Original Article

Received/Accepted Dates

12.01.2026/15.02.2025

DOI 10.52096/usbd.10.42.09

International Journal of Social Sciences

IJSS

Uluslararası Sosyal Bilimler Dergisi

www.sobider.net e-ISSN: 2548-0685

Araştırma Makalesi

Konu: Hukuk

Reconfiguring Judicial Legitimacy: Articial Intelligence, Cognitive Bias, and The Rise of Algorithmic Authority

Assoc. Prof. Dr. Fikret ERKAN

fikreterkan@hotmail.com, ORCID: 0009-0004-9462-7572

Abstract

The increasing integration of artificial intelligence (AI)–based decision-support systems into judicial processes compels a fundamental re-examination of judicial impartiality and constitutional legitimacy. While existing scholarship has primarily addressed either human cognitive bias or algorithmic discrimination as separate phenomena, this article argues that judicial bias in the digital age must be understood as a structurally layered constitutional problem emerging from the interaction between human cognition and algorithmic authority.

First, the study systematizes major types of cognitive bias affecting judicial reasoning, demonstrating that impartiality is not a natural attribute of adjudication but a fragile institutional ideal requiring continuous protection. Second, it analyzes algorithmic bias as a socio-technical phenomenon rooted in historical data inequalities, model architecture, and optimization objectives. Contrary to the widespread narrative of computational neutrality, algorithmic systems embed normative choices that may conflict with equality, due process, and non-discrimination principles.

The article advances a central theoretical claim: the rise of algorithmic decision-support systems does not merely introduce new technical risks but transforms the source of judicial authority itself. When statistical accuracy begins to compete with legal justification as a legitimacy foundation, the rational-legal structure of adjudication is normatively destabilized. Statistical prediction cannot substitute constitutional reasoning.

To address this transformation, the article proposes an original four-layered framework—the Erkan Constitutional Algorithmic Safeguard Model (ECASM). The model conceptualizes algorithmic systems not as neutral administrative tools but as constitutional risk-producing instruments subject to multi-layered oversight: normative compatibility review, transparency and explainability requirements, institutional accountability safeguards, and behavioral integrity controls. By integrating constitutional law, human rights doctrine, judicial theory, behavioral science, and algorithmic governance, the model provides a structured safeguard architecture against the delegation of judicial authority to statistical systems.

Makaleye Atıf Bilgisi

Erkan, F. (2026). **Reconfiguring Judicial Legitimacy: Articial Intelligence, Cognitive Bias, and The Rise of Algorithmic Authority** *International Journal of Social Science (IJSS Journal), (e-ISSN:2548-0685)* Vol:10, Issue:42; s. 159-220.

DOI: 10.52096/usbd.10.42.09

Gönderim: 12.01.2026

Kabul: 15.02.2026

SOSYAL BİLİMLER DERGİSİ (ULUSLARARASI HAKEMLİ DERGİ)

editorusbd@gmail.com

The article concludes that in the AI era, judicial impartiality cannot be reduced either to human virtue or to algorithmic accuracy. Rather, it must be reconstructed as a dynamic constitutional equilibrium between human normative responsibility and technologically mediated decision-making. Algorithms may assist adjudication, but they cannot embody justice.

Keywords: Algorithmic Bias, Judicial Impartiality, Algorithmic Authority, Artificial Intelligence in Courts, Human–Machine Interaction, Constitutional Legitimacy, Multilayered Algorithmic Oversight

Introduction

Judicial impartiality constitutes one of the foundational pillars of the rule of law and the right to a fair trial. Impartiality does not merely signify that a judge is free from personal opinions or predispositions; it also requires that the judicial decision-making process be objective, predictable, reasoned, and consistent with constitutional values. In the modern constitutional state, this ideal has traditionally been safeguarded through the ethical responsibility of individual judges, professional training, and institutional guarantees. However, the increasing integration of artificial intelligence–based algorithmic systems into judicial processes has reopened debate over both the scope and the normative foundations of impartiality.

Today, the risks to judicial impartiality no longer stem solely from the conscious or unconscious biases of the human judge. Historical inequalities embedded in datasets, design choices in algorithmic architecture, optimization objectives, and machine learning–based decision-support systems have become invisible yet influential components of the judicial process. These systems are often presented as “objective” and “neutral” tools. Yet algorithmic bias may generate serious risks with respect to equality, the prohibition of discrimination, and the right to a reasoned decision.

The issue, however, is not limited to the technical accuracy of algorithms or the quality of data. At a deeper level, the problem concerns the source of judicial authority itself. The rise of algorithmic systems carries the potential to shift the legitimacy of judicial decisions from legal normativity toward claims of statistical accuracy. This development calls for a reassessment of rational-legal authority and the normative foundations of judicial legitimacy. The distinction between statistical prediction and normative justification has thus become a central line of debate in determining the place of the judiciary in the age of artificial intelligence.

This article approaches the problem of judicial impartiality in the AI era from a multilayered perspective. First, drawing on cognitive psychology and behavioral law scholarship, it systematically analyzes the types of bias that emerge in human judicial decision-making. Second, it examines the impact of algorithmic bias on judicial processes, particularly in the context of risk assessment tools and algorithmic decision-support systems. The study argues that human cognitive bias and algorithmic bias are not independent phenomena; rather, they frequently interact in ways that reinforce and scale one another.

The original contribution of this article lies in conceptualizing algorithmic bias not merely as a technical malfunction but as a normative issue situated within the broader transformation of judicial legitimacy. In this regard, the study proposes a Multilayered Algorithmic Oversight Model that evaluates the integration of algorithmic systems into judicial processes through a constitutional filtering architecture. Comprising layers of normative compliance, explainability, institutional accountability, and behavioral impact assessment, the model seeks to preserve the balance between technological efficiency and constitutional legitimacy.

In conclusion, this study contends that judicial impartiality can no longer be reduced either to human virtue or to technical accuracy. It has become a dynamic normative ideal that requires the reconfiguration of human–machine interaction within constitutional boundaries. In the age of artificial intelligence, safeguarding the legitimacy of the judiciary does not depend on rejecting algorithmic tools, but on subjecting them to the normative framework of law.

I. ARTIFICIAL INTELLIGENCE, THE RULE OF LAW, AND JUDICIAL IMPARTIALITY

1. The Concept of Artificial Intelligence, Its Areas of Use, and Its Purpose

Artificial intelligence (AI) is an interdisciplinary field that refers to the simulation of cognitive processes characteristic of human intelligence—such as learning, reasoning, problem-solving, and decision-making—through algorithms and computational models. The term “artificial intelligence” was first introduced by John McCarthy at the Dartmouth Conference in 1956 (Russell and Norvig 2010). While early AI systems were primarily based on symbolic logic,

contemporary developments are dominated by machine learning and, in particular, deep learning approaches.[1]

Definitions of AI in international legal instruments provide an important normative framework for understanding the functions and societal impact of these systems (Erkan 2025). In this context, AI systems have been defined both in the Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law[2] and in the European Union's Artificial Intelligence Act (Regulation (EU) 2024/1689)[3].

According to Article 2 of the Council of Europe Convention, an AI system is:

"…a machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."

Similarly, Article 3(1) of Regulation (EU) 2024/1689 defines an AI system as:

"a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments."

Both definitions make clear that AI systems are not merely technical tools but structures capable of affecting human action and social environments with a degree of autonomy.

---

[1] AI also encompasses subfields such as machine learning (ML) and deep learning (DL). Machine learning enables systems to learn from experience and improve their performance without being explicitly programmed for each specific task. Deep learning, by contrast, employs larger datasets and complex neural network architectures to generate more advanced and layered learning processes.

[2] The Convention was adopted by the Committee of Ministers of the Council of Europe on 17 May 2024. It was opened for signature on 5 September 2024 in Vilnius, Lithuania, during the Council of Europe Conference of Ministers of Justice. Council of Europe. (2024). *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*. Strasbourg. Retrieved from https://www.coe.int/en/web/artificial-intelligence/convention.

[3] On 12 July 2024, Regulation (EU) 2024/1689 on Artificial Intelligence was published in the Official Journal of the European Union, becoming the first comprehensive horizontal legal framework for the regulation of AI systems across the EU. The Regulation will enter into force on 1 August 2024 in all 27 EU Member States, and most of its provisions will become applicable starting from 2 August 2026.
For the full text of the Regulation, see:
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689 *(Accessed 01/07/2025)*

Modern AI systems learn by identifying statistical relationships within large datasets rather than by following explicitly pre-programmed rules. This characteristic transforms AI from a purely automated instrument into a partially autonomous decision-support mechanism. Supervised and unsupervised learning models generate predictions about future outcomes based on past data, and such predictive outputs are increasingly used in public decision-making processes.

1.1. Areas of Use of Artificial Intelligence in the Legal Field

Artificial intelligence has transformed sectors such as healthcare, finance, automotive industries, and education. In the legal field, its application has become increasingly widespread across jurisdictions worldwide (Erkan 2025)[4].

AI applications in law predominantly take the form of decision-support systems. The principal areas of use identified in national and international scholarship can be summarized as follows:

1.1.1. Case Outcome Prediction Systems

AI applications aimed at predicting case outcomes have developed primarily in the United States and Europe. These systems analyze past judicial decisions, case types, party profiles, and judicial tendencies in order to generate forecasts regarding likely outcomes.

In the United States, models developed by Katz, Bommarito, and Blackman demonstrated that decisions of the U.S. Supreme Court could be predicted with accuracy rates approaching 70 percent (Katz, Bommarito, and Blackman 2017)[5]. Such findings strengthened the thesis that judicial behavior is not entirely unpredictable but exhibits identifiable patterns.

In Europe, research conducted in France and the Netherlands has attracted particular attention. Studies applying machine learning techniques to judgments of the European Court of Human Rights (ECtHR) have shown that violations under certain Convention articles can be predicted with relatively high accuracy (Aletras et al. 2016). While these findings highlight the data-

[4] For the concept, tools, and classification of artificial intelligence, see Erkan (2025).

[5] In the conclusion and future research section of the study, the following assessment is made: "Building on prior work in judicial prediction, we present the first generalized, consistent, and out-of-sample applicable machine learning model for forecasting decisions of the Supreme Court of the United States. Our model, generating predictions across nearly two centuries of cases, achieves 70.2% accuracy at the case outcome level and 71.9% accuracy at the justice voting level."

driven analyzability of judicial decisions, they also raise normative concerns regarding judicial independence and the individualized nature of adjudication.

Since the 2020s, AI-based case outcome prediction research has moved beyond classical statistical models and achieved high levels of performance using state-of-the-art architectures[6] and large-scale datasets.

What do these systems do in practice?

- They process case texts within a coherent semantic framework.
- They establish relationships between factual background, legal qualification, and outcome.
- They generate predictions regarding case outcomes and, in some instances, extract patterns in judicial reasoning.

For example, new architectures have predicted appellate decisions of Brazilian federal courts more successfully than human experts (Menezes-Neto and Clementino 2022). Advanced CNN- and GAN-based models have reportedly achieved accuracy rates exceeding 90 percent in datasets from the Supreme Court of Nigeria (Liu 2024)[7]. Research is also being conducted in different regions, including the Arab world, on the performance of large language model (LLM)-based systems in judicial prediction (Ammar et al. 2024)[8]. Field studies in China indicate that judges interact with LLMs in drafting judicial reasoning, and the results suggest that AI interaction may have both positive and potentially risky effects on decision quality (Zhang and Li 2025)[9].

[6] This statement refers to the highest-performing and most advanced classes of artificial intelligence models recognized in the literature. In the context of judicial decision prediction, this no longer denotes classical machine learning techniques (such as logistic regression or support vector machines), but rather advanced architectures such as Transformer- and Large Language Model (LLM)–based systems, as well as deep learning–based hybrid architectures.

[7] Since the 2020s, AI-based judicial decision prediction models have evolved into complex architectures that not only forecast outcomes but also increasingly focus on new dimensions such as model explainability and legal feature extraction.

[8] Recent studies examine the impact of Large Language Models (LLMs) across different legal systems and evaluate their decision-prediction performance within Arab legal systems.

[9] Artificial intelligence is used not only for outcome prediction, but also in assisting judges in drafting legal reasoning. While this development carries significant positive potential, it also entails risks of amplifying cognitive errors.

These developments demonstrate that judicial decision prediction research is ongoing and becoming increasingly sophisticated.

1.1.2. Risk Assessment and Recidivism Prediction Systems

One of the most controversial applications of artificial intelligence in criminal justice is the use of risk assessment and recidivism prediction systems. These systems calculate the likelihood that a defendant or convicted individual will reoffend by means of statistical models and are used as decision-support tools in processes concerning detention, bail, parole, and sentencing regimes. Particularly in Anglo-Saxon legal systems, such algorithmic instruments have become increasingly integrated into judicial decision-making.

One of the most well-known examples in this field is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system, widely used in the United States. COMPAS generates risk scores based on defendants' demographic characteristics, criminal records, and socioeconomic data. However, following an extensive investigation conducted by ProPublica, the system was heavily criticized for systematically assigning higher risk scores to Black defendants (Angwin et al., 2016). These findings triggered a global academic and legal debate concerning the neutrality and fairness of algorithmic decision-support systems.

Post-2017 scholarship has approached the COMPAS case not merely as a technical malfunction but as indicative of the structural limits of algorithmic justice. In particular, studies demonstrating that equal error rates and equal predictive accuracy across different demographic groups cannot be simultaneously achieved have rendered the claim of algorithmic "fairness" theoretically problematic (Chouldechova, 2017). This line of research shows that algorithmic risk scores may inevitably produce disproportionate impacts on certain groups.

Empirical studies published after 2020 have further questioned the assumption that algorithmic risk assessment systems are more accurate or fair than human decision-makers. An experimental study by Dressel and Farid demonstrated that both laypersons and algorithms exhibit similar error rates in predicting recidivism, thereby weakening claims of the objective superiority of algorithmic assessments[10].

[10] The assumption that algorithmic risk assessment systems are more accurate or fair than human decision-makers has been seriously challenged by the experimental study conducted by Dressel and Farid. Their findings

Comparable risk assessment tools are also employed in other jurisdictions, including **OASys (Offender Assessment System)**[11] in the United Kingdom, LSI-R (Level of Service Inventory–Revised) in Canada, and various risk classification mechanisms within the German penal enforcement system. In the United Kingdom, OASys is used primarily in probation and correctional contexts as a structured risk assessment instrument and does not formally constitute a binding decision-making mechanism[12]. Nevertheless, the literature on risk-based decision-making widely acknowledges that risk scores and classifications may effectively influence judicial discretion and shape evaluative processes in practice[13].

In Canada, strong critiques have emerged arguing that risk assessment instruments may produce discriminatory outcomes, particularly with respect to Indigenous peoples. These critiques suggest that such tools do not neutrally measure individual dangerousness but instead statistically reproduce existing socioeconomic and structural inequalities[14]. This indicates that

demonstrate that widely used recidivism prediction tools do not substantially outperform lay human predictions and raise important questions concerning claims of algorithmic superiority (Dressel & Farid, 2018).

[11] The Offender Assessment System (OASys) is a tool used to evaluate the risk of harm and reoffending and to inform decision-making processes in both sentencing and parole. It now incorporates artificial intelligence techniques to profile thousands of offenders on a weekly basis. Ministry of Justice (UK), *Offender Assessment System (OASys): User Guidance*, https://assets.publishing.service.gov.uk/media/5a7f676fed915d74e33f6380/research-analysis-offender-assessment-system.pdf (accessed February 10, 2026).

The OASys assessment considers both static (immutable) risk factors—such as criminal history and age—and dynamic (modifiable) risk factors, including accommodation, employability, relationships, lifestyle and environment, drug use, alcohol use, emotional well-being, thinking and behavior patterns, and attitudes. Different risk factors (criminogenic needs) are assigned different weights according to their predictive significance. The human assessor exercises discretion when assigning an overall score for dynamic risk factors. The system subsequently integrates data concerning the individual's criminal record and demographic characteristics, producing two distinct risk evaluations: the likelihood of violent reoffending and the likelihood of non-violent reoffending.

[12] In addition to, or as an alternative to OASys, another predictive instrument frequently used is the Offender Group Reconviction Scale (OGRS). OGRS estimates the probability of reoffending within two years by considering variables such as the individual's age at sentencing, gender, number of prior criminal sanctions (including cautions and convictions), age at first sanction, and current offense data, together with a "Copas rate" measuring the volume and frequency of the individual's alleged criminal career. OGRS provides separate predictors for violent and non-violent reoffending. "Offender Assessment System (OASys)" and "Offender Group Reconviction Scale (OGRS)," *Technology and Justice Project*, https://www.techandjustice.bsg.ox.ac.uk/ (accessed February 10, 2026).

[13] Official guidance on OASys is provided by the UK Ministry of Justice. Ministry of Justice (UK), *Offender Assessment System (OASys): User Guidance*.

[14] The broader theoretical debate on actuarial sentencing and risk-based penal decision-making has been critically examined in the scholarship of Kelly Hannah-Moffat and Paula Maurutto, who question the normative and procedural implications of embedding actuarial logics within sentencing practices (Moffat, 2013 and Moffat & Maurutto, 2010).

risk assessment systems may reflect historical and structural disadvantages within algorithmic decision-making processes rather than purely individualized assessments of risk.

Across continental Europe, a more cautious approach toward risk assessment systems has generally been adopted. The 2020 Ethical Charter of the European Commission for the Efficiency of Justice (CEPEJ) under the Council of Europe emphasizes that AI systems used in judicial contexts must be transparent, explainable, and subject to oversight, and explicitly states that such tools should not replace judicial discretion[15]. In European scholarship, algorithmic risk assessment systems are frequently examined within the framework of the right to a fair trial and the prohibition of discrimination, with the legal legitimacy of automated decision-support mechanisms being critically scrutinized.

In the context of the Indian legal system, although there are limited examples of predictive risk assessment tools being directly integrated into judicial decision-making, studies indicate that AI has found broader application as a supportive administrative instrument and legal information access system. For instance, AI-based platforms such as SUPACE provide judges with relevant legal arguments and prior case law, while language processing and document management functions are implemented through systems such as SUVAS. These applications focus primarily on enhancing research and data-processing capacity rather than on recidivism prediction or risk assessment. This suggests that the digital transformation of the Indian judiciary remains at an early stage (Sharma, 2023).

In conclusion, although risk assessment and recidivism prediction systems are presented as decision-support tools within criminal justice processes, contemporary academic literature strongly challenges claims regarding their neutrality and objectivity. Rather than eliminating human bias, algorithms risk reproducing past social and structural inequalities through

[15] The European Commission for the Efficiency of Justice (CEPEJ), operating under the auspices of the Council of Europe, adopted in December 2018 the first European instrument setting out ethical principles for the use of artificial intelligence (AI) in judicial systems. This Charter provides a framework of principles intended to guide policymakers, legislators, and justice professionals confronting the rapid development of AI within national judicial processes. See: Council of Europe, CEPEJ, *European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and their Environment*, https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment (accessed February 10, 2026).

mathematical models. This underscores the necessity of subjecting the use of artificial intelligence in criminal justice to strict legal and ethical safeguards.

1.1.3. Decision-Support Tools in Sentencing

In certain legal systems, AI-based sentencing support systems are being developed to enhance consistency in criminal justice across similar cases. These tools analyze past comparable cases and provide judges with advisory sentencing ranges, thereby aiming to establish a data-driven reference framework within the decision-making process.

Particularly in the context of federal criminal law in the United States, statistical analyses are incorporated within the framework of the Sentencing Guidelines system, and in some states algorithmic recommendation and risk-based assessment tools are applied in practice. However, the existence of such systems also raises concerns that judicial discretion may be effectively narrowed in the face of algorithmic recommendations.

Although there is no single nationwide "AI sentencing system" operating at the federal level in the United States, the influence of algorithmic recommendation and risk assessment tools on sentencing decisions has been extensively examined in recent years. Contemporary scholarship focuses especially on the following questions:

- Under what conditions can algorithmic recommendation systems be considered sufficiently successful to substitute or approximate human judicial reasoning?
- Is technical accuracy sufficient to ensure normative legitimacy?
- By which criteria should justice, explainability, and freedom from bias be measured?

Studies published in 2024 emphasize that the evaluation of algorithmic sentencing recommendation systems should not rely solely on performance metrics; principles such as the right to a fair trial, equality before the law, and transparency must also be taken into account. This approach clarifies the distinction between technical adequacy and legal legitimacy, demonstrating that algorithmic tools must not only be "accurate" but also "normatively defensible."[16]

[16] Ryberg's 2024/2025 study emphasizes that the difference between the performance of algorithmic predictions and that of human judicial decisions should not be assessed solely through technical metrics, but also in light of

The literature on system design further highlights several essential safeguards for algorithmic decision-support tools:

- Transparent model architecture
- The right to explanation
- Judicial reviewability
- The principle that the final decision must remain with a human judge

In this context, critiques developed around risk assessment tools used in the United States (such as COMPAS and similar systems) demonstrate that algorithmic recommendations, even when formally non-binding, may influence judicial decision-making processes. This dynamic reveals the tension between the principle of individualized sentencing and the tendency toward algorithmic standardization in criminal justice.

In Europe, AI applications in sentencing are addressed within a more cautious framework. In the Netherlands, experimental decision-support systems provide judges only with statistical trends derived from similar cases and do not produce binding recommendations. Nevertheless, due to the potential of algorithmic outputs to function as cognitive anchors within judicial reasoning, such systems have become subjects of academic and ethical debate (Barkat & Busuioc, 2022).

Within the French legal system, the use of algorithmic decision-support tools in criminal justice has been confined within explicit legal boundaries. In particular, the profiling of judges' decision patterns or individual tendencies through automated data processing techniques has been prohibited. This regulation aims to prevent judicial discretion from being subjected to algorithmic standardization and to avoid the emergence of indirect normative pressure on judges.

In the United Kingdom, AI tools are employed at various stages of criminal proceedings. Police authorities use predictive analytics and facial recognition systems, while courts utilize AI systems for case management, legal research, and draft preparation. Recent disclosures by some judges that AI has been used in the preparation of draft judgments indicate that algorithmic

ethical and legal criteria. This approach transcends the classical claim of "objectivity" and situates performance evaluation within a broader normative framework (Ryberg, 2025).

tools are beginning to penetrate even the judicial writing process. Nevertheless, these systems do not assume responsibility for the final decision and function only as supportive instruments[17].

In India, AI applications in sentencing do not take the form of predictive recommendation systems; rather, they are primarily used for case management, document analysis, and legal research support. This suggests that algorithmic integration into sentencing remains at an early stage in India (Sharma, 2023).

In China, the digitalization of the judiciary has been adopted at the level of state policy and expanded under the "Smart Court" reform initiative, which promotes AI-supported judicial information systems. Although these systems do not replace judges, they play an active role in case management, document analysis, and legal data processing. Academic literature indicates that this reform has transformed not only administrative functioning but also judicial supervision and institutional oversight mechanisms. Examples such as the "206 System" implemented in Shanghai provide judges with analytical recommendations while simultaneously raising concerns regarding transparency, accuracy, and algorithmic bias[18].

A comparative assessment demonstrates that in no legal system has sentencing been fully automated. The prevailing trend favors a model of "human-centered algorithmic support" rather than one of "automated decision-making." Fundamental guarantees such as the principle of individualized sentencing, judicial autonomy, and the right to a fair trial normatively constrain the possibility of fully automated sentencing.

Accordingly, the core issue is not whether artificial intelligence should be used in sentencing, but under what limits and safeguards such use can be considered legitimate.

1.1.4. Case Law Analysis and Text Mining

The most widespread and comparatively least controversial application of artificial intelligence in the judicial field is case law analysis and text mining. Through natural language processing (NLP), machine learning, and large-scale data analytics, millions of judicial decisions can be classified; patterns of similar disputes can be identified; normative argumentation structures

[17] https://www.techandjustice.bsg.ox.ac.uk (Accessed 11/2/2026).
[18] https://english.court.gov.cn/2022-12/12/c_838810.htm?utm_source=chatgpt.com.

can be analyzed; and even predictive models regarding the likely direction of decisions can be developed.

These tools do not directly render judgments; however, they transform the epistemic infrastructure of legal research and judicial reasoning.

1.1.4.1. United States: Commercial Legal Analytics Platforms

In the United States, the transformation of case law analysis has largely been driven by the private sector. Platforms such as Westlaw Edge and Lexis+ have moved beyond traditional keyword searches by developing advanced legal analytics tools. These systems are capable of visualizing:

- Judge-specific decision tendencies
- Success rates of particular legal arguments
- Citation networks and precedential relationships

The work of Kevin D. Ashley demonstrates that AI-supported legal reasoning has made significant progress, particularly in the fields of case-based reasoning and computational argumentation modeling.

Although these tools do not directly issue decisions, they substantially accelerate strategic legal research processes. Nevertheless, two principal critiques emerge in the literature:

1. The statistical reinforcement of past decision patterns may narrow the dynamic and evolving character of law.
2. Algorithmic prioritization may direct judicial attention toward particular lines of precedent, thereby generating an invisible anchoring effect.

Thus, the normative impact is indirect, yet structural.

1.1.4.2. Europe: A Publicly Oriented Digital Case Law Ecosystem

In Europe, case law analysis is generally approached from a public policy perspective. In countries such as Estonia, Finland, and the Netherlands, the systematic digitalization of judicial decisions has been completed, and AI-supported search and classification tools have been developed through publicly accessible databases.

In Estonia, pilot projects concerning small-value disputes have demonstrated that draft decisions may be prepared algorithmically. However, these systems do not generate fully automated criminal judgments; rather, they function as template generators and consistency-check mechanisms.

The ethical principles articulated within the framework of the Council of Europe's CEPEJ emphasize that AI use must remain human-centered and subject to oversight.

A key principle underlying these approaches can be summarized as follows: Artificial intelligence does not generate norms; it organizes access to norms.

1.1.4.3. Theoretical Integration: Judicial Cognition and Algorithmic Anchoring

At first glance, case law analysis tools appear to be merely technical research facilitators. However, when evaluated in light of judicial cognition literature, they reveal deeper implications.

Research in cognitive psychology demonstrates that decision-makers, particularly under conditions of uncertainty, are susceptible to cognitive tendencies such as:

- The anchoring effect
- Automation bias

Algorithmic case-mapping systems, by rendering certain decision patterns more visible, may direct judicial attention toward statistically dominant trends. In this way, even a system that does not directly produce norms may indirectly generate normative weight.

At this point, artificial intelligence intervenes not in the content of the decision itself, but in the epistemic framing of the decision.

Accordingly, case law analysis tools should be assessed not only from the perspective of technical efficiency, but also in terms of judicial independence, freedom of reasoning, and cognitive autonomy.

1.1.5. Judicial Administration and Workload Management

Artificial intelligence technologies are used in judicial systems not only as decision-support tools but also in the organization of judicial administration and workload management. The use of algorithmic systems in administrative processes such as file classification, case prioritization, hearing schedule planning, and the allocation of judges and court staff is becoming increasingly widespread.

Although these applications do not directly intervene in the adjudicative process, they generate indirect yet structural effects by reshaping the institutional framework within which judicial decisions are produced. The issue, therefore, concerns not merely technical efficiency but the manner in which the institutional infrastructure of judicial impartiality is configured.

1.1.5.1. Singapore: Judicial Efficiency and the Digital Court Model

Singapore has been an early adopter of data-driven management systems in judicial administration under its "e-litigation" and "Smart Courts" strategy. Digital case management platforms developed within the Supreme Court of Singapore perform functions such as automated file classification, detection of procedural deficiencies, and prediction of case duration.

Algorithms designed to forecast case timelines and optimize hearing schedules are particularly noteworthy. These systems analyze court capacity, number of parties, case type, and historical average durations to generate scheduling recommendations. In the literature, the Singaporean model is frequently cited as an example of "data-driven rationalization in judicial governance."

Nevertheless, some academic assessments argue that algorithmic planning systems may create implicit normative preferences in case prioritization. For instance, the systematic acceleration or postponement of certain categories of cases—while seemingly a technical scheduling decision—may in practice establish a normative hierarchy. This raises concerns regarding the indirect justice implications of administrative design.

1.1.5.2. United Kingdom: The HMCTS Reform Programme

In the United Kingdom, the digital transformation programme conducted by HM Courts & Tribunals Service (HMCTS) incorporates extensive use of data analytics in case management

and resource allocation processes. Online filing systems, automated case-routing mechanisms, and workload-based planning tools have been developed[19].

These systems are particularly influential in the following areas:

- Automated case routing according to case type
- Recommendations for judicial assignment based on workload distribution
- Performance monitoring based on average decision timelines

The primary objective is to enhance the courts' caseflow management capacity. However, a significant line of critique within British scholarship contends that performance measurement tools may generate indirect pressure on judges, and that a "quantitative efficiency" approach risks narrowing the space of judicial discretion.

Within this framework, algorithmic management tools, even without intervening in the substance of decisions, may produce indirect normative consequences by shaping the temporal and institutional conditions under which decisions are rendered.

1.1.5.3. Australia and Canada: Case Allocation and Workload Optimization

In Australia, digital workflow systems are used within the Federal Court and several state courts for case allocation and management. These systems generate automated recommendations by considering judges' areas of expertise, current workloads, and case categories.

In Canada, particularly in the provinces of British Columbia and Ontario, electronic case management systems have been implemented, allowing dynamic file allocation based on caseload intensity. A notable feature of the Canadian model is that algorithmic systems generate "administrative recommendations," while the final assignment decision remains with the chief justice or court administration.

In both countries, the primary objective is to enhance judicial efficiency. However, the literature discusses several associated risks:

- Specialization-based file allocation may gradually produce de facto judicial profiling.

[19] Gary O'Reilly, "HMCTS is accelerating the responsible adoption of artificial intelligence (AI) to transform the courts and tribunals," Inside HMCTS (3 September 2025).

- Certain case types may become closely associated with particular judges.
- Performance metrics may influence judicial career trajectories.

Although these effects do not directly intervene in the content of judicial decisions, they are significant with regard to the dimension of "appearance of impartiality" within judicial independence.

Interim Conclusion

Comparative examples demonstrate that artificial intelligence transforms the institutional infrastructure of judicial systems before it transforms the substantive content of decisions.

Planning and efficiency-oriented systems in Singapore and the United Kingdom, as well as workload optimization tools in Australia and Canada, have become widespread. These systems do not directly render judgments; however, they reorganize the structure within which judicial decisions are produced.

Accordingly, the use of artificial intelligence in judicial administration cannot be assessed merely as a form of technical modernization. Through file allocation, time management, and performance analytics, a structural transformation is taking place—one that may affect the institutional dimension of judicial impartiality. For this reason, the design process must be evaluated not solely from the perspective of efficiency, but also within the framework of institutional impartiality and the right to a fair trial.

1.2. What Do Artificial Intelligence Systems Do in Law?

Contemporary AI applications in the legal field have moved beyond traditional keyword-based search mechanisms. In particular, models based on deep learning and transformer architectures process legal texts not merely at the lexical level, but within the framework of context, normative structure, and justificatory relationships.

These systems aim not at modeling the formal surface of legal reasoning, but rather its semantic and structural dimensions.

1.2.1. Semantic Analysis of Case Texts as Contextual Wholes

Modern AI systems are capable of analyzing petitions, defenses, and judicial decisions by considering the relationships between factual narratives, party claims, and applicable legal norms. Instead of fragmenting the text into isolated units, contextual language models treat legal documents as integrated wholes.

The “attention” mechanism of transformer-based architectures identifies legally determinative expressions within a text and assigns greater weight to specific sections. Through contextual prioritization, the model can distinguish facts and arguments that carry particular legal significance.

Within this framework, artificial intelligence processes the text not merely as a formal dataset but as a legal narrative.

1.2.2. Establishing the Relationship Between Facts, Legal Qualification, and Outcome

One of the most significant features of contemporary AI models is their capacity to establish a relationship between factual narratives and legal qualification. Systems learn, through statistical pattern recognition, which types of factual scenarios are associated with which legal norms and which outcomes these associations tend to produce.

For example, studies conducted on the case law of the European Court of Human Rights have demonstrated that models can match specific factual elements with relevant provisions of the European Convention on Human Rights and, on that basis, predict findings of violation or non-violation.

This development indicates that artificial intelligence attempts to model not only the final outcome but also the normative chain leading to that outcome. However, unlike human reasoning, this modeling remains probabilistic and pattern-based.

1.2.3. Outcome Prediction and Extraction of Justification Patterns

Post-2020 scholarship demonstrates that AI systems are capable not only of predicting judicial outcomes but also of identifying recurring patterns in judicial reasoning. Large language models can analyze justificatory structures used in similar cases and detect correlations between particular argumentative strategies and specific outcomes.

Field studies conducted in jurisdictions such as China and Brazil report that such systems have identified recurring argumentative structures in appellate decisions and have provided judges with "reasoning suggestions."

While this development holds positive potential in terms of consistency, it simultaneously raises concerns regarding the automation of justification and the possible erosion of individualized reasoning.

1.3. The Purpose of Artificial Intelligence and the Normative Tension

The common denominator of these applications is the claim of enhancing efficiency, speed, and consistency within judicial systems. European policy documents frequently emphasize the potential of artificial intelligence to reduce workload and prevent procedural delays.

However, a fundamental normative question arises:

If algorithms learn from past decision patterns, is it not inevitable that they will reproduce past cognitive and structural biases?

This question forms a natural transition to the subsequent discussion on algorithmic bias.

The objective of AI applications in the legal domain is not merely technical efficiency. Judicial decision-making is a normative, value-laden activity deeply intertwined with human rights. Consequently, the advantages of speed and standardization do not necessarily produce outcomes compatible with justice, fairness, and impartiality.

The literature highlights several major lines of critique:

- Data-driven models may reproduce existing biases embedded in historical data (Barocas & Selbst, 2016)[20].

[20] The study by Barocas and Selbst argues that big data techniques such as data mining and machine learning, even when deployed with the aim of eliminating human bias, carry the risk of reproducing pre-existing social inequalities and patterns of discrimination. The underlying reason for this risk lies in the fact that the datasets on which algorithms are trained may embed historical biases, beginning with choices regarding definitions and labeling practices. The authors provide a comprehensive analysis demonstrating how outputs generated by previously biased systems may be learned by predictive models and subsequently reproduced in future decision-making processes.

- The "black box" character of algorithmic systems may conflict with the right to a reasoned judgment and principles of judicial transparency (Pasquale, 2015)[21].
- Excessive reliance on algorithmic recommendations may weaken the judge's obligation of independent assessment.

In this respect, artificial intelligence constitutes a dual-character instrument in relation to judicial impartiality:

On the one hand, it possesses the potential to reduce arbitrariness and inconsistency arising from human decision-making.
On the other hand, through algorithmic bias stemming from design choices, data selection, and objective functions, it may generate new forms of risk.

Accordingly, the use of artificial intelligence in the legal domain should not be treated as a merely technical advancement, but as a structural transformation that must be subjected to normative evaluation in light of the rule of law and the principle of judicial impartiality.

II. THE RULE OF LAW AND JUDICIAL INDEPENDENCE

1. The Concept of the Rule of Law

The principle of the rule of law, in its classical sense, refers to the supremacy of law and the subjection of state power to legal constraints. However, this principle is not limited merely to the existence of laws; it also encompasses normative elements such as the equal application of the law, the prevention of arbitrariness, the existence of independent oversight mechanisms, and the provision of judicial safeguards.

According to the definition[22] provided by the United Nations Secretary-General, the rule of law refers to a system of governance in which all persons, institutions, and entities—public and private alike—are accountable to laws that are publicly promulgated, equally enforced, and

[21] As Frank Pasquale emphasizes in *The Black Box Society*, the "black box" character of modern artificial intelligence systems—namely, the opacity of their internal functioning—may directly conflict with the right to a reasoned judgment and the principle of judicial transparency. This opacity restricts the ability to meaningfully trace and scrutinize the reasoning process underlying a decision, thereby generating serious concerns regarding legal legitimacy.

[22] United Nations Secretary-General, *Report of the Secretary-General: The Rule of Law and Transitional Justice in Conflict and Post-Conflict Societies*, UN Doc. S/2004/616 (23 August 2004), quoted in United Nations, "What is the Rule of Law?" webpage, https://www.un.org/ruleoflaw/what-is-the-rule-of-law/ (accessed February 2026).

independently adjudicated. This definition emphasizes both the procedural dimension of the rule of law (adherence to legal norms and procedural guarantees) and its normative dimension (equality and independent adjudication).

Legal scholarship offers diverse interpretations regarding the substantive content of the rule of law. For instance, A.V. Dicey's conception of the rule of law stresses the importance of legally established and predictable norms, the subjection of administrative authorities to law, and the preservation of ordinary courts. Contemporary approaches, by contrast, treat the rule of law not merely as the dominance of legislation but as a framework that includes equal application of laws, protection of fundamental rights, and the guarantee of the right to a fair trial.

Within this framework, judicial independence is regarded as one of the core components of the rule of law. The realization of the rule of law requires the limitation of arbitrary actions by the legislative and executive branches while making the independent and impartial functioning of the judiciary a prerequisite. Accordingly, an independent judiciary is not a consequence of the rule of law; it is a precondition for its existence.

2. Judicial Independence: Conceptual Content and Definitions in the Literature

Judicial independence refers to the capacity of the judiciary to render legal decisions free from the influence of the legislative and executive branches. This concept is not confined to the autonomy of individual judges; it also encompasses institutional and structural safeguards.

In international legal scholarship, judicial independence is not treated as a one-dimensional concept but as a multilayered institutional phenomenon. Modern approaches emphasize that independence depends both on normative guarantees (de jure independence) and on practical implementation (de facto independence)[23].

---

[23] Leading theorists emphasize that judicial independence encompasses not only the personal autonomy of the judge, but also internal (organizational) independence, institutional or administrative independence, and decisional independence. Huchhanavar (2023) classifies these components as individual, internal, and institutional independence, demonstrating that the administrative and institutional structure of the judiciary constitutes an inseparable element of independence itself.

Similarly, the analysis presented in *The Cambridge Handbook of Constitutional Theory* (Kosař & Spáč) conceptualizes judicial independence across three dimensions—*de jure* institutional independence, *de facto* institutional independence, and decisional independence—thereby underscoring both the importance of formal legal guarantees and the autonomy exercised in practice.

In this context, the literature commonly distinguishes three principal dimensions:

- **Personal independence:**
  Security of tenure, protection against arbitrary removal or disciplinary sanctions, and adequate financial and social guarantees for judges.
- **Decisional independence:**
  The absence of external pressure, direction, or hierarchical instruction in the adjudication of individual cases.
- **Institutional independence:**
  The autonomous position of the judiciary vis-à-vis the legislative and executive branches with respect to budgetary authority, appointments, disciplinary mechanisms, and administrative structure.

This tripartite structure demonstrates that judicial independence is not merely an individual privilege but a systematic and structural requirement for the proper application of law.

Accordingly, judicial independence refers not only to the psychological freedom of the individual judge but also to the insulation of the entire institutional architecture of decision-making from external interference.

3. Judicial Independence in the Case Law of the European Court of Human Rights

The European Court of Human Rights (ECtHR) considers judicial independence to be one of the constitutive elements of the right to a fair trial under Article 6 §1 of the European Convention on Human Rights (ECHR).

Article 6 §1 of the ECHR guarantees that everyone is entitled to have their case heard by "an independent and impartial tribunal established by law." This formulation demonstrates that judicial independence is not merely an institutional organizational principle; it is also a standard of individual rights protection (Erkan, 2023).

Taken together, these frameworks demonstrate that judicial independence is not confined to the autonomy of the individual judge; it also encompasses the administrative organization of the judiciary as an institution and its structural separation from the executive and legislative branches.

3.1. The Principle of a "Tribunal Established by Law"

According to the ECtHR, the requirement of a "tribunal established by law" ensures that the judicial system is not left to the discretion of the executive and is regulated by the legislature on the basis of democratic legitimacy. The Court does not interpret this concept narrowly. It requires not only the existence of a normative legal basis but also that the court be constituted in accordance with the rules governing its establishment, jurisdiction, and functioning[24].

Accordingly, failure to comply with the rules governing the appointment of judges may result in a finding that a tribunal was not "established by law," thereby constituting a violation of Article 6[25].

This approach shows that independence is not merely a formal matter but a question of structural legitimacy.

3.2. Criteria of Independence

In its case law, the ECtHR assesses independence primarily in relation to independence from the executive and from the parties. The Court has identified the following elements as relevant in determining whether a body can be regarded as "independent"[26]:

- The manner of appointment of its members
- The duration of their term of office and security of tenure
- The existence of safeguards against external pressure
- The appearance of independence

The element of "appearance of independence" is particularly significant. According to the Court, not only actual independence but also the absence of legitimate doubt in the eyes of a reasonable observer is required.

The existence of an organic or functional link with the executive may be incompatible with independence. If the final decision-making authority lies with the executive, or if a judicial body is bound by the executive's binding opinion, this may conflict with Article 6. For example,

[24] Zand v. Austria, no. 7360/76, Commission decision, 15 DR 70, 80, 1978
[25] (Posokhov v. Russia, no. 63486/00, 2003-V)
[26] These criteria were explicitly formulated by the European Court of Human Rights in *Campbell and Fell v. the United Kingdom* (Applications nos. 7819/77 and 7878/77, judgment of 28 June 1984, § 78).

in *Benthem v. the Netherlands* (no. 8848/80, 23 October 1985), the Court held that vesting final decision-making authority in a minister was incompatible with Article 6.

Similarly, in *Sramek v. Austria* (no. 8790/79, 22 October 1984), the Court accepted that a body subject to the executive's binding opinion could not be regarded as independent.

3.3. Security of Tenure and De Facto Protection

The Court considers security of tenure to be one of the fundamental components of judicial independence. Protection of judges against arbitrary removal during their term of office constitutes a decisive criterion in assessing independence[27].

This approach demonstrates that independence must be safeguarded not only at the normative (de jure) level but also at the practical (de facto) level. Even where the legal framework appears sufficient, independence cannot be said to exist if, in practice, the executive exerts influence.

3.4. Impartiality: The Subjective and Objective Tests

The ECtHR examines impartiality through both subjective and objective criteria:

1. **Subjective test:**
   This test investigates whether the judge harbored any personal prejudice or bias.
2. **Objective test:**
   This test evaluates whether sufficient guarantees exist to exclude any legitimate doubt from an external perspective[28].

Objective impartiality becomes particularly significant where the judge has participated in earlier stages of the proceedings, has an organic link with one of the parties, or occupies a structural position capable of generating doubt[29].

This approach demonstrates that impartiality is not merely an internal psychological state; it is also a matter of institutional design and appearance.

---

[27] Engel and Others v. the Netherlands 5100/71., 8 June 1976).
[28] Hauschildt v. Denmark kararında Mahkeme, tarafsızlığın bu iki tesst aracılığıyla belirlenmesi gerektiğini ifade etmiştir (no. 10486/83, 24 May 1989, §46).
[29] In *Piersack v. Belgium*, the Court stated that the participation of a judge in the trial chamber who had previously served in the prosecutorial authority could give rise to legitimate doubts as to impartiality (Application no. 8692/79, judgment of 1 October 1982, § 30).

Theoretical Significance

This jurisprudential framework reveals that judicial independence is not confined to the individual will of the judge. Rather, it encompasses a broad structural domain extending from appointment procedures and institutional organization to security of tenure and protection against external pressures.

At this point, the critical question emerges:

If algorithmic systems are integrated into the judicial decision-making process, how should their design and use be assessed in light of the independence and impartiality criteria established by the ECtHR?

This question directly prepares the ground for the next section, which will examine the relationship between algorithmic bias and structural independence.

3.5. Reinterpreting the ECtHR Standard in the Context of AI-Supported Decision Systems

The criteria of judicial independence and impartiality developed in the case law of the European Court of Human Rights (ECtHR) have historically been shaped around the institutional position and individual conduct of the human judge. However, the increasing use of AI-based decision support systems in judicial processes necessitates that these criteria be assessed not only at the personal and institutional levels, but also at the level of technological infrastructure.

The use of algorithmic tools in areas such as evidence analysis, risk assessment, case classification, precedent matching, or draft decision generation raises new constitutional questions concerning the scope of the principle of an "independent and impartial tribunal" guaranteed under Article 6 §1 of the European Convention on Human Rights (ECHR). In this context, what is required is not a narrowing of the ECtHR standards, but rather their structural expansion.

3.5.1. The Principle of a "Tribunal Established by Law" and Algorithmic Infrastructure

According to the ECtHR, the principle of a "tribunal established by law" encompasses not only the normative basis of a court's existence, but also the lawful regulation of its functioning (see

Zand v. Austria). The Court interprets this principle in conjunction with the concepts of democratic legitimacy and legal certainty.

This approach requires that algorithmic systems playing a determinative role in judicial decision-making:

- have a clear legal basis;
- be subject to accessible and foreseeable rules;
- be normatively regulated with respect to their design, updating, and oversight processes.

If an algorithmic system carrying decisive weight in a judicial decision does not rest on a clear and democratically enacted legal framework, this may result in a violation of the "lawfulness" requirement. In particular, leaving the technical control of the algorithm to the discretion of the executive would be incompatible with the ECtHR's understanding of separation of powers and democratic legitimacy.

Accordingly, the principle of a "tribunal established by law" calls for an expanded interpretation under which the algorithmic infrastructure itself is subject to normative oversight. The issue is no longer merely whether the court has been lawfully constituted, but whether the entirety of the decision-generating mechanism is bound by law.

3.5.2. Institutional Independence and the Risk of "Technological Dependency"

In its assessment of independence, the ECtHR places particular emphasis on safeguards against external interference and on institutional separation from the executive (see *Campbell and Fell v. the United Kingdom*; *Sramek v. Austria*).

Where AI systems:

- are developed by private companies,
- are procured by the executive, or
- rely on datasets shaped by public policy priorities,

a new dimension is added to the analysis of institutional independence.

In this context, several questions become crucial:

- Who determines the dataset and parameters of the algorithm?
- Does the court possess real and effective authority to review algorithmic outputs?
- Is the system susceptible to indirect influence by the executive?

If algorithmic output becomes de facto binding, or if the court—due to a lack of technical capacity—is unable to meaningfully scrutinize the algorithmic assessment, this may amount to an indirect transfer of the judicial function to the executive. This possibility transposes the problem of executive dependency identified in *Benthem v. the Netherlands* to the technological sphere.

Therefore, the analysis of institutional independence must inevitably incorporate the dimension of "technological dependency." Judicial independence must be protected not only against political interference, but also against indirect influence exercised through technical infrastructure.

3.5.3. Impartiality and Algorithmic Bias

The subjective and objective tests of impartiality developed by the ECtHR (see *Piersack v. Belgium*; *Hauschildt v. Denmark*) can be functionally applied to algorithmic systems.

In the context of subjective impartiality, instead of the personal prejudice of a human judge, the focus shifts to systematic bias embedded in data and model design (algorithmic bias). Datasets that reflect historical inequalities, generate statistically adverse trends for certain groups, or produce discriminatory outcomes due to model architecture become the new focal points of impartiality analysis.

With respect to objective impartiality, the "reasonable observer" standard remains decisive. If an algorithm:

- relies on non-transparent datasets,
- is capable of generating discriminatory outcomes, or
- lacks effective auditability,

a reasonable observer may entertain legitimate doubts as to the impartiality of the tribunal.

Under the ECtHR's established emphasis on the "appearance of impartiality," not only actual impartiality but also the perception of impartiality must be safeguarded. Algorithms operating as opaque "black boxes" pose a serious risk to this appearance.

For this reason, algorithmic transparency and explainability are not merely technical preferences; they constitute structural components of the guarantee of objective impartiality under Article 6.

3.5.4. The Right to a Reasoned Judgment and Algorithmic Explainability

The ECtHR considers the obligation of courts to provide adequate and relevant reasoning as an element of the right to a fair trial (e.g., *Ruiz Torija v. Spain*).

If algorithmic analysis plays a determinative role in the decision-making process, the court must clarify in its reasoning:

- how the algorithmic output was assessed;
- which criteria were taken into account;
- whether the algorithm was binding or merely supportive;
- under what circumstances the judge may depart from the algorithmic recommendation.

Otherwise, the parties will be unable to understand the logical basis of the decision and to exercise effectively their rights of defense and appeal. This situation may undermine not only the requirement of reasoning, but also the principles of equality of arms and adversarial proceedings.

Algorithmic explainability, therefore, emerges as a technological complement to the right to a reasoned judgment.

3.5.5. The Principle of Structural Protection and the Baka Approach

In *Baka v. Hungary*, the ECtHR made clear that judicial independence must be protected not only through formal guarantees but also at the practical and structural levels. This approach is equally applicable to algorithmic systems.

If algorithmic tools:

- rely on datasets reflecting executive policy priorities;
- generate risk scores aligned with punitive policy orientations;
- systematically steer judges toward particular outcomes,

this may amount to a de facto erosion of independence.

Judicial independence cannot be indirectly undermined through technological mediation. In a democratic society, independence must be safeguarded not only against political interference but also against structural influences embedded within decision-generating mechanisms.

General Assessment

In conclusion, the criteria developed in ECtHR case law must be reinterpreted in the context of AI-supported decision systems. This reinterpretation entails the following expansions:

- The principle of a "tribunal established by law" → normative regulation of algorithmic infrastructure in a clear and foreseeable manner;
- Institutional independence → decision architecture free from technological dependency;
- Objective impartiality → effective control of algorithmic bias;
- The right to a reasoned judgment → ensured algorithmic explainability.

From this perspective, the use of AI systems in judicial processes does not narrow the ECtHR standard. On the contrary, it necessitates a stricter regime of independence, transparency, and accountability.

In a democratic state governed by the rule of law, judicial independence must now be protected not only against political power, but also against the structural effects that technological mediation and algorithmic systems may generate.

III. BIAS IN THE JUDICIARY AND TYPES OF BIAS

The judiciary is obliged to render decisions not only on the basis of legality, but also in accordance with the principles of impartiality and objectivity. Impartiality is recognized as one of the constitutive elements of the right to a fair trial in both constitutional legal orders and international human rights law. Nevertheless, judicial decision-making is not a purely

mechanical application of legal norms; it is grounded in the processes of evaluation, comparison, and reasoning inherent in the human mind.

Accordingly, no matter how experienced, expert, or ethically conscientious a judge may be, decision-making processes are not entirely independent from the cognitive functioning of the human mind. Modern decision theory and the behavioral law literature demonstrate that decision-makers—even in highly specialized fields—rely on cognitive shortcuts (heuristics) and are subject to systematic deviations (cognitive biases) (Tversky & Kahneman, 1974; Kahneman, 2011). These deviations are not necessarily conscious acts of partiality; yet they may systematically influence outcomes.

In contemporary scholarship, therefore, the concept of "bias" does not refer solely to ethical misconduct or intentional favoritism. It also encompasses cognitive, structural, and contextual tendencies that shape the decision-making process. Empirical research has shown that judges and jurors are not immune to these cognitive mechanisms (Guthrie, Rachlinski & Wistrich, 2001; Rachlinski et al., 2009).

For this reason, bias must be examined not only within the framework of normative legal rules, but also in light of cognitive psychology, behavioral economics, and decision theory.

1. Bias in the Judiciary: Conceptual Framework

1.1. Bias in the Legal Sense

In traditional legal doctrine, bias refers to personal interests, relationships, prejudices, or connections that undermine a judge's impartiality. Particularly in Anglo-Saxon legal systems, this approach has been structured around two principal tests:

- **Actual bias**
- **Apprehended bias / reasonable apprehension of bias**

The first test examines whether the judge possesses a concrete and demonstrable prejudice. The second test shifts the assessment of impartiality away from the judge's subjective mental state and anchors it in the perspective of an objective observer.

The High Court of Australia, in *Ebner v Official Trustee in Bankruptcy*, formulated the bias test as follows: Would a fair-minded lay observer reasonably apprehend that the judge might not

bring an impartial mind to the resolution of the case? The Court further emphasized that it must be demonstrated how the alleged interest or connection could logically influence the decision.

Similarly, the Supreme Court of Ireland, in *Bula Ltd v Tara Mines Ltd (No 6)*, adopted the criterion of whether a reasonable and informed person would have a legitimate doubt regarding impartiality.

This approach parallels the "appearance of impartiality" standard developed in the case law of the European Court of Human Rights (see *Hauschildt v. Denmark*). According to this standard, not only actual impartiality but also the appearance of impartiality must be preserved.

Thus, in classical doctrine, bias is grounded either in the existence of actual prejudice or in the objective impairment of the appearance of impartiality.

1.2. The Behavioral and Cognitive Dimension

Modern decision theory extends beyond the classical legal approach. The "heuristics and biases" theory developed by Amos Tversky and Daniel Kahneman demonstrated that individuals systematically deviate from the rational choice model under conditions of uncertainty (Tversky & Kahneman, 1974). The human mind employs shortcuts to simplify complex problems; however, these shortcuts frequently produce systematic errors.

Empirical research in the judicial context indicates that judges are likewise susceptible to these cognitive mechanisms (Guthrie, Rachlinski & Wistrich, 2001). The following types of deviations are particularly notable:

- **Anchoring:** The disproportionate influence of initially presented information on the final decision.
- **Confirmation bias:** The tendency to assign greater weight to evidence that confirms a pre-existing belief.
- **Hindsight bias:** The assumption, after learning an outcome, that it was predictable ex ante.
- **Framing effect:** Variations in decision outcomes resulting from different presentations of the same information.

For example, experimental studies have demonstrated that the sentence requested by a prosecutor in a criminal case can systematically influence the judge's final sentencing decision. This finding reveals that bias does not always arise from conscious favoritism; it may stem from the cognitive architecture of the decision-making process itself.

1.3. The Tension Between Normative Law and Behavioral Reality

The classical legal approach tends to define bias primarily in terms of conscious, interest-based, or ethically wrongful deviations. By contrast, the behavioral literature demonstrates that unconscious and systematic tendencies can also influence decision outcomes. This divergence creates a structural tension between the normative standard of impartiality and psychological reality.

This tension points to two important implications:

1. Impartiality is not merely an ethical obligation; it is a structural issue that requires awareness of cognitive limitations.
2. Institutional and procedural safeguards within the judicial system must be designed not only to prevent external interference, but also to mitigate systematic tendencies inherent in human cognition.

Accordingly, bias should be understood not solely as a category of legal violation, but also as a behavioral phenomenon grounded in decision science.

2. Types of Bias

Bias may arise not only from human sources, but also through algorithmic systems. Contemporary analysis must therefore examine bias along two principal axes:

- **Human-based bias**
- **Algorithmic (technological) bias**

This section first offers a systematic classification of human-based bias.

2.1. Human-Based Bias: A Systematic Classification

Judicial bias cannot be reduced to individual moral deviation. Modern scholarship demonstrates that bias is a multilayered phenomenon. Human-based bias may be examined under four main categories:

1. Cognitive bias
2. Social/cultural bias
3. Institutional bias
4. Strategic/political bias

This classification enables a comprehensive analysis of judicial decision-making processes across psychological, sociological, and structural dimensions.

2.1.1. Cognitive Bias

Cognitive bias refers to systematic deviations in decision-making arising from the information-processing mechanisms of the human mind. Its theoretical foundation lies in the "heuristics and biases" framework developed by Tversky and Kahneman. This approach demonstrates that human judgment is not always rational and may generate systematic deviations under conditions of uncertainty.

Empirical research in the judicial context shows that judges are not immune to such cognitive mechanisms (Guthrie, Rachlinski & Wistrich, 2001). In particular, the following biases have been shown to affect judicial assessment:

- **Anchoring**
- **Confirmation bias**
- **Hindsight bias**
- **Framing effect**

These deviations are often unconscious; yet they may systematically shape outcomes. Cognitive bias therefore differs from the classical notion of "actual bias." The issue is not deliberate favoritism, but structural tendencies rooted in the cognitive architecture of decision-making.

Why Is It Important?

Even if cognitive bias does not formally distort the structure of legal reasoning, it may lead to systematic deviations in outcomes. This can negatively affect:

- The realization of individual justice,
- Consistency across similar cases,
- Public trust in the judiciary.

Society presumes that courts operate objectively and impartially. However, scientific research demonstrates that judges, as human beings, are subject to the inherent limitations of human cognition. This reality compels a reconsideration of both the internal mechanisms of the judicial system and its broader legitimacy in the eyes of the public.

2.1.1.1. Prominent Cognitive Biases in Judicial Decision-Making

Cognitive biases do not arise from a conscious violation of legal norms. Rather, they consist of mental tendencies that systematically influence how information is perceived, selected, interpreted, and weighted during the decision-making process. Emerging through heuristics identified in cognitive psychology, these tendencies may indirectly—and often imperceptibly—undermine the rational coherence of judicial reasoning.

Although such biases do not constitute ethical misconduct in the classical sense, they may weaken the epistemic quality of decisions and erode the normative ideal of impartiality. The most frequently discussed and empirically supported forms of cognitive bias in judicial decision-making are examined below.

i. Hindsight Bias

Hindsight bias refers to the retrospective perception, after learning an outcome, that the result was predictable or even inevitable. First systematically identified by Baruch Fischhoff, research demonstrates that once individuals possess outcome knowledge, they systematically underestimate the degree of prior uncertainty (Fischhoff, 1975).

In judicial contexts, hindsight bias is particularly significant in assessments concerning:

- Fault,

- Negligence,
- Foreseeability,
- The reasonable person standard.

Empirical studies show that judges evaluating events after their occurrence may expect from the actor a level of foresight that would not have been reasonable ex ante (Rachlinski, 1998; Guthrie, Rachlinski & Wistrich, 2007). This tendency may lead, particularly in tort and criminal law, to a retrospective tightening of the "reasonable person" standard.

For this reason, many higher courts emphasize that conduct must be assessed from an ex ante perspective. This doctrinal insistence functions as a legal corrective mechanism against the systematic cognitive error identified in psychological research.

ii. Racial Bias and Implicit Bias

Racial bias often manifests not in explicit or conscious discrimination, but through implicit and automatic associations. The theory of implicit bias developed by Greenwald and Banaji demonstrates that individuals may hold unconscious associations concerning social categories even when they consciously endorse egalitarian values (Greenwald & Banaji, 1995).

Legal scholarship acknowledges that judges are not immune to these implicit cognitive processes. Kang and colleagues argue that implicit racial bias may produce measurable effects particularly in:

- Pretrial detention decisions,
- Bail determinations,
- Sentencing outcomes (Kang et al., 2012).

In the European context, the European Court of Human Rights has recognized the relevance of indirect and structural discrimination in its anti-discrimination jurisprudence. This approach is notably reflected in *Nachova v. Bulgaria*. Although the Court does not explicitly employ the terminology of "implicit bias," its attention to structural and indirect discrimination aligns conceptually with implicit bias theory.

Racial bias in this sense extends beyond individual ill intent; it reflects the automatic influence of socially constructed categories on judicial cognition.

iii. The Rhyme-as-Reason Effect

The rhyme-as-reason effect refers to the phenomenon whereby rhyming or linguistically fluent statements are perceived as more persuasive and credible, independently of their substantive accuracy. Experimental research by McGlone and Tofighbakhsh demonstrates that between two statements conveying identical meaning, the rhyming version is systematically judged as more truthful (McGlone & Tofighbakhsh, 2000).

This effect is closely related to the concept of processing fluency in cognitive psychology. Information that is easier to process is more likely to be perceived as true and reliable.

In judicial contexts, this tendency may indirectly influence decision-making through:

- Oral advocacy,
- Slogan-like legal arguments,
- Repetitive media framing.

Such dynamics risk conflating rhetorical fluency with logical validity. Yet the literature clearly establishes that cognitive fluency does not constitute an epistemic guarantee of truth. The distinction between rhetorical force and legal correctness must therefore be consciously maintained.

iv. Egocentric Bias and the "Bias Blind Spot"

Egocentric bias refers to the tendency of individuals to overestimate their own objectivity while more readily identifying bias in others. The concept of the "bias blind spot," developed by Pronin, Lin, and Ross, demonstrates that individuals exhibit systematic blindness to their own cognitive biases (Pronin et al., 2002).

In judicial settings, this tendency may reduce the perceived need for self-critical reflection, as judges may assume their own neutrality. The phenomenon may be particularly salient in single-judge proceedings.

Modern judicial systems incorporate several institutional mechanisms that may function as counterbalances to egocentric bias:

- The obligation to provide reasoned judgments (externalization and reviewability of reasoning),
- Collegial court structures (collective deliberation),
- Appellate and cassation review (hierarchical oversight mechanisms).

These mechanisms represent structural design solutions aimed at compensating for individual cognitive limitations at the institutional level.

Interim Conclusion

These forms of cognitive bias arise not from conscious favoritism, but from limitations inherent in human cognition. Judicial impartiality, therefore, is not merely an ethical or disciplinary issue. It is a multilayered problem engaging cognitive psychology, behavioral law, and institutional design.

Recognizing this dimension is essential for understanding the structural conditions under which judicial independence and impartiality can be meaningfully safeguarded in contemporary legal systems.

2.1.2. Social and Cultural Bias

Social and cultural bias refers to situations in which a judge—consciously or unconsciously—bases evaluative judgments on social stereotypes, group identities, or cultural codes. While closely related to the literature on implicit bias, this category extends beyond individual psychology to encompass the influence of broader social structures.

Psychological research demonstrates that individuals may unconsciously develop evaluative patterns based on categories such as:

- Gender,
- Ethnicity,
- Religion,
- Socioeconomic status (Greenwald & Krieger, 2006).

Empirical studies in American legal scholarship indicate that, in criminal proceedings, the race or social status of the defendant may exert statistically measurable effects on sentencing outcomes (Rachlinski et al., 2009). In the European context, the European Court of Human

Rights has indirectly acknowledged the relevance of social context by placing particular emphasis on the principle of "appearance" in its impartiality analysis (see *Hauschildt v. Denmark*).

Social and cultural bias thus reflects not merely individual psychology, but the influence of social structures and cultural codes on judicial consciousness. Accordingly, it cannot be addressed solely through individual ethical safeguards; it requires educational initiatives, awareness programs, and institutional diversity policies.

2.1.3. Institutional Bias

Institutional bias refers to systematic tendencies arising from the structural characteristics of the judicial system. Unlike individual judicial misconduct, this form of bias is associated with organizational design, hierarchy, performance metrics, and workload pressures.

Examples include:

- The de facto pressure exerted by higher court precedents on lower courts;
- Promotion and performance systems that incentivize particular decision patterns;
- Excessive caseloads leading to superficial reasoning in favor of speed.

In this context, the concept of **internal independence** becomes particularly significant. A judge must be independent not only from the executive, but also from hierarchical pressures within the judiciary. Otherwise, the institutional structure may indirectly shape the outcome of decisions.

Institutional bias does not necessarily involve conscious direction or manipulation. Rather, it reflects structural tendencies embedded in systemic design.

2.1.4. Strategic and Political Bias

Strategic or political bias arises when a judge makes a conscious strategic choice by taking into account political context, anticipated public reaction, or concerns regarding institutional legitimacy.

Judicial behavior literature, particularly in relation to constitutional and supreme courts, has examined models of strategic interaction (Epstein & Knight, 1998). According to this

perspective, judges may consider whether their decisions will provoke institutional conflict with legislative or executive authorities.

Although the European Court of Human Rights does not explicitly employ the term "political bias," its strong emphasis on protecting the judiciary from executive influence (see *Baka v. Hungary*) underscores that strategic pressures may pose serious risks to independence.

Unlike cognitive bias, strategic bias may be grounded in conscious and rational calculation. For this reason, it constitutes a more visible and normatively assessable risk under classical impartiality standards.

2.2. Algorithmic Bias: A New Risk Domain in Judicial Decision-Making

Algorithmic bias refers to situations in which artificial intelligence and automated decision-support systems produce systematic and foreseeable outcomes disadvantaging certain individuals or groups due to the datasets they employ, their model architectures, optimization objectives, or the normative assumptions embedded in their design (Barocas & Selbst, 2016; Selbst et al., 2019).

Unlike classical human prejudice, this form of bias often becomes obscured under the rhetoric of "computational objectivity" and "mathematical neutrality." Yet the literature clearly demonstrates that algorithms:

- Learn from historical data;
- Are optimized to maximize specific performance metrics;
- Reflect the normative choices of their designers.

Algorithmic decision systems are therefore not merely technical instruments; they constitute sociotechnical governance mechanisms (Jasanoff, 2004; Pasquale, 2015). Algorithmic bias should thus be understood as a technologically encoded and scaled form of human bias.

2.2.1. Structural Sources of Algorithmic Bias

There is relative consensus in international scholarship that algorithmic bias arises from three principal sources: data, model architecture, and objective function. Together, these elements define the normative framework of algorithmic decision-making.

(i) Data-Driven Bias

Machine learning systems extract patterns from historical data. However, particularly in areas such as criminal justice, public safety, and law enforcement, historical datasets often reflect:

- Overrepresentation of certain groups;
- Structural inequalities;
- Selective policing practices.

As emphasized by Barocas and Selbst, when datasets contain historical discrimination, algorithms may reproduce inequality through proxy variables even if they do not explicitly include discriminatory features (Barocas & Selbst, 2016). This dynamic is commonly summarized as the principle of "bias in, bias out."

The problem extends beyond merely reflecting the past. Algorithmic systems project historical inequalities into the future, thereby functioning as mechanisms of normative reproduction. Data thus assumes not only a descriptive, but also a regulatory role.

(ii) Model Architecture and Variable Selection

The choice of which variables an algorithm considers, excludes, or weights may appear technical, yet it is fundamentally normative (Selbst et al., 2019).

For example:

- Prior criminal record,
- Age,
- Residential location

may appear legally neutral. However, due to their structural correlations with particular social groups, they may generate indirect discrimination.

For this reason, the "fairness through unawareness" approach—excluding sensitive variables from the model—has been strongly criticized in the literature. Discrimination often operates not directly, but through statistical correlations and proxy indicators (Kleinberg, Mullainathan & Raghavan, 2017).

Model architecture, therefore, is not merely an engineering problem; it is a value-laden design choice.

(iii) Objective Function and Optimization Logic

Algorithms are designed to maximize specific performance criteria, such as:

- Accuracy,
- Reduction of error rates,
- Cost minimization,
- Speed and efficiency.

Yet the foundational principles of legal systems—equality, individualization, proportionality, and reason-giving—cannot be reduced to statistical performance metrics.

As demonstrated by Kleinberg and colleagues, it is mathematically impossible in many circumstances to satisfy multiple fairness metrics simultaneously (Kleinberg et al., 2017). This impossibility reveals that algorithmic design inevitably entails normative choices. The question of which type of error (false positive or false negative) is more tolerable is not technical, but legal and ethical in nature.

Algorithmic bias, therefore, emerges not as an accidental malfunction, but as a structural consequence of data selection, model design, and optimization logic.

2.2.2. Algorithmic Bias and Human Rights Risks in the Judicial Context

The use of algorithmic systems in the judicial field is becoming increasingly widespread in areas such as:

- pre-trial detention assessments,
- sentencing decisions,
- risk analysis, and
- case prioritization.

This development elevates algorithmic bias to a constitutional and human rights issue.

(i) Prohibition of Discrimination

The case law of the European Court of Human Rights defines indirect discrimination as situations in which seemingly neutral practices produce disproportionate effects on particular groups (e.g., *D.H. and Others v. Czech Republic*; *Nachova v. Bulgaria*).

Even where algorithmic systems do not explicitly use variables such as race or gender, they may reproduce structural inequalities through indirect indicators such as:

- region,
- level of education,
- income level, and
- family structure.

For this reason, algorithmic bias directly engages with the ECtHR's doctrine of indirect discrimination.

(ii) Equality of Arms and Procedural Fairness

The opacity of "black box" algorithms, whose internal logic cannot be examined by the defense, creates a serious information asymmetry in criminal proceedings (Pasquale, 2015).

The principle of equality of arms under Article 6 of the European Convention on Human Rights requires that parties be able to effectively present their claims and defenses. However, if:

- the parameters of the algorithmic model,
- its error rates, and
- its threshold values

cannot be tested or contested by the defense, procedural fairness may be compromised.

This situation poses a significant risk not only to substantive justice, but also to the adversarial and reviewable character of judicial proceedings.

(iii) The Right to a Reasoned Judgment and the Problem of Explainability

The European Court of Human Rights recognizes the right to a reasoned judgment as one of the core elements of a fair trial (see *Hadjianastassiou v. Greece*).

However, deep learning–based systems in particular raise serious concerns regarding explainability. If a judge bases a decision on an algorithmic risk score, a fundamental question arises:

Can a statistical output replace legal reasoning?

Legal reasoning is not merely the announcement of a result; it is the transparent presentation of the chain of reasoning. An algorithmic system lacking explainability may, in practice, render the obligation to provide reasons meaningless.

2.2.3. The Collapse of the "Neutral Technology" Myth

Algorithms were long perceived as neutral and objective tools. This understanding has been critically challenged in contemporary scholarship.

Pasquale's *Black Box Society* approach argues that algorithmic decision-making systems generate a new form of power that evades democratic accountability (Pasquale, 2015). Similarly, the literature in science and technology studies characterizes algorithms as sociotechnical structures that:

- encode human values,
- reflect institutional power relations, and
- produce political consequences (Jasanoff, 2004).

Accordingly, the core issue is not technical accuracy, but legal legitimacy, democratic oversight, and compatibility with human rights.

2.3. The Structural Relationship Between Cognitive Bias and Algorithmic Bias

The original contribution of this article lies in revealing the structural continuity between cognitive and algorithmic bias.

Cognitive bias is:

- individual,
- psychological,
- limited in scale, and

- partially controllable through reasoning and appellate review.

Algorithmic bias, by contrast, is:

- institutional,
- encoded,
- scalable,
- standardized, and
- difficult to challenge and render visible.

Algorithmic bias represents the version of cognitive prejudice that has been:

- translated into technical language,
- automated, and
- rendered applicable on a large scale.

If human prejudice constitutes a "micro-deviation," algorithmic bias is its "macro-replicated" form. For this reason, the classical debate on judicial impartiality shifts from the level of individual ethics to the level of institutional design and technological governance.

In conclusion, unlike classical cognitive biases, algorithmic bias emerges not merely at the individual level, but at the institutional and structural level, and it operates on a much broader scale. This reality requires the judicial system to focus not only on the impartiality of judges, but also on the design, supervision, and limitation of the technological tools employed.

In this context, algorithmic bias should be regarded as a new-generation impartiality problem for modern judicial systems.

2.4. Algorithmic Bias within the Framework of the European Union and the Council of Europe

Legal debates concerning algorithmic bias have, in recent years, gained significant prominence within European human rights and public law frameworks. Although the European Court of Human Rights (ECtHR) has not yet developed direct case law specifically addressing algorithmic decision-making systems, its established jurisprudence on the prohibition of discrimination, the right to a fair trial, and the obligation to provide reasoned judgments offers strong normative foundations for the scrutiny of algorithmic decisions.

In particular, the ECtHR's approach grounded in indirect discrimination and proportionality analysis brings into focus the State's positive obligations in situations where seemingly neutral practices produce disproportionate effects on certain groups. Within this framework, algorithmic bias should be regarded not merely as an isolated rights violation, but as a structural human rights problem.

At the level of the European Union, Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act) classifies AI systems used in the judiciary and law enforcement as "high-risk" and imposes comprehensive obligations. The Regulation, in particular, addresses algorithmic bias through requirements concerning:

- ensuring the representativeness of datasets,
- prior assessment of bias risks,
- human oversight,
- technical documentation and traceability, and
- transparency obligations.

Through these regulatory mechanisms, algorithmic bias is positioned not as a mere technical malfunction, but as a normative risk domain.

Similarly, the Council of Europe's work on artificial intelligence and human rights emphasizes the legal accountability of algorithmic systems and the principle of human-centered design. This approach reflects an effort to establish a normative balance between technological innovation and the protection of fundamental rights.

In conclusion, algorithmic bias is not an alternative to cognitive bias; rather, it is its institutionalized and scaled form. Therefore, debates concerning the use of artificial intelligence in judicial processes should be conducted not primarily in terms of technical efficiency, but within the framework of legal legitimacy, human rights, and democratic accountability.

At this juncture, the fundamental question is:

Is judicial impartiality alone sufficient, or must the technological tools upon which judges rely also be subjected to constitutional standards of impartiality?

This question compels a reconsideration of the literature on judicial independence and impartiality within the context of the twenty-first century.

IV. THE RISE OF ALGORITHMIC AUTHORITY AND THE TRANSFORMATION OF JUDICIAL LEGITIMACY

(The Source of Authority: Human or Algorithm?)

Algorithmic bias cannot be assessed merely as a technical weakness of the principle of judicial impartiality. At a deeper level, the issue concerns the transformation of the source of authority in judicial decision-making. The increasing use of algorithmic systems has the potential to reshape the traditional foundations that underpin the normative legitimacy of the judiciary.

The fundamental question is:

Does judicial authority derive from law, or from a claim of statistical accuracy?

1. The Rational–Legal Foundation of Judicial Authority

According to Max Weber's classical typology of authority, the legitimacy of the modern state and judiciary rests on "rational–legal authority." This form of authority is built upon:

- adherence to normative rules,
- institutional authorization, and
- public justification.

The legitimacy of a judicial decision derives not solely from the correctness of its outcome, but from its grounding in law, its reviewability, and its reasoned justification. The European Court of Human Rights likewise recognizes the right to a reasoned judgment as a core component of the right to a fair trial[30].

By contrast, the legitimacy claims of algorithmic decision-support systems often rely on:

- accuracy rates,
- error minimization, and

[30] Hadjianastassiou vs. Greece, App. No. 12945/87 (ECtHR, 1992)

- statistical consistency.

Thus, a structural tension emerges between legal normativity and computational accuracy.

2. Statistical (Algorithmic) Authority and Justice

Algorithms generate probabilities; they do not generate normative reasons. A risk score may predict future behavior, but the legal values through which that prediction is interpreted remain a matter of normative choice.

As demonstrated by Jon Kleinberg and his colleagues, it is often mathematically impossible to satisfy different fairness metrics simultaneously. This finding indicates that algorithmic fairness is not purely technical; it rests on normative preferences.

According to Jürgen Habermas's conception of legitimacy, the legitimacy of a legal decision stems from its capacity to be publicly justified. Legitimacy derives not merely from a correct outcome, but from the discursive defensibility of its reasoning. Yet many algorithmic systems function as "black boxes," incapable of adequately explaining the logical chain behind their outputs (Pasquale).

In such circumstances, statistical accuracy risks replacing normative justification. This transformation may weaken the rational–legal foundation of the judiciary.

3. Algorithmic Authority and Democratic Accountability

Frank Pasquale argues that algorithmic systems generate a form of invisible authority. This authority:

- is grounded in technical expertise,
- may be resistant to external scrutiny, and
- can evade traditional mechanisms of accountability.

Judicial power, however, constitutes a constitutionally non-delegable form of public authority. If the decisive reference point in a judge's decision becomes an algorithmic score, a de facto shift of authority may occur. As Danielle Keats Citron has observed, automated decision systems risk undermining procedural safeguards.

Accordingly, algorithmic decision-support tools are not merely technical assistants; they are actors that shape the normative framework of the decision itself.

4. Automation Bias and Judicial Independence

Psychological research shows that individuals tend to over-rely on automated systems. This phenomenon, known as "automation bias," may lead human decision-makers to perceive algorithmic outputs as more objective and reliable.

Behavioral law studies further demonstrate that judges can be influenced by numerical reference points (the anchoring effect). In this context, an algorithmic score may function as an anchor that indirectly guides independent judicial reasoning.

The issue here is not only bias, but the transformation of judicial autonomy.

5. The Redefinition of Judicial Legitimacy

Ultimately, the debate on algorithmic bias converges on the question of legitimacy. The problem is not solely whether a decision is fair, but upon which authority the decision claims legitimacy.

If judicial authority begins to rely:

- on statistical accuracy instead of normative justification,
- on technical expertise instead of public accountability, and
- on probabilistic modeling instead of legal argumentation,

the rational–legal legitimacy of the judiciary undergoes a structural transformation.

For this reason, the integration of algorithmic systems into judicial processes should not be assessed solely in terms of efficiency and accuracy. It must be evaluated within the framework of constitutional legitimacy, democratic accountability, and the principles of normative justification.

In this sense, the problem of algorithmic bias is not merely a technical modeling issue; it is a debate about the ontological position of the judiciary in the twenty-first century.

V. A SOLUTION TO BIAS PROBLEMS:

The Erkan Constitutional Algorithmic Safeguard Model (ECASM)

In the previous sections of this study:

- human-based cognitive bias types were systematically categorized,
- the structural sources of algorithmic bias were identified,
- its relationship with European human rights standards was analyzed, and
- the transformative impact of algorithmic authority on judicial legitimacy was discussed.

In this section, the study moves beyond descriptive analysis and advances a constructive model. The objective is to propose a constitutionally grounded, normatively coherent, and internationally applicable oversight architecture addressing the problem of algorithmic bias.

1. The Multi-Layered Algorithmic Oversight Model

(MLAOM – Erkan Constitutional Algorithmic Safeguard Model)

Core Thesis

The integration of algorithmic decision-support systems into judicial processes must not be subjected merely to technical accuracy testing; it must be placed under a multi-layered, normative, and institutional constitutional review.

The point of departure of the model is as follows:

"Algorithmic decision-support systems should not be supervised as ordinary administrative acts, but as normative instruments that generate constitutional risk."

Table 1: Erkan Constitutional Algorithmic Safeguard Model (ECASM)

Layer	Type of Review	Primary Objective	Review Questions	Legal Basis	Type of Risk
I	**Normative Compliance Review**	To ensure the algorithm's conformity with constitutional and human rights principles	- Is there a risk of discrimination? - Has proportionality been observed? - Is the right to a fair trial protected?	ECHR Art. 6, Art. 14; constitutional equality principle	Structural discrimination
II	**Transparency and Explainability Review**	To ensure that algorithmic outputs are capable of reasoned justification	- Is the decision logic explainable? - Can parties challenge the system? - Does it function as a black box?	Right to a reasoned judgment; procedural safeguards	Black-box authority
III	**Institutional Responsibility Review**	To safeguard the non-delegable nature of judicial authority	- Who bears final responsibility? - Is the algorithm binding? - Is there meaningful human control?	Judicial independence; separation of powers	Authority shift
IV	**Behavioral Impact Review**	To protect the cognitive independence of the judge	- Is there automation bias? - Does the algorithm create a reference (anchoring) effect?	Principle of impartial adjudication; independent judicial reasoning	Cognitive steering

Within this framework, the model consists of four layers.

I. Layer: Normative Compatibility Review

(Normative Compatibility Layer)

The purpose of this layer is to test the algorithm's conformity with constitutional values and human rights norms.

The core questions are:

- Do the variables used entail a risk of indirect discrimination?
- Is the system compatible with the right to a fair trial and the prohibition of discrimination?
- Have the principles of equality and proportionality been observed?

This layer treats the algorithm not as a purely technical instrument, but as a normative object subject to constitutional review.

Proposed mechanisms:

- Ex ante constitutional compliance review
- Mandatory Human Rights Impact Assessment

Through this layer, algorithmic systems are subjected to a constitutional filter before entering judicial practice.

II. Layer: Transparency and Explainability Review

(Transparency & Explainability Layer)

Explainability constitutes the core of algorithmic legitimacy.

This layer examines the following criteria:

1. Is the decision logic understandable?
2. Can the parties effectively challenge the algorithmic assessment?
3. Is the algorithm's contribution explicitly indicated in the reasoned judgment?

If a system:

- functions as a black box, or
- is kept closed on the grounds of trade secrecy,

its judicial use should be constitutionally restricted.

This layer represents the digital-age extension of the right to a reasoned judgment and aims to safeguard procedural fairness guarantees.

III. Layer: Institutional Responsibility and Authority Review

(Institutional Accountability Layer)

One of the original dimensions of the model is its institutional responsibility architecture.

The central question is: Who is responsible for an algorithmic error?

The model envisages a tripartite responsibility structure:

1. Developer liability
2. Administrative authority responsibility
3. Final judicial responsibility

Within this framework:

- The algorithm cannot replace the judge.
- It cannot have binding force.
- It must be explicitly positioned in the judgment as an "auxiliary tool."

This layer safeguards the principle of the non-delegability of judicial power and aims to prevent de facto shifts of authority.

IV. Layer: Behavioral Impact Review

(Behavioral Integrity Layer)

Algorithms do not merely generate outputs; they may also influence the cognitive framework of the decision-maker.

For this reason, the model explicitly incorporates behavioral risks into the oversight structure, including:

- automation bias testing,
- anchoring effect analysis, and
- independent reasoning verification.

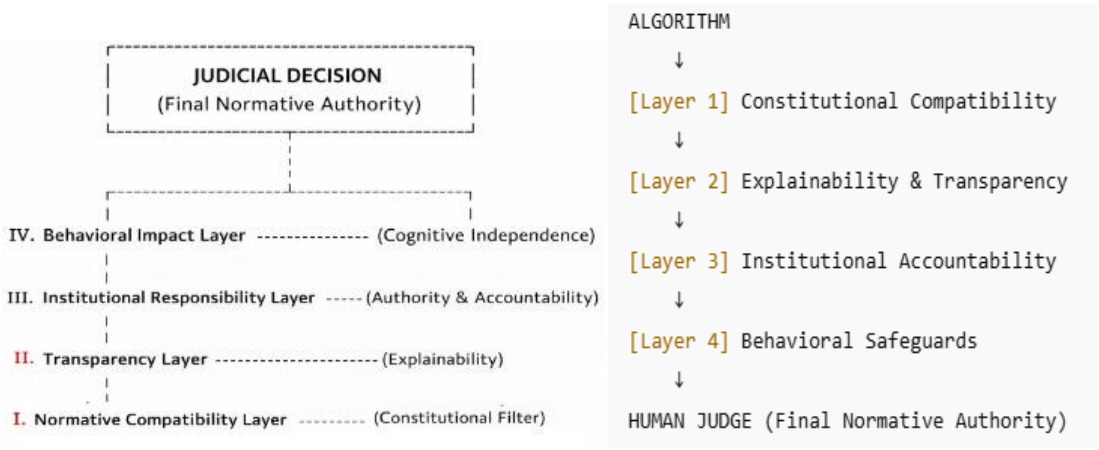
Proposed practical safeguards:

- Requiring the judge to draft the decision outline before viewing the algorithmic score
- Conducting alternative scenario testing
- Explicitly demonstrating independent human reasoning within the decision text

By addressing cognitive and algorithmic bias within an integrated framework, this layer offers a theoretical contribution to the literature.

2. The Theoretical Positioning of the Model

The Erkan Constitutional Algorithmic Safeguard Model conceptualizes the algorithmic system through a "layered constitutional filter" approach.



The core principles of the model are as follows:

- The algorithm does not decide.
- Final normative authority rests with the human decision-maker.
- Technology must pass through a constitutional filter.

This approach does not prohibit algorithmic systems; rather, it situates them within constitutional boundaries.

3. The Academic Innovation of the Model

The model distinguishes itself from existing approaches in the literature in several respects:

- It does not rely solely on fairness metrics.
- It does not reduce the issue to data protection law.
- It is not merely a declaration of ethical principles.
- It integrates the behavioral law dimension into institutional oversight.
- It positions judicial independence as the central normative principle.

In this respect, the model proposes a constitutional safeguard architecture against the risk of "Algorithmic Delegation of Judicial Power."

4. Interdisciplinary Integration

The model provides a structural framework that integrates:

- constitutional law,
- human rights law,

- judicial theory,
- behavioral law, and
- algorithmic governance.

In this sense, it is not merely a technical oversight model, but a constitutional protection architecture.

5. Universal Applicability

The model is applicable:

- in Anglo-Saxon systems where risk assessment tools are used,
- in Continental European systems undergoing digital judicial reforms, and
- in legal systems with developing institutional capacity.

This is because the model is grounded not in a specific technological preference, but in the preservation of legal legitimacy.

Concluding Principle

Algorithmic systems should not be categorically prohibited. However, an unsupervised algorithm transforms not only judicial impartiality, but judicial legitimacy itself.

The Multi-Layered Algorithmic Oversight Model seeks not merely to reduce algorithmic bias, but to ensure that judicial authority remains within constitutional limits.

The future of the judiciary will inevitably be digital; yet its legitimacy can be preserved only within the normative responsibility of human reasoning.

Algorithms may support decisions; they cannot represent justice.

VI. CONCLUSION

The increasing integration of artificial intelligence–assisted decision systems into judicial processes compels a fundamental reassessment of classical assumptions concerning judicial independence and impartiality. This study has demonstrated that judicial impartiality can no longer be understood solely within the framework of the individual judge's ethical responsibility or traditional institutional safeguards. Rather, it has evolved into a multi-layered

constitutional problem field that must be examined in conjunction with cognitive psychology, behavioral law, and algorithmic decision architecture.

The systematic classification of cognitive bias types undertaken in this study reveals that human judicial decision-making may be shaped by heuristic shortcuts, contextual influences, and implicit prejudices. This finding confirms that judicial impartiality is not a self-executing condition, but a fragile constitutional ideal requiring continuous protection at both institutional and cognitive levels.

Yet the issue extends beyond human-based biases. The study has shown that algorithmic bias introduces a less visible, yet structurally embedded, dimension of risk into judicial processes. Algorithms are frequently presented as "objective" and "neutral" instruments. However, the reproduction of historical inequalities within datasets, normative choices embedded in model design, and optimization objectives may generate serious risks of violating the prohibition of discrimination and the right to a fair trial. In this context, risk assessment systems illustrate that algorithmic decisions are not merely questions of technical accuracy; they are directly connected to the principles of equality, reasoned justification, and equality of arms.

The central contribution of this study lies in conceptualizing algorithmic bias not as a mere technical malfunction or data problem, but as a normative issue situated within the broader transformation of judicial authority. The rise of algorithmic systems carries the potential to shift the source of legitimacy of judicial decisions from legal normativity to claims of statistical accuracy. Yet statistical accuracy cannot substitute for normative justification; algorithmic computation does not generate constitutional legitimacy.

Accordingly, the fundamental question is not limited to whether algorithms are biased. The deeper concern is whether judicial authority will remain anchored in normative responsibility or be reshaped around claims of technical correctness. This question necessitates a rethinking of the classical rational–legal foundation of the judiciary within the context of the digital age.

Guided by this theoretical framework, the study has proposed an original Multi-Layered Algorithmic Oversight Model. The model does not categorically reject the integration of algorithmic systems into judicial processes; rather, it situates them within a constitutional filter architecture. Comprising four layers—normative compatibility, explainability, institutional

accountability, and behavioral integrity—the model provides for the supervision of algorithmic tools in accordance with human rights, judicial independence, and democratic accountability.

This approach treats algorithmic systems not as ordinary technical instruments, but as normative actors capable of generating constitutional risk. In doing so, the proposed model seeks not merely to mitigate discrimination, but to preserve the institutional foundations of legitimacy within an increasingly digitalized judicial environment.

The study further demonstrates that human cognitive bias and algorithmic bias are not independent processes; on the contrary, they often operate in mutually reinforcing interaction. Behavioral mechanisms such as automation bias and the anchoring effect may lead judges to over-rely on algorithmic outputs, thereby subordinating normative reasoning to computational suggestion. Judicial impartiality must therefore be reconsidered not only in light of the limitations of the human mind, but also in relation to the holistic design of human–machine interaction.

In conclusion, in the age of artificial intelligence, judicial impartiality can be reduced neither to the virtue of the individual judge nor to the technical accuracy of algorithmic systems. Impartiality emerges as a dynamic constitutional ideal that requires the relationship between human reasoning and algorithmic computation to be reconstructed in a manner that is transparent, reviewable, justifiable, and compatible with human rights.

The digitalization of the judiciary is inevitable; yet its legitimacy will continue to rest upon the constitutional reasoning of human decision-makers who bear normative responsibility. Algorithms may support decisions, but they cannot represent justice.

The sustainability of the rule of law in the digital age depends upon preserving the equilibrium between technological efficiency and normative legitimacy. The Multi-Layered Algorithmic Oversight Model should therefore be regarded as an institutional proposal aimed at securing this balance on constitutional grounds.

REFERENCES

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective." *PeerJ Computer Science* 2 (2016): e93.

Ammar, A., A. Koubaa, B. Benjdira, O. Nacar, and S. Sibaee. "Prediction of Arabic Legal Rulings Using Large Language Models." *Electronics* 13, no. 4 (2024): 764.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine Bias." *ProPublica*, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Ashley, Kevin D. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age.* Cambridge: Cambridge University Press, 2017.

Baka v Hungary, App. No. 20261/12 (ECtHR, 23 June 2016).

Baker, Steven. "Ensuring Procedural Fairness in AI-Driven Criminal Sentencing: A Focus on Transparency and the Right to Explanation." *National High School Journal of Contemporary Scholarship*, February 27, 2025.

Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." *California Law Review* 104, no. 3 (2016): 671–732.

Busuioc, Madalina, and B. Alon. "Human–AI Interactions in Public Sector Decision Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice." *Journal of Public Administration Research and Theory* (2022). https://doi.org/10.1093/jopart/muac007.

Chalkidis, Ilias, et al. "Legal-BERT: The Muppets Straight Out of Law School." *Findings of EMNLP* (2020): 2898–2904.

Chouldechova, Alexandra. "Fair Prediction with Disparate Impact." *Big Data* 5, no. 2 (2017): 153–163.

Citron, Danielle Keats. "Technological Due Process." *Washington University Law Review* 85 (2008): 1249–1313.

Dressel, Julia, and Hany Farid. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Science Advances* 4, no. 1 (2018).

Ebner v Official Trustee in Bankruptcy (2000) 205 CLR 337 (HCA).

Epstein, Lee, and Jack Knight. *The Choices Justices Make.* Washington, DC: CQ Press, 1998.

Erkan, Fikret. "Scope And Legal Binding Of The European Court Of Human Rights Decisions", *Journal of Economics and Social Research*, Number 20, p. 24-31, 2023.

Erkan, Fikret. "Artificial Intelligence and Liability for Damages." *International Journal of Social Sciences* 9, no. 12 (2025): 297–343.

European Commission for the Efficiency of Justice (CEPEJ). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems.* Strasbourg: Council of Europe, 2020.

European Data Protection Board. "Guidelines on Automated Individual Decision-Making and Profiling." 2018.

Fischhoff, Baruch. "Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty." *Journal of Experimental Psychology: Human Perception and Performance* 1, no. 3 (1975): 288–299.

Greenwald, Anthony G., and Mahzarin R. Banaji. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102, no. 1 (1995): 4–27.

Greenwald, Anthony G., and Linda Hamilton Krieger. "Implicit Bias: Scientific Foundations." *California Law Review* 94 (2006): 945–967.

Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. "Inside the Judicial Mind." *Cornell Law Review* 86 (2001): 777–830.

Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. "Blinking on the Bench: How Judges Decide Cases." *Cornell Law Review* 93 (2007): 1–44.

Habermas, Jürgen. *Between Facts and Norms.* Cambridge, MA: MIT Press, 1996.

Hadjianastassiou v Greece, App. No. 12945/87 (ECtHR, 16 December 1992).

Hauschildt v Denmark, App. No. 10486/83 (ECtHR, 24 May 1989).

Hildebrandt, Mireille. "Law as Computation in the Era of Artificial Legal Intelligence." *Osgoode Hall Law Journal* 57, no. 1 (2020): 101–128.

Hillman, Noel L. “The Use of Artificial Intelligence in Gauging the Risk of Recidivism.” *The Judges’ Journal* (2025).

Huchhanavar, Shivaraj S. “Conceptualising Judicial Independence and Accountability from a Regulatory Perspective.” *Oslo Law Review* 9, no. 2 (2023): 110–148.

Jasanoff, Sheila. *States of Knowledge: The Co-Production of Science and Social Order.* London: Routledge, 2004.

Jesper Ryberg, “Artificial Intelligence at Sentencing: When Do Algorithms Perform Well Enough to Replace Humans?,” *AI and Ethics* 5 (2025): 1009–1018.

Kahneman, Daniel. *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux, 2011.

Katz, Daniel Martin, Michael J. Bommarito II, and Josh Blackman. “A General Approach for Predicting the Behavior of the Supreme Court of the United States.” *PLoS ONE* 12, no. 4 (2017): e0174698.

Kelly Hannah-Moffat, “Actuarial Sentencing: An ‘Unsettled’ Proposition,” *Justice Quarterly* 30, no. 2 (2013): 270–296, esp. 281–288;

Kelly Hannah-Moffat and Paula Maurutto, “Re-Contextualizing Pre-Sentence Reports,” *Punishment & Society* 12, no. 3 (2010): 262–286.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. “Inherent Trade-Offs in the Fair Determination of Risk Scores.” *Proceedings of the 8th Innovations in Theoretical Computer Science Conference* (2017).

Kosař, David, and Samuel Spáč. “Judicial Independence.” In *The Cambridge Handbook of Constitutional Theory*, edited by Richard Bellamy and Jeff King. Cambridge: Cambridge University Press, 2025.

Lum, Kristian, and William Isaac. “To Predict and Serve?” *Significance* 13, no. 5 (2016): 14–19.

Menezes-Neto, E. J. de, and M. B. M. Clementino. “Using Deep Learning to Predict Outcomes of Legal Appeals Better than Human Experts.” *PLoS ONE* 17, no. 7 (2022): e0272287.

Oswald, Marion. "A Three-Pillared Approach to Ensuring the Trustworthy, Accountable and Transparent Use of Artificial Intelligence in Policing." *European Journal of Law and Technology* 13, no. 1 (2022).

Papagianneas, Straton. "Automating Intervention in Chinese Justice: Smart Courts and Supervision Reform." *Asian Journal of Law and Society* 10, no. 3 (2023): 463–489.

Parasuraman, Raja, and Victor Riley. "Humans and Automation: Use, Misuse, Disuse, Abuse." *Human Factors* 39, no. 2 (1997): 230–253.

Pasquale, Frank. *The Black Box Society.* Cambridge, MA: Harvard University Press, 2015.

Piersack v Belgium, App. No. 8692/79 (ECtHR, 1 October 1982).

Pronin, Emily, Daniel Y. Lin, and Lee Ross. "The Bias Blind Spot." *Personality and Social Psychology Bulletin* 28, no. 3 (2002): 369–381.

Rachlinski, Jeffrey J. "A Positive Psychological Theory of Judging in Hindsight." *University of Chicago Law Review* 65 (1998): 571–625.

Rachlinski, Jeffrey J., Andrew J. Wistrich, and Chris Guthrie. "Can Judges Make Reliable Numeric Judgments?" *Indiana Law Journal* 90 (2009): 695–739.

Russell, Stuart J., and Peter Norvig. *Artificial Intelligence: A Modern Approach.* 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2010.

Selbst, Andrew D., et al. "Fairness and Abstraction in Sociotechnical Systems." *Proceedings of FAT* (2019).

Sharma, M. "India's Courts and Artificial Intelligence: A Future Outlook." *LeXonomica* 15, no. 1 (2023): 99–120.

Singapore Judiciary. "Media Release: New Generative AI-Powered Case Summarisation Tool." September 10, 2025.

Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* 104, no. 3 (2016): 671–732.

Sunstein, Cass R., et al. *Punitive Damages: How Juries Decide.* Chicago: University of Chicago Press, 2002.

Surden, Harry. "Machine Learning and Law." *Washington Law Review* 89, no. 1 (2014): 87–115.

Tversky, Amos, and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185, no. 4157 (1974): 1124–1131.

Vaswani, Ashish, et al. "Attention Is All You Need." *Advances in Neural Information Processing Systems* (2017).

Weber, Max. *Economy and Society.* Berkeley: University of California Press, 1978.

Wistrich, Andrew J., Chris Guthrie, and Jeffrey J. Rachlinski. "Heart Versus Head." *Texas Law Review* 93 (2015): 855–923.

Zhang, Y., and X. Li. "How Do Judges Use Large Language Models? Evidence from Shenzhen." *Journal of Legal Analysis* (2025).