoring of subgroup error rates before any wider rollout. Taken together, these elements offer a practical template for national courts: keep AI strictly assistive, design for bias detection and correction from the outset, and make continued use contingent on demonstrable fairness and control.

At the Council of Europe level, the European Ethical Charter on the use of AI in judicial systems (2018) provides five justice-specific guardrails: respect for fundamental rights; non-discrimination with attention to data quality; quality and security; transparency, impartiality and auditability; and the "under user control" principle that keeps AI subordinate to human decision-makers. Crucially, the Charter treats bias as a foreseeable risk: data must be representative and of verified quality; systems should not enable deterministic or black-box outcomes; and design must enable external scrutiny so discriminatory effects can be detected and corrected.[33] Building on this, CEPEJ has shifted

32 Court of Justice of the European Union, *Artificial Intelligence Strategy* (Directorate-General for Information 2023).

33 CEPEJ, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (Council of Europe 2019).

from principles to practice. Its Assessment Tool (2023, updated 2025) operationalises the Charter through concrete checks for discriminatory risk arising from data selection, annotation or training choices, unclear criteria or hidden weightings, and it prescribes mitigation steps and documentation.[34] Where training data or source code cannot be audited, or where bias cannot be ruled out, the guidance counsels against deployment in adjudicative contexts. Complementary 2024 guidance on the online publication of judicial decisions seeks to improve the quality,[35] accessibility, and traceability of the datasets that feed court-facing AI, while the AI Advisory Board's 2025 reporting emphasises regular evaluation and inventory of tools in use.[36]

In parallel, the Council of Europe Framework Convention on Artificial Intelligence (2024) is the first binding, horizontal treaty requiring that AI lifecycle activities comply with human rights, democracy and the rule of law.[37] It expressly recognises that AI can create or aggravate inequalities and therefore imposes duties of risk and impact management, documentation, testing and ongoing monitoring to prevent and mitigate discriminatory outcomes, alongside safeguards for judicial independence and effective remedies for affected persons. Finally, CCJE Opinion No. 26 (2023) sets courtroom-specific guardrails: technology may assist but must not replace adjudication; decision-making remains a human judicial act; design and operation must be non-discriminatory, transparent and intelligible; judges must retain oversight of procurement, design and control; and "judge-analytics" tools that profile or predict individual judges' behaviour are out of bounds.[38] Taken together, these instruments converge on a clear standard: court-facing AI is acceptable only in an assistive, transparent, and auditable role under judicial control, and is not deployed where bias cannot be meaningfully assessed or mitigated.[39]

In EU Member States, standalone AI framework laws are generally unnecessary, as the EU Artificial Intelligence Act already regulates the field and applies directly. What matters, therefore, is how national judiciaries are translating that baseline into practice. Two French

34 CEPEJ, *Assessment Tool for the Operationalisation of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (CEPEJ(2023)16 final, Council of Europe 2023).

35 CEPEJ, *Guidelines for the Online Publication of Judicial Decisions Aiming at Furthering Legal Knowledge* (CEPEJ(2024)9, Council of Europe 2024).

36 CEPEJ, CEPEJ-GT-Cyberjust and CEPEJ-AIAB, *First Artificial Intelligence Advisory Board (AIAB) Report on the Use of Artificial Intelligence (AI) in the Judiciary* (CEPEJ-AIAB(2024)4Rev5, Council of Europe 2025).

37 Council of Europe, Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (5 September 2024) [2024] CETS 225.

38 CCJE Opinion No 26 (2023) Moving Forward—The Use of Assistive Technology in the Judiciary (1 December 2023) <https://rm.coe.int/ccje-opinion-no-26-2023-final/1680adade7> accessed 1 September 2025.

39 Tetiana Tsuvina, 'Artificial Intelligence Technologies in the Judiciary: European Standards and Ukrainian Practice' (2025) 139(2) Foreign Trade: Economics, Finance, Law 4. doi:10.31617/3.2025(139)03.

examples stand out. First, the Cour de cassation's joint guidelines for magistrates and lawyers cast generative AI as strictly assistive and treat bias as a primary, foreseeable risk.[40] They require human oversight, independent verification of citations, auditable workflows, user training on bias mechanisms, and the readiness to suspend tools that exhibit discriminatory drift. Second, the Paris Commercial Court's Charter approaches AI "in the spirit of French humanism": it keeps judges responsible for decisions, limits AI to supportive functions, mandates transparency to parties, provides for training and monitored deployment, and centres fairness and non-discrimination.[41]

Ukraine's current AI policy is anchored in cross-sector instruments led by the Ministry of Digital Transformation. The Governmental Concept for AI Development set the baseline for promoting AI R&D, data infrastructure, skills, and ethical principles (transparency, non-discrimination).[42] Still, it did not create sector-specific rules for courts or specify bias controls in adjudication. The Ministerial White Paper on AI Regulation in Ukraine advances regulatory options aligned with EU law, signalling convergence with the EU Artificial Intelligence Act's risk-based model and fundamental rights safeguards.[43] It treats bias as a general risk (fairness, non-discrimination) and anticipates sectoral tailoring, yet it does not provide judiciary-specific operational guidance. The in-progress AI Strategy 2030 (public consultation in 2025) similarly prioritises EU alignment, data/standards, sandboxes, and high-risk governance,[44] but contains no dedicated chapter on the justice sector and no bespoke bias-mitigation regime for court-facing tools.

The research shows that bias in judicial AI is well recognised within the current framework. Hard law (the EU AI Act's high-risk controls and data-protection duties) and soft law (CEPEJ Charter and Assessment Tool, the Council of Europe Framework Convention,

40 Conseil consultatif conjoint de déontologie de la relation magistrats–avocats, 'Intelligence Artificielle Generative: Et vigilance déontologique dans l'exercice professionnel des magistrats et des avocats et de leurs équipes' (*Cour de cassation and Conseil national des barreaux*, 24 October 2025) <https://www.courdecassation.fr/toutes-les-actualites/2025/10/24/magistrats-et-avocats-lere-de-lia-faire-vivre-une-deontologie> accessed 1 September 2025.

41 Vincent Fauchoux, 'From the Paris Edict of 1563 to the AI Act of 2023: How the Paris Commercial Court Regulates Artificial Intelligence in the Spirit of French Humanism' (*DDG Avocats*, 11 September 2025) <https://www.ddg.fr/actualite/from-the-paris-edict-of-1563-to-the-ai-act-of-2023-how-the-paris-commercial-court-regulates-artificial-intelligence-in-the-spirit-of-french-humanism> accessed 1 September 2025..

42 Order of the Cabinet of Ministers of Ukraine No 1556-p 'On the Approval of the Concept for the Development of Artificial Intelligence in Ukraine' (2 December 2020) [in Ukrainian] <https://zakon.rada.gov.ua/go/1556-2020-%D1%80> accessed 1 September 2025.

43 Ministry of Digital Transformation of Ukraine, *White Paper on Artificial Intelligence Regulation in Ukraine: Vision of the Ministry of Digital Transformation of Ukraine* (Version for Consultation, Reforms Delivery Office of the Cabinet of Ministers of Ukraine 2024).

44 Ministry of Digital Transformation of Ukraine, 'The Future of AI in Ukraine Starts Here—Join the Survey' (*Digital State UA*, 27 June 2025) <https://digitalstate.gov.ua/news/govtech/the-future-of-ai-in-ukraine-starts-here-join-the-survey> accessed 1 September 2025.

HLEG "Trustworthy AI"/ALTAI, the CJEU's AI Strategy, and court charters) converge on the same toolkit: rigorous data governance and representativeness checks; meaningful human oversight to counter automation bias; explainability and logging for audit; disclosure to parties; red-lines with a readiness not to deploy where bias cannot be mitigated.

Given these risks, courts are confining AI to assistive roles (research, retrieval, drafting hygiene), not outcome setting or sentencing. Even so, doing assistive AI properly is resource-intensive: it requires upfront investment in secure infrastructure, dataset curation, user training, governance and accountability structures, and then continuous bias monitoring, audits, incident handling, and periodic re-validation over the tool's life cycle.

6 CONCLUSIONS

Bias is not a fringe risk in judicial AI; it is structural. As this paper showed, large language models inherit training-data bias, exhibit inductive/model bias, and are shaped by contextual, retrieval/index, and interface (automation) biases. The case studies (COMPAS/Loomis; Ewert) demonstrate that these mechanisms translate into measurable, group-level disparities and due-process concerns. "Technological neutrality" is therefore a myth: without explicit controls, court-facing AI will reflect and sometimes amplify the unevenness of the legal record from which it learns.

The current hard-and-soft-law framework converges on the same answer: assistive, not determinative use under real human control. The EU AI Act classifies administration-of-justice systems as high-risk and requires data governance, transparency, logging, human oversight, robustness, and post-market monitoring; Council of Europe instruments (CEPEJ Charter/Assessment Tool, Framework Convention) and judicial policies (e.g., CJEU strategy; French court guidance) operationalise these requirements in practice. In short: no black boxes in adjudication; validate on subgroups; disclose limitations; keep an auditable trail; and be prepared not to deploy where bias cannot be meaningfully mitigated.

Done properly, even assistive AI is resource-intensive. It demands secure infrastructure; curated, traceable datasets; gold-standard evaluation (including subgroup performance); procurement with audit rights; user training; and continuous monitoring with a kill-switch. For many systems, especially resource-constrained judiciaries, the rational path is to target low-risk, high-yield uses (retrieval, summarisation, drafting hygiene, anonymisation) and invest first in data quality and basic digitalisation, without which AI will underperform, and budgets will be wasted.

For the Ukrainian realities, in the near term, the most valuable and realistic uses are assistive ones that complement existing national infrastructure rather than replace it. Concretely: (i) case retrieval across Ukrainian jurisprudence, with AI used to surface and cluster relevant Supreme Court lines (not to opine on outcomes)—noting the Supreme Court's increasingly

refined public database; (ii) norm retrieval and collation from the Codes and secondary legislation, recognising that commercial services (e.g., Liga) already deliver strong coverage, so AI should aggregate and cross-check rather than duplicate; (iii) checking judicial drafts for clarity, structure, internal consistency, and citation fidelity; and (iv) "missed-norms" prompts in the Legal Basis section, where a RAG-first tool compares the draft against an authoritative, up-to-date canon of norms used in similar cases and flags omitted but commonly applied provisions for the judge's review. All of these should run inside a logged, court-controlled environment, expose their sources, and leave the final legal reasoning and responsibility solely with the judge.

However, a further constraint is financial. As there are no public disclosures on the cost of judicial-AI deployments anywhere, estimates must be inferred from comparable high-stakes public-sector systems, where initial deployment typically requires €10–20 million, with €1–3 million annually for maintenance, monitoring, and audit.[45] On this basis, the current budget of the Ukrainian judiciary cannot realistically absorb such expenditures, particularly given ongoing reconstruction demands and the still-uncertain marginal benefit that LLM tools would bring to judges' and court staff's workload.

REFERENCES

1. Akhlaghi M, 'Navigating AI in the Courts: Lessons from Singapore, South Korea, and Australia' (*Laboratoire de Cyberjustice*, 21 July 2025) <https://www.cyberjustice.ca/en/2025/07/21/navigating-ai-in-the-courts-lessons-from-singapore-south-korea-and-australia/> accessed 1 September 2025
2. Bender EM and others, 'On the Dangers of Stochastic Parrots: Can Language Models be Too Big?' (FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency) 610. doi:10.1145/3442188.3445922
3. Bommasani R and others, 'On the Opportunities and Risks of Foundation Models' (*arXiv preprint*, 12 July 2022) arXiv:2108.07258. doi:10.48550/arXiv.2108.07258
4. Brennan T and Dieterich Wi, 'Correctional Offender Management Profiles for Alternative Sanctions (COMPAS)' in Singh JP and others (eds), *Handbook of Recidivism Risk/Needs Assessment Tools* (Wiley-Blackwell 2018) 49. doi:10.1002/9781119184256.ch3
5. Chen Q, 'Improving the Trial Efficiency of Criminal Cases with the Assistance of Artificial Intelligence' (2025) 5 Discover Artificial Intelligence 110. doi:10.1007/s44163-02500353-2

45 Vincenzo Piccolo, 'Cost of Implementing AI in Healthcare in 2025' (*Callin.io*, 2025) <https://callin.io/cost-of-implementing-ai-in-healthcare/> accessed 6 September 2025.

6. Danziger S, Levav J and Avnaim-Pesso L, 'Extraneous Factors in Judicial Decisions' (2011) 108(17) Proceedings of the National Academy of Sciences 6889. doi:10.1073/pnas.1018033108

7. Devine PG, 'Stereotypes and Prejudice: Their Automatic and Controlled Components' (1989) 56(1) Journal of Personality and Social Psychology 5. doi:10.1037/0022-3514.56.1.5

8. Dressel J and Farid H, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4(1) Science Advances eaao5580. doi:10.1126/sciadv.aao5580

9. Edmond G and Martire KA, 'Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making' (2019) 82(4) Modern Law Review 633. doi:10.1111/1468-2230.12424

10. Fauchoux V, 'From the Paris Edict of 1563 to the AI Act of 2023: How the Paris Commercial Court Regulates Artificial Intelligence in the Spirit of French Humanism' (*DDG Avocats*, 11 September 2025) <https://www.ddg.fr/actualite/from-the-paris-edict-of-1563-to-the-ai-act-of-2023-how-the-paris-commercial-court-regulates-artificial-intelligence-in-the-spirit-of-french-humanism> accessed 1 September 2025.

11. Grigorov D, *Introduction to Python and Large Language Models* (Apress 2024). doi:10.1007/979-8-8688-0540-0

12. Groves M, 'The Rule Against Bias' (2009) 39 Hong Kong Law Journal 485

13. Henry J, 'Judges Admit to Using AI After Made-Up Rulings Called Out (1)' (*Bloomberg Law*, 23 October 2025) <https://news.bloomberglaw.com/business-and-practice/judges-called-out-for-nonfactual-rulings-admit-to-use-of-ai> accessed 23 October 2025

14. Izarova I and others, 'Advancing Sustainable Justice Through AI-Based Case-Law Analysis' (2024) 7(1) Access to Justice in Eastern Europe 127. doi:10.33327/AJEE-18-7.1-a000123

15. Jabri I, 'The Use of Artificial Intelligence in the Dutch Courtroom' (Master thesis, TU Delft 2022)

16. Jialin L, 'Exploring Bias Formation Mechanisms in Legal LLMs from a Cognitive Science Perspective' in Meng X and others (eds), *Big Data and Social Computing: BDSC 2025* (Springer 2025) 290. doi:10.1007/978-981-95-0880-8_24

17. Jones E and Steinhardt J, 'Capturing Failures of Large Language Models Via Human Cognitive Biases' (NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems) 11785

18. Kaplina O and others, 'Application of Artificial Intelligence Systems in criminal Procedure: Key Areas, Basic Legal Principles and Problems of Correlation with Fundamental Human Rights' (2023) 6(3) Access to Justice in Eastern Europe 147. doi:10.33327/AJEE-18-6.3-a000314

19. Koenecke A and others, 'Tasks and Roles in Legal AI: Data Curation, Annotation, and Verification' (*arXiv preprint*, 2 April 2025) arXiv:2504.01349. doi:10.48550/arXiv.2504.01349

20. Kois LE and Chauhan P, 'Criminal Responsibility: Meta-Analysis and Study Space' (2018) 36(3) Behavioral Sciences & the Law 276. doi:10.1002/bsl.2343

21. Krištofík A, 'Bias in AI (Supported) Decision Making: Old Problems, New Technologies' (2025) 16(1) International Journal for Court Administration 3. doi:10.36745/ijca.598

22. Larson J and others, 'How We Analyzed the Compas Recidivism Algorithm' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> accessed 7 September 2025

23. Moraes Sousa M and Moraes TMS, 'Institutionalization of Innovation: The Perception of Actors in the Brazilian Labor Court with Artificial Intelligence' (2025) 27(142) Revista Jurídica da Presidência 293. doi:10.20499/2236-3645.RJP2025v27e142-3215

24. Nickerson RS, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises' (1998) 2(2) Review of General Psychology 175. doi:10.1037/1089-2680.2.2.175

25. O'Neil C, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016)

26. Piccolo V, 'Cost of Implementing AI in Healthcare in 2025' (*Callin.io*, 2025) <https://callin.io/cost-of-implementing-ai-in-healthcare/> accessed 6 September 2025.

27. Raff E, Farris D and Biderman S, *How Large Language Models Work* (Manning 2025)

28. Tsuvina T, 'Artificial Intelligence Technologies in the Judiciary: European Standards and Ukrainian Practice' (2025) 139(2) Foreign Trade: Economics, Finance, Law 4. doi:10.31617/3.2025(139)03

29. Tversky A and Kahneman D, 'Judgment under Uncertainty: Heuristics and Biases' (1974) 185(4157) Science 1124. doi:10.1126/science.185.4157.1124

30. Wang X and others, 'Self-Consistency Improves Chain of Thought Reasoning in Language Models' (*arXiv preprint*, 7 March 2023) arXiv:2203.11171. doi:10.48550/arXiv.2203.11171

AUTHORS INFORMATION

Nataliia Mazaraki*
Dr.Sc. (Law), Prof., International, civil and commercial law, State University of Trade and Economics, Kyiv, Ukraine
Competition law, Max Planck Institut für Innovation und Wettbewerb, Munchen, Germany
n.mazaraki@knute.edu.ua
https://orcid.org/0000-0002-1729-7846
Corresponding author, responsible for conceptualization, research methodology, supervising and writing – original draft.

Dmytro Honcharuk
Master of Law, International, civil and commercial law, State University of Trade and Economics, Kyiv, Ukraine
d.honcharuk@knute.edu.ua
https://orcid.org/0009-0000-8468-6959
Co-author, responsible for data collection and writing – original draft.

Competing interests: No competing interests were disclosed. Any potential conflict of interest must be disclosed by authors.

Disclaimer: The authors declare that their opinion and views expressed in this manuscript are free of any impact of any organizations.

RIGHTS AND PERMISSIONS

EDITORS

Managing Editor – Mag. Yuliia Hartman. **English Editor** – Julie Bold.
Ukrainian language Editor – Mag. Liliia Hartman.

ABOUT THIS ARTICLE

Cite this article
Mazaraki N and Honcharuk D, 'Judicial AI and the Irreparable Bias Problem' (2025) 8(Spec) Access to Justice in Eastern Europe 1-21 <https://doi.org/10.33327/AJEE-18-8.S-c000159> Published Online 11 Dec 2025

DOI: https://doi.org/10.33327/AJEE-18-8.S-c000159

Keywords: *artificial intelligence, EU law, judiciary, access to justice, bias.*

DETAILS FOR PUBLICATION

Date of submission: 02 Nov 2025
Date of acceptance: 24 Nov 2025
Online First Publication: 11 Dec 2025
Last date of publication: December 2025
Whether the manuscript was fast tracked? - No
Number of reviewer report submitted in first round: 2 reports (2 external reviewers)
Number of revision rounds: 1 round

Technical tools were used in the editorial process:
Plagiarism checks - Turnitin from iThenticate https://www.turnitin.com/products/ithenticate/
Scholastica for Peer Review https://scholasticahq.com/law-reviews

AI DISCLOSURE STATEMENT

The corresponding author confirms that artificial intelligence tools (such as Grammarly or language editors) were used solely for minor proofreading and readability improvements. All content and intellectual contributions are original and authored by the listed researchers.

АНОТАЦІЯ УКРАЇНСЬКОЮ МОВОЮ

Тематичне дослідження

ШІ В СУДОЧИНСТВІ ТА ПРОБЛЕМА НЕВИПРАВНОЇ УПЕРЕДЖЕНОСТІ

Наталія Мазаракі* та Дмитро Гончарук

АНОТАЦІЯ

Вступ: *Суди дедалі частіше експериментують з великими мовними моделями (LLM) для таких завдань, як пошук правової інформації, допомога в складанні документів, анонімізація та сортування. Однак обіцянка ефективності стикається зі структурною проблемою: упередженістю. Судочинство, яке здійснюється людьми, вже відображає когнітивні та інституційні упередження; LLM, навчені на основі попередніх судових рішень та юридичних текстів, успадковують, а іноді й посилюють ці упередження. У цій статті ставиться конкретне питання: якщо штучний інтелект взагалі має місце в судах, то який шлях є безпечним, законним і корисним, особливо, з*

огляду на упередження? В основі дослідження – гарантії справедливого судочинства та нові регуляторні очікування.

Методи: *У роботі використовується поетапний аналіз, що ґрунтується на правових зобов'язаннях та враховує відповідні технічні характеристики. По-перше, відображаються джерела людської та судової упередженості, а також моменти, у яких LLM вводять або посилюють упередженість. По-друге, синтезуються жорсткі та м'які правові бар'єри, що стосуються контролю упередженості у сфері правосуддя. По-третє, дві показові судові справи — COMPAS/Loomis (США) та Еверт проти Канади — розглядаються, щоб продемонструвати, як диспропорції щодо певних груп людей, та непрозорість моделей можуть створювати ризики для належної правової процедури, та визначити засоби правового захисту, які можна перенести на робочі процеси, що підтримуються LLM. Нарешті, розроблено та застосовано операційний план для визначення низькоризикових та високоефективних способів використання для України.*

Результати та висновки: *Аналіз показує, що повністю неупереджені результати ШІ недосяжні в судовому розгляді; упередженість неможливо усунути, але її можна обмежити. Для України раціональним шляхом є спочатку інвестувати в управління даними, безпечну інфраструктуру, можливості оцінювання та закупівлі послуг ШІ з правами аудиту, а також обмежити використання ШІ пошуком, зіставленням норм, редагуванням та підказками про «пропущені норми». Це дослідження є внеском у план управління, який пов'язує конкретні режими збоїв LLM з юридичними обов'язками, що підлягають виконанню, та практичними гарантіями, пропонуючи судам надійний, орієнтований на вирішення питання упередженості шлях для ШІ, який покращує обслуговування, водночас дотримуючись правових засад.*

Ключові слова: *штучний інтелект, законодавство ЄС, судова система, доступ до правосуддя, упередженість.*