

**INTERNATIONAL JOURNAL OF LAW
MANAGEMENT & HUMANITIES**
[ISSN 2581-5369]

Volume 9 | Issue 5

2026

© 2026 *International Journal of Law Management & Humanities*

Follow this and additional works at: <https://www.ijlmh.com/>

Under the aegis of VidhiAagaz – Inking Your Brain (<https://www.vidhiaagaz.com/>)

This article is brought to you for free and open access by the International Journal of Law Management & Humanities at VidhiAagaz. It has been accepted for inclusion in the International Journal of Law Management & Humanities after due review.

In case of **any suggestions or complaints**, kindly contact support@vidhiaagaz.com.

To submit your Manuscript for Publication in the **International Journal of Law Management & Humanities**, kindly email your Manuscript to submission@ijlmh.com.

ANCHOR Guidelines: A Constitutional Blueprint for Human-Centred Judicial AI in India

ADARSH RAY* AND SUHANI ANAND**

ABSTRACT

In a rapidly advancing technological era, there is an urgent need to address how Artificial Intelligence (AI) can be integrated into India's overloaded courts without forgoing constitutional safeguards. To that end, this article surveys and analyses thirty years of the development of AI in the legal field, empirical research on bias and public confidence, and the regimes of the United States, China and Europe, in order to identify the gaps in explainability, accountability and doctrinal congruence. Anchored in Articles 14 and 21, the doctrine of the separation of powers, and judicial review under Articles 32 and 226, the article presents the ANCHOR (Accountability, Non-delegation, Constitutional conformity, Human Oversight, and Review) Guidelines to govern AI as a co-judge and not a surrogate. ANCHOR requires a risk-based liability framework, mandatory explainability requirements, periodic bias audits and effective human-override procedures. To test the framework, the article sets out a pilot in three diverse courts, with metrics on case resolution rates, judge-AI agreement and litigant confidence. By combining constitutional theory with a clear roadmap for implementation, the work offers a feasible, rights-upholding template for the use of AI in Indian courts.

Keywords: ANCHOR, Artificial Intelligence, Co-Judge, Explainability, Accountability

I. INTRODUCTION

The administration of justice around the world is undergoing a profound shift as Artificial Intelligence (AI) is incorporated, at an accelerating pace, into judicial systems. The Wisconsin Supreme Court's 2016 decision upholding the sentence of Eric Loomis, which had been informed in part by an opaque algorithmic risk score,¹ and the use of machine learning in China's internet courts, including Beijing's, to generate judgments automatically for judges,² illustrate this shift. These developments reflect the growing presence of AI in the administration of justice.

* Author is a Student at Vivekananda Institute of Professional Studies, New Delhi, India.

** Author is a Student at Vivekananda Institute of Professional Studies, New Delhi, India.

¹ *State v. Loomis*, 2016 WI 68, 881 N.W.2d 749 (Wis. 2016).

² Changqing Shi, Tania Sourdin & Bin Li, *The Smart Court – A New Pathway to Justice in China?*, 12 INT'L J. FOR CT. ADMIN. art. 4 (2021), <https://doi.org/10.36745/ijca.367>.

This integration, however, is taking place amid an institutional crisis confronting conventional court systems. Administrative inefficiency, case backlogs and budgetary shortfalls have strained critical judicial infrastructure and limited timely access to constitutional remedies. In this setting, AI is emerging not merely as an amenity but as a force capable of substantially improving the efficiency, consistency and accessibility of the legal process.

This technological convergence also raises important questions, notably about transparency and human discretion. In IBM's Global AI Adoption Index 2023, 83% of IT professionals at companies exploring or deploying AI said that being able to explain how their AI reached a decision was important to their business,³ a demand that becomes more urgent still when AI outputs affect life and liberty. Among Indian legal professionals, surveys conducted by Manupatra found that only 4.1% fully trusted AI output without human verification and that 51.5% were concerned about the quality of AI-generated content.⁴

This paper addresses these issues directly, arguing that AI systems must be deliberately designed and deployed to supplement, and not to undermine or replace, the values of fairness and openness inherent in judicial systems. To address gaps in the current literature, the research is guided by one primary question: how can AI be responsibly integrated into judicial systems to enhance efficiency and access while maintaining judicial legitimacy and human accountability?

The need for this study is evident from stark realities. In India, undertrial prisoners make up more than 70% of the prison population (75.8% at the end of 2022),⁵ a severe denial of swift justice. A UNESCO survey published in 2024 found that 44% of the judicial operators who

³ IBM, IBM GLOBAL AI ADOPTION INDEX 2023 (2023) (survey of 8,584 IT professionals conducted by Morning Consult in November 2023); see IBM, *Data Suggests Growth in Enterprise Adoption of AI Is Due to Widespread Deployment by Early Adopters* (Jan. 10, 2024), <https://newsroom.ibm.com/2024-01-10-Data-Suggests-Growth-in-Enterprise-Adoption-of-AI-is-Due-to-Widespread-Deployment-by-Early-Adopters> (83% of IT professionals at companies exploring or deploying AI say that being able to explain how their AI reached a decision is important to their business).

⁴ MANUPATRA ACADEMY, AI ADOPTION AND ITS IMPACT IN THE LEGAL INDUSTRY: AN INDIAN PERSPECTIVE (2024), <https://www.manupatracademy.com/assets/pdf/AI-Adoption-and-Its-Impact-in-the-Legal-Industry-An-Indian-Perspective.pdf> (51.5% concerned about the quality of AI-generated content); MANUPATRA, ADOPTION OF AI IN THE INDIAN LEGAL LANDSCAPE (2025) (227 respondents; 4.1% fully trust AI outputs without human verification), as summarised in *Manupatra Conducts a Survey on AI-Adoption in the India Legal Landscape, First of Its Kind in India*, BUSINESS STANDARD (June 2, 2025), https://www.business-standard.com/content/press-releases-ani/manupatra-conducts-a-survey-on-ai-adoption-in-the-india-legal-landscape-first-of-its-kind-in-india-125060201166_1.html.

⁵ Jayanth K. Krishnan & C. Raj Kumar, *Delay in Process, Denial of Justice: The Jurisprudence and Empirics of Speedy Trials in Comparative Perspective*, 42 GEO. J. INT'L L. 747 (2011); NATIONAL CRIME RECORDS BUREAU, PRISON STATISTICS INDIA 2022, at xii (2023), <https://www.ncrb.gov.in/uploads/nationalcrimerecordsbureau/custom/psiyarwise2022/1701613297PSI2022ason01122023.pdf> (undertrials 75.8% of all prisoners on 31 December 2022).

responded had used AI tools for work-related activities.⁶ The need for principled and ethically sound AI integration has never been more urgent.

This paper presents the ANCHOR Guidelines, a policy and governance blueprint designed to uphold constitutional values, ensure transparency and integrate ethical reasoning at every level. ANCHOR stands for Accountability, Non-delegation, Constitutional Conformity, Human Oversight, and Review, the foundational principles of a benchmark for judicial AI systems. It envisages AI as an intelligent and integral addition that supplements human judgment instead of replacing it. ANCHOR is not a technological prototype but a design for governance: a normative agenda intended to anchor judicial AI within constitutional safeguards and administrative constraints.

In brief, this work draws on doctrinal analysis and policy design to inform policymakers, lawyers and technologists, and seeks to provide legal and institutional guidelines for future AI tools.

II. LITERATURE REVIEW

A. Historical evolution of AI in legal reasoning

Early legal AI systems made explanation an integral part of their architecture, so that users could understand how decisions were reached. Two distinct paradigms were prominent at the first International Conference on Artificial Intelligence and Law (ICAIL), held in Boston in 1987: rule-based reasoning (RBR) systems, which encode legal rules in formal logic, and case-based reasoning (CBR) systems, which draw analogies from prior cases.⁷ RBR offers transparency through explicit rule chains but can struggle with ambiguous, ‘open-textured’ norms; CBR excels at analogical flexibility but often lacks clear explanatory paths.

HYPPO, described by Edwina Rissland and Kevin Ashley at the first ICAIL in 1987, introduced reasoning with legal cases to the field.⁸ A key feature of HYPPO is its ability to generate hypothetical cases in order to test the sensitivity of an argument to the absence or presence of certain facts, to ‘flesh out’ sparse areas of its case knowledge base, and to formulate refinements

⁶ UNESCO, *UNESCO Survey Uncovers Critical Gaps in AI Training Among Judicial Operators* (June 19, 2024), <https://www.unesco.org/en/articles/unesco-survey-uncovers-critical-gaps-ai-training-among-judicial-operators-0>; see Gutiérrez, *infra* note 32, at 7.

⁷ Edwina L. Rissland & Kevin D. Ashley, *A Case-Based System for Trade Secrets Law*, in PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW 60 (ACM 1987), <https://doi.org/10.1145/41735.41743>.

⁸ Guido Governatori et al., *Thirty Years of Artificial Intelligence and Law: The First Decade*, 30 ARTIFICIAL INTELLIGENCE & L. 481 (2022), <https://doi.org/10.1007/s10506-022-09329-4>.

and refutations of an argument.⁹ In 1991, the CABARET system was developed as a successor to HYPO.¹⁰ CABARET, built by Edwina L. Rissland and David B. Skalak, was a hybrid system integrating the two principal reasoning paradigms, RBR and CBR. Its primary purpose was to interpret and apply legal rules containing ambiguous, poorly defined or undefined terms by appealing to previous constructions of those terms in relevant cases. In 1999, the development and evaluation of Split Up¹¹ was reported: an automated legal reasoning system designed to predict outcomes in Australian family law cases involving property proceedings. The central problem it targeted was the pervasive judicial discretion in such areas of law, which poses a major challenge to conventional AI paradigms.

More recently, AI has undergone a dramatic shift, both generally and in legal applications, from symbolic techniques to machine learning approaches, particularly those that work on natural language texts instead of feature sets.¹²

A 2020 paper by Medvedeva and colleagues investigated how natural language processing (NLP) and machine learning can be used to predict judicial decisions, using data from the European Court of Human Rights (ECtHR).¹³ Using NLP tools, the study achieved an average accuracy of 75% in predicting violations of nine articles of the European Convention on Human Rights. When decisions in later cases were predicted from earlier ones, accuracy fell to between 58% and 68%, and it dropped further as the gap between the training and testing periods widened: for the three articles tested in this way, average accuracy was 77% on mixed-year data but only 58% with a ten-year gap.

This path reflects the historical development of legal AI from early, transparent and argumentative systems to today's data-driven predictive models, and it opens an inquiry into the trade-offs between explanation and algorithmic capability in the justice system that accompanied that shift.

⁹ Edwina L. Rissland & Kevin D. Ashley, *Hypotheticals as Heuristic Device*, in PROCEEDINGS OF THE FIFTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE (AAAI-86) 289, 289 (1986), <https://cdn.aaai.org/AAAI/1986/AAAI86-048.pdf>.

¹⁰ Edwina L. Rissland & David B. Skalak, *CABARET: Rule Interpretation in a Hybrid Architecture*, 34 INT'L J. MAN-MACHINE STUD. 839 (1991).

¹¹ Andrew Stranieri et al., *A Hybrid Rule-Neural Approach for the Automation of Legal Reasoning in the Discretionary Domain of Family Law in Australia*, 7 ARTIFICIAL INTELLIGENCE & L. 153 (1999).

¹² Serena Villata et al., *Thirty Years of Artificial Intelligence and Law: The Third Decade*, 30 ARTIFICIAL INTELLIGENCE & L. 561 (2022), <https://doi.org/10.1007/s10506-022-09327-6>.

¹³ Masha Medvedeva, Michel Vols & Martijn Wieling, *Using Machine Learning to Predict Decisions of the European Court of Human Rights*, 28 ARTIFICIAL INTELLIGENCE & L. 237, 237, 258 (2020).

B. Ethical challenges of AI in the legal system

The application of AI brings efficiency and accessibility to the judiciary but poses grave ethical, and therefore legal, challenges that erode core legal principles. This balance must be analysed closely so that technology can advance without compromising justice. Contemporary AI, often based on machine learning algorithms, can reach decisions in ways that resemble human intuition. Such systems no longer follow pre-written instructions but generate solutions from patterns in data that are invisible to human perception.

Herein lies a paradox. It can be impossible, even for its creators, to understand how an AI that has internalised vast quantities of data reaches its decisions. This is known as the ‘black box’ character of AI systems.¹⁴ AI systems built on deep neural networks can be as impenetrable as the human mind, and there is no easy way to understand how they arrive at their outputs.

Although the ‘black box’ character of deep learning systems defeats the doctrines of intention and causation, among others, advances in Explainable Artificial Intelligence (XAI) offer a partial counterweight. Techniques such as LIME and SHAP provide local interpretability by approximating complex predictions in human-readable terms.¹⁵ Broader taxonomies place these alongside surrogate models and rule-based or hybrid methods tailored to legal settings. Proxy explanation techniques, including decision-tree surrogates, provide global interpretability by simulating black-box behaviour in simpler logical forms.¹⁶ Nevertheless, an analysis of the explanation requirements of European law, including fiduciary duties, the GDPR and product safety and liability law, concludes that the current state of XAI ‘is not fully equipped to fulfill all legal requirements’, particularly as regards the correctness of explanations and confidence estimates.¹⁷

This poses a grave danger to two foundational legal doctrines, intention and causation, which traditionally rely on evidence of human conduct, foresight, planning and knowledge. Unlike humans, who leave evidence of their reasoning, AI cannot explain itself, and its outputs may rest on patterns in biased data that lie beyond human perception.¹⁸ As a result, even those who

¹⁴ Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889 (2018).

¹⁵ Ahmed M. Salih et al., *A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME*, 7 ADVANCED INTELLIGENT SYS. art. 2400304 (2025), <https://doi.org/10.1002/aisy.202400304>.

¹⁶ Karen McGregor Richmond et al., *Explainable AI and Law: An Evidential Survey*, 3 DIGITAL SOC’Y art. 1 (2024), <https://doi.org/10.1007/s44206-023-00081-z>.

¹⁷ Benjamin Fresz et al., *How Should AI Decisions Be Explained? Requirements for Explanations from the Perspective of European Law*, 7 PROC. AAAI/ACM CONF. ON AI, ETHICS, & SOC’Y 438, 447 (2024), <https://doi.org/10.1609/aies.v7i1.31648>.

¹⁸ Sandy Steel, *Legal Causation and AI*, in THE CAMBRIDGE HANDBOOK OF PRIVATE LAW AND ARTIFICIAL INTELLIGENCE 189 (Ernest Lim & Phillip Morgan eds., 2024), <https://doi.org/10.1017/9781108980197.010>.

created or deployed such systems may be unable to anticipate, or meaningfully account for, the outcomes they generate.

This has profound implications, of which ‘algorithmic bias’ is perhaps the most discussed ethical issue in the integration of AI into the legal system. Despite claims that algorithmic techniques such as data mining eliminate human bias, they can inherit the prejudices of prior decision-makers, reflect widespread societal biases, or uncover existing patterns of exclusion and inequality.¹⁹ The resulting discrimination is almost always an unintended, emergent by-product of the algorithm’s use, which makes its origin very difficult to trace or to explain to a court. The goal must be to ensure that AI serves as a tool for justice rather than as a hidden engine of discrimination or unreasoned outcomes.

C. Public trust and perception of AI in the legal system

The incorporation of AI into the judicial system also poses particular challenges of public acceptance and institutional legitimacy. Public opinion and trust become overriding considerations that ultimately determine whether AI technologies can be used effectively in the judicial process.

A 2025 study found that the public consistently preferred judges who relied solely on professional expertise over those who used AI, whether exclusively or in combination with human judgment.²⁰ This held for both bail and sentencing decisions, though AI use was viewed less favourably in sentencing. This suggests that members of the public see the use of AI as diminishing human judgment in more serious decisions such as sentencing. Its authors note, however, that their online convenience sample of United States residents limits the generalisability of the findings.

Similarly, a 2024 study found that a significant majority of judges (64%) did not believe that AI holds promise for removing bias from bail and sentencing decisions.²¹ Their written comments largely revealed a negative attitude towards AI in these procedures. This points to the widespread phenomenon of ‘algorithm aversion’, in which judges were reluctant to use, or resistant to, algorithms even where the high accuracy of AI suggested that it might outperform human expertise in uncertain forecasting tasks.

¹⁹ Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671 (2016).

²⁰ Anna Fine, Emily R. Berthelot & Shawn Marsh, *Public Perceptions of Judges’ Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice*, 15 BEHAV. SCI. 476 (2025), <https://doi.org/10.3390/bs15040476>.

²¹ Anna Fine, Stephanie Le & Monica K. Miller, *Content Analysis of Judges’ Sentiments Toward Artificial Intelligence Risk Assessment Tools*, 18 RUSS. J. ECON. & L. 246 (2024), <https://doi.org/10.21202/2782-2923.2024.1.246-263>.

Another study, published in 2025, however, revealed a ‘complex relationship’ between stated trust and behavioural trust among law enforcement and legal professionals.²² Despite their collective scepticism about algorithms, these practitioners were open to taking advice from AI under certain conditions. One respondent said, for instance, ‘I didn’t trust the algorithm’s advice because I have no idea what it’s calculating’, and another that ‘Algorithms aren’t super reliable for human behaviors’. Yet in the experiments they still gave significant, sometimes complete, weight to the algorithm’s recommendations in their final decisions. Conversely, some who said they trusted algorithms gave them less weight in practice.

Indian surveys provide a much-needed local perspective on the uptake of AI in a rapidly digitising legal system. A 2024 Manupatra survey found a strong willingness to experiment with AI: 54.5% of respondents reported that they, their organisations or their teams were currently using AI tools, and 76.7% expected generative AI to reduce the time spent on mundane but necessary tasks.²³ Manupatra’s 2025 survey of 227 respondents found, however, that only 4.1% fully trusted AI output without human verification. The chief impediments were unreliable output quality (58.1%), hallucinated or incorrect content (51.2%) and a lack of India-specific legal context (42.4%). These findings point to a persistent trust gap, driven by the accuracy and contextual limitations of AI systems when applied to the Indian legal system.

While the productivity advantages of AI are evident, concerns about reliability, the local legal framework, data security and regulatory constraints must be addressed before it can be more widely accepted within the Indian legal system.

D. Literature gaps

Although the existing literature provides in-depth analyses of the development of AI in legal reasoning, of ethical issues such as algorithmic opacity and bias, and of the perceptions of civil society and legal practitioners about the use of AI in adjudication, significant gaps persist, including the following:

- **Longitudinal impact studies:** A notable gap in the literature on AI in judicial systems is the lack of longitudinal evaluations, which are indispensable for understanding the long-term effects of AI on legal outcomes, institutional legitimacy and foundational principles of justice. Most current research consists of cross-sectional studies that

²² Ryan Kennedy, Lydia Tiede, Amanda Austin & Kenzy Ismael, *Law Enforcement and Legal Professionals’ Trust in Algorithms*, 2 J.L. & EMPIRICAL ANALYSIS 77 (2025), <https://doi.org/10.1177/2755323x251325594>.

²³ MANUPATRA ACADEMY, *supra* note 4 (54.5% currently using AI tools; 76.7% expecting generative AI to reduce time spent on mundane but necessary tasks); MANUPATRA, *supra* note 4 (2025 survey: 4.1%; 58.1%, 51.2% and 42.4%).

reveal neither how the impact of AI develops over time, nor its cumulative effects, nor gradual changes in principles of justice such as due process and equality.

- **Comprehensive theoretical frameworks for autonomous AI in the judicial context:** The implications of highly autonomous or hybrid forms of AI in judicial decision-making have not yet been examined through any comprehensive conceptual model. Although the role of AI in legal settings is extensively debated, research lags behind the conceptual models that the regulation of highly autonomous systems would require. Questions of constitutional legitimacy and public trust arise alongside.
- **Addressing algorithmic bias:** Current analyses tend to overlook the way AI models mirror and amplify societal biases in training data, which creates an urgent need to neutralise biased algorithmic outputs. There is a marked lack of actionable approaches to detecting, measuring and mitigating bias in judicial AI systems. The literature lacks rich, empirically tested methodologies that address bias in legal data, such as bias in historical records of policing, charging and sentencing, or that suggest practical interventions on which legal professionals can act.

These gaps mark areas that require urgent attention if sophisticated AI is to be integrated into the legal system effectively and responsibly.

III. CALL FOR ACTION

Judicial systems across the globe are hampered not only by administrative inefficiency but by a fundamental failure to resolve disputes in time. This failure undermines the very concept of justice. The ‘cost of inaction’²⁴ is the product of this crisis, visible in institutional dysfunction, poor allocation of resources and persistent limits on human capacity.

Judicial systems worldwide suffer from grave structural problems that delay the resolution of disputes, and India exemplifies this crisis. By the Government’s own measure, India had about 21 judges per million people at the end of 2021 (counting sanctioned posts against the 2011 Census), far short of the 50 per million that the Law Commission recommended in 1987, when it recorded 107 per million in the United States.²⁵ This scarcity of judges has had a severe

²⁴ Vanshika Sirohi, *Judicial Delays: A Global Challenge in Delivering Justice* (SSRN, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5079448.

²⁵ PRESS INFORMATION BUREAU, MINISTRY OF LAW & JUSTICE, *Inadequate Fast Track Courts and Vacancies in Courts* (Feb. 10, 2022), <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1797201> (judge-population ratio of 21.03 per million ‘as on 31.12.2021’); LAW COMMISSION OF INDIA, ONE HUNDRED TWENTIETH REPORT ON MANPOWER PLANNING IN JUDICIARY: A BLUEPRINT 3 (1987).

consequence: at the end of 2022, 434,302 of the 573,220 people held in India's prisons, or 75.8%, were undertrials, many of them detained for long periods.²⁶

Prolonged pre-trial detention undermines the presumption of innocence, and court backlogs carry heavy financial costs, estimated in a study of United States state courts at millions of dollars for each jurisdiction.²⁷ Proceedings are further stalled by procedural inefficiency, such as repeated adjournments, which exhausts litigants and erodes public confidence; an investigation in England and Wales, for instance, documented serious criminal trials adjourned for months because of staffing and budget cuts.²⁸ Beyond a lack of funding, political interference, budgetary pressure and public pressure campaigns can all contribute to the erosion of judicial independence.²⁹ Antiquated legal frameworks and inadequate technology aggravate these shortcomings, producing systems that prioritise procedural compliance over substantive justice.

The Indian judiciary has consistently recognised the gravity of mounting pendency. In 2012, the Supreme Court's judgment in *Imtiyaz Ahmad v. State of U.P.* recorded that 'unduly long delay has the effect of bringing about blatant violation of the rule of law and adverse impact on the common man's access to justice', that access to justice is a fundamental right under Article 21, and that its denial 'undermines public confidence in the justice delivery system'.³⁰ More recently, in *Yashpal Jain v. Sushila Devi* (2023), the Court warned that the 'litigant public may become disillusioned with judicial processes due to inordinate delay', noting that some cases instituted more than fifty years earlier were still pending.³¹ These pronouncements show institutional acknowledgement that chronic delay erodes public confidence and imperils the constitutional order, so that greater judicial efficiency is a necessity.

AI offers significant potential to reduce judicial delay through faster legal research, document processing, case management and improved access to legal information. Existing Indian and foreign initiatives show that AI can reduce administrative burdens and help judges handle large volumes of material. These benefits do not, however, resolve the separate questions of accuracy,

²⁶ Krishnan & Kumar, *supra* note 5; NATIONAL CRIME RECORDS BUREAU, *supra* note 5, at xi–xii.

²⁷ THOMSON REUTERS, THE COST OF DELAYED JUSTICE: A META-ANALYTIC DRIVEN STUDY INTO THE PROBLEMS AND SOLUTIONS SURROUNDING COURT BACKLOGS 5, 26 (2022), <https://legal.thomsonreuters.com/content/dam/ewp-m/documents/legal/en/pdf/reports/the-cost-of-delayed-justice-report.pdf>.

²⁸ Maeve McClenaghan, *Justice Delayed: Administrative Problems Delay Serious Trials*, BUREAU OF INVESTIGATIVE JOURNALISM (July 31, 2013), <https://www.thebureauinvestigates.com/stories/2013-07-31/justice-delayed-administrative-problems-delay-serious-trials> (Crown Courts in England and Wales).

²⁹ Ronald M. George, *Challenges Facing an Independent Judiciary*, 80 N.Y.U. L. REV. 1345 (2005).

³⁰ *Imtiyaz Ahmad v. State of U.P.*, (2012) 2 SCC 688 (India).

³¹ *Yashpal Jain v. Sushila Devi*, 2023 INSC 948, ¶¶ 1, 18 (India), https://api.sci.gov.in/supremecourt/2020/4113/4113_2020_8_1506_47917_Judgement_20-Oct-2023.pdf; see also *Supreme Court's Concerns Over Pendency of Cases: Says May Affect Public Confidence*, INDIA TODAY (Oct. 21, 2023).

bias, explainability and constitutional accountability. The analysis that follows therefore examines how AI is currently being deployed and the safeguards required before its role in adjudication can be expanded.

IV. CURRENT LANDSCAPE

A. Global adoption trends

Between September and December 2023, UNESCO surveyed judicial operators in 96 countries on their use of AI systems.³² The survey sought to understand how judicial operators use AI tools in work-related activities, the extent of that use in legal tasks, and the risks they associate with it. It found that 93% of judicial operators were familiar with AI to some degree: 31% considered themselves experts or very familiar, 41% moderately familiar and 20% slightly familiar, while only 7% knew nothing about the topic. Some 69% acknowledged potential negative issues in using AI chatbots for legal work. The key concerns were:

- Output quality (27%), particularly inaccuracy, falsehood and unreliability of information.
- Privacy, personal data protection and data security (18%).
- Copyright infringement, integrity and originality of output (17%).
- Lack of transparency (14%) about the data used for training, development and operation.

B. Contemporary AI implementations: global case studies

The deployment of AI in legal systems worldwide offers a range of examples that show both the disruptive potential and the limitations of these technologies. They offer insight into the interplay between technological capability, human choice and societal notions of fairness and legitimacy.

i. COMPAS: risk assessment in criminal justice

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system, developed by Northpointe (now Equivant), is one of the most prominent and controversial AI applications in criminal justice.³³ This actuarial risk assessment tool examines more than 130 variables, including criminal history, employment status, history of substance abuse and

³² Juan David Gutiérrez, UNESCO GLOBAL JUDGES' INITIATIVE: SURVEY ON THE USE OF AI SYSTEMS BY JUDICIAL OPERATORS 6–7, 9 (UNESCO 2024).

³³ Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES eao5580 (2018), <https://doi.org/10.1126/sciadv.aao5580>.

psychological evaluations, to produce risk scores that inform judicial decisions on pre-trial release, sentencing and parole.

A 2016 ProPublica investigation of more than seven thousand people arrested in Broward County, Florida, revealed significant racial disparities in the system's performance.³⁴ Among defendants who did not reoffend, Black defendants were nearly twice as likely as white defendants to be labelled higher risk (45% against 23%). Conversely, among defendants who did reoffend within two years, white defendants were nearly twice as likely as Black defendants to be labelled low risk (48% against 28%).³⁵

Flores, Bechtel and Lowenkamp later challenged ProPublica's methodology, arguing that it had tested on pretrial defendants an instrument whose recidivism scales were developed for people on post-disposition supervision.³⁶ Their re-analysis found that Black defendants had higher general (52% against 39%) and violent (21% against 12%) recidivism base rates than white defendants, and concluded that COMPAS predicted with similar accuracy for white (AUC 0.69) and Black (AUC 0.70) defendants, with an overall AUC of 0.71, so that a given COMPAS score translates into a roughly equivalent likelihood of recidivism regardless of race.

In *State v. Loomis*,³⁷ decided by the Wisconsin Supreme Court in 2016, Eric Loomis's high COMPAS risk score had informed his sentence.³⁸ The Court upheld the use of the tool at sentencing but imposed significant restrictions: a risk score may not be used to determine whether an offender is incarcerated or the severity of the sentence, it may not be the determinative factor, and a presentence report that contains a COMPAS assessment must carry written warnings about its limitations. The decision reflects concerns about algorithmic opacity and the preservation of judicial discretion.

ii. SUPACE: India's hybrid approach

The Supreme Court Portal for Assistance in Court Efficiency (SUPACE), launched on 6 April 2021, is a pioneering effort in the digitisation and modernisation of the Indian judiciary. As a flagship initiative of the Supreme Court's Artificial Intelligence Committee, set up in 2019 to

³⁴ Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, in ETHICS OF DATA AND ANALYTICS 254 (Kirsten Martin ed., 2022).

³⁵ Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

³⁶ Anthony W. Flores, Kristin Bechtel & Christopher T. Lowenkamp, *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias"*, 80 FED. PROB. 38 (2016), <https://www.uscourts.gov/federal-probation-journal/2016/09/false-positives-false-negatives-and-false-analyses-rejoinder>.

³⁷ *Loomis*, 881 N.W.2d 749.

³⁸ Note, *State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*, 130 HARV. L. REV. 1530 (2017).

explore the possibilities of AI in judicial proceedings, SUPACE reflects a strategic vision of judicial efficiency that keeps the core principle of human decision-making in adjudication intact.³⁹

SUPACE is designed as a multi-dimensional, AI-driven platform to reduce judicial workload and streamline case management. Its stated aim is to help judges grasp the factual matrix of a case through an intelligent search of the precedents.⁴⁰ Crucially, the platform is conceived as an assistive tool rather than a decision-making authority, so that human judgment is retained in every judicial determination.

SUPACE is a hybrid AI strategy that responds to public demand for human judgment in high-stakes cases while using technology to improve efficiency and effectiveness. Its deployment nonetheless remains limited: the Government informed Parliament in March 2025 that SUPACE was still in an experimental stage of development and that no AI tools were being used by the Supreme Court in its decision-making process.⁴¹ At its launch in April 2021, the Supreme Court's Artificial Intelligence Committee had resolved to put SUPACE to use on an experimental basis, initially with judges dealing with criminal matters in the Bombay and Delhi High Courts.⁴²

Alongside it, the Supreme Court Vidhik Anuvaad Software (SUVAS), a machine-assisted translation tool trained by AI, and the AI-assisted translation work that followed had rendered 36,271 Supreme Court judgments into Hindi and 17,142 into 16 other regional languages by 5 August 2024.⁴³ Phase III of the eCourts Project, approved by the Union Cabinet on 13

³⁹ *Artificial Intelligence Committee, Supreme Court of India Launches SUPACE*, BAR & BENCH (Apr. 6, 2021), <https://www.barandbench.com/news/artificial-intelligence-committee-supreme-court-of-india-launches-supace-live-updates> (reporting Justice L. Nageswara Rao's account of 'the Supreme Court constituting the Artificial Intelligence committee in 2019'); PRESS INFORMATION BUREAU, MINISTRY OF LAW & JUSTICE, *Use of AI in Supreme Court Case Management* (Mar. 20, 2025), <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2113224>.

⁴⁰ PRESS INFORMATION BUREAU, MINISTRY OF LAW & JUSTICE, *supra* note 39 (SUPACE 'aimed at developing a module to understand the factual matrix of cases with an intelligent search of the precedents apart from identifying the cases').

⁴¹ PRESS INFORMATION BUREAU, MINISTRY OF LAW & JUSTICE, *supra* note 39 (SUPACE "is in an experimental stage of development for its testing"; "no AI and ML based tools are being used by the Supreme Court of India in decision making process").

⁴² *Artificial Intelligence Committee, Supreme Court of India Launches SUPACE*, *supra* note 39 (reporting Justice L. Nageswara Rao's statement that the Committee 'has resolved to put SUPACE tool to use on an experimental basis initially with the judges dealing with criminal matters in Bombay and Delhi High Courts').

⁴³ Khara Jyoti Khimjibhai & Sarvari Joshi, *Artificial Intelligence and Judicial Efficiency: Accelerating Trial Processes in Indian Courts*, 11 INT'L EDUC. & RSCH. J., no. 3 (2025), <https://doi.org/10.5281/zenodo.15583828>; MINISTRY OF LAW & JUSTICE, Lok Sabha Unstarred Question No. 3197, *Artificial Intelligence in Judiciary* (answered Aug. 9, 2024) (as on 5 August 2024, 36,271 Supreme Court judgments translated into Hindi and 17,142 into 16 other regional languages); PRESS INFORMATION BUREAU, MINISTRY OF LAW & JUSTICE, *Action Plan for Simple, Accessible, Affordable and Speedy Justice* (Aug. 10, 2023), <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1947490> (describing SUVAS).

September 2023 with an outlay of ₹7,210 crore over four years, earmarks ₹53.57 crore for ‘future technological advancements’, an amount the Government has since described as earmarked for the integration of AI and blockchain technologies across High Courts, and envisages the use of AI, machine learning, optical character recognition and natural language processing in court processes, demonstrating India’s commitment to comprehensive judicial modernisation.⁴⁴

iii. Chinese internet courts: large-scale AI integration

China’s specialised internet courts in Hangzhou (August 2017), Beijing and Guangzhou (both September 2018) represent one of the world’s most extensive AI-supported judicial systems.⁴⁵ As part of the ‘smart court’ reforms officially launched in 2016, these courts handle internet-related disputes, such as those over online loans, e-commerce contracts and product liability, domain names and online copyright, and make extensive use of AI-based technology.⁴⁶

Technology is embedded across the litigation process: cases are registered online, facial recognition confirms litigants’ identities, blockchain preserves evidence, and machine learning generates judgments for judges.⁴⁷ According to the Supreme People’s Court’s 2019 white paper, the three internet courts had accepted close to 120,000 cases by 31 October 2019 and took an average of 38 days to conclude a case, roughly half the time previously taken.⁴⁸

iv. European implementations: ethics-driven approaches

European countries have adopted prudent, ethics-led approaches to the use of AI in the judiciary that emphasise human control, transparency and fundamental rights. One notable development is the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and

⁴⁴ Mohit Sharma, *India’s Courts and Artificial Intelligence: A Future Outlook*, 15 LEXONOMICA 99 (2023), <https://doi.org/10.18690/lexonomica.15.1.99-120.2023>; PRESS INFORMATION BUREAU, CABINET, *Cabinet Approves eCourts Phase III for 4 Years* (Sept. 13, 2023), <https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=1956919>; PRESS INFORMATION BUREAU, MINISTRY OF LAW & JUSTICE, *Digital Transformation of Justice: Integrating AI in India’s Judiciary and Law Enforcement* (Feb. 25, 2025), <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2106239> (‘₹53.57 Crore is specifically earmarked for the integration of AI and Blockchain technologies across High Courts in India’).

⁴⁵ Meirong Guo, *Internet Court’s Challenges and Future in China*, 40 COMPUTER L. & SEC. REV. 105522 (2021), <https://doi.org/10.1016/j.clsr.2020.105522>.

⁴⁶ Shi, Sourdin & Li, *supra* note 2 (the ‘smart court’ concept was ‘first officially raised’ in a 2016 Supreme People’s Court Work Report; quoting Richard Susskind on disputes over ‘online loans, e-commerce (contractual and product liability issues), domain name disputes, and online copyright issues’).

⁴⁷ Shi, Sourdin & Li, *supra* note 2; Tara Vasdani, *Robot Justice: China’s Use of Internet Courts*, LAW360 CANADA (Feb. 5, 2020), <https://www.lexisnexis.ca/en-ca/ihc/2020-02/robot-justice-chinas-use-of-internet-courts.page>.

⁴⁸ SUPREME PEOPLE’S COURT OF CHINA, CHINESE COURTS AND INTERNET JUDICIARY (2019); *China’s Online Courts Reduce Case Handling Time by Half: White Paper*, XINHUA (Dec. 5, 2019), http://www.xinhuanet.com/english/2019-12/05/c_138605947.htm.

Their Environment,⁴⁹ adopted by the Council of Europe's European Commission for the Efficiency of Justice (CEPEJ) in December 2018. These values align with the broader ethical concerns addressed in the Ethics Guidelines for Trustworthy AI, published in 2019 by the European Commission's High-Level Expert Group on Artificial Intelligence. The Guidelines emphasise human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability.⁵⁰

This approach has been reinforced by the comprehensive regulatory framework of the EU AI Act, which entered into force on 1 August 2024.⁵¹ The Act is the first comprehensive legal framework on AI and takes a risk-based approach, classifying AI systems according to their potential effect on fundamental rights and safety. Its provisions on high-risk AI systems apply directly in judicial contexts: AI systems intended to assist a judicial authority in researching and interpreting facts and the law, and in applying the law to a concrete set of facts, are classified as high-risk.⁵²

This diversity of approaches brings to light the central problem of harnessing the capabilities of technology without compromising the essential features of a functioning judiciary.

V. DOCTRINAL BASELINES FOR JUDICIAL AI IN INDIA

The integration of any AI system into the judiciary should be viewed not as a policy innovation but as a natural extension of settled principles of Indian constitutional and administrative law, so that the use of AI in adjudication strengthens judicial legitimacy rather than undermining it.

The incorporation of AI into India's legal system must be rooted in constitutional principle so that efficiency is not achieved at the cost of legitimacy. At their core, such frameworks must align with Article 21 of the Constitution, which, as the Supreme Court held in *Maneka Gandhi v. Union of India*,⁵³ requires every procedure affecting life and liberty to be 'fair, just and reasonable'. The provision has been broadly interpreted to include the right to a speedy and fair trial and the efficient administration of justice. Central to this fairness is the duty to give reasoned orders: the judiciary is accountable to litigants and owes them reasoned decisions. In

⁴⁹ EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (2018), <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

⁵⁰ HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Ethics Guidelines for Trustworthy AI* (Apr. 8, 2019), <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

⁵¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), 2024 O.J. (L 1689) art. 113.

⁵² Regulation (EU) 2024/1689, *supra* note 51, art. 6(2), annex III, ¶ 8(a).

⁵³ *Maneka Gandhi v. Union of India*, (1978) 1 SCC 248 : AIR 1978 SC 597 (India).

S.N. Mukherjee v. Union of India,⁵⁴ the Court held that the recording of reasons excludes arbitrariness and makes effective judicial review possible, and in *Kranti Associates Pvt. Ltd. v. Masood Ahmed Khan*⁵⁵ it reaffirmed the duty and described reason-giving as a requirement of judicial accountability and transparency.

This requirement supports a practice in which every AI recommendation must be explainable and traceable to its source data, and judges must give reasons whenever they adopt or depart from an algorithmic recommendation. AI assistance thereby becomes consistent with the requirement of ‘speaking orders’, a protection that the Vidhi Centre for Legal Policy has applied to judicial AI, observing that opaque ‘black box’ systems stand ‘in direct contravention to the judicial norm of reasoned orders’.⁵⁶

Equally important are the separation of powers and the bar on unlawful delegation of judicial power. Constitutional restrictions on the sub-delegation of judicial authority must be strictly observed, in keeping with the maxim *delegatus non potest delegare* (a delegate cannot further delegate). In *In re Delhi Laws Act*, the Supreme Court held that the legislature may delegate ancillary and subordinate functions but cannot abdicate its essential legislative function;⁵⁷ by analogy, this paper argues, the core adjudicatory function cannot be handed over to a third party. An AI design aligned with ANCHOR avoids this pitfall by casting AI as an aide, not a surrogate judge: its outputs are never binding, are always open to override by a judge, and serve only to assist, not to replace, human judgment.

Algorithmic systems trained on biased or unrepresentative data may produce different outcomes for different social groups, raising concerns under Article 14. Conventional AI systems are often ‘black boxes’; what is needed is an AI framework that confronts this opacity directly. In *S.P. Gupta v. Union of India*,⁵⁸ the Court recognised the right to know and the principle of open government, and rejected a blanket claim of secrecy over the records relevant to the case before it, underlining that access to information is essential to effective judicial scrutiny.

Article 14 of the Constitution imposes a further constraint on the design and deployment of AI in the judiciary. The Supreme Court held in *E.P. Royappa v. State of Tamil Nadu*⁵⁹ that

⁵⁴ *S.N. Mukherjee v. Union of India*, (1990) 4 SCC 594 : AIR 1990 SC 1984 (India).

⁵⁵ *Kranti Assocs. Pvt. Ltd. v. Masood Ahmed Khan*, (2010) 9 SCC 496 (India).

⁵⁶ VIDHI CTR. FOR LEGAL POL’Y, RESPONSIBLE ARTIFICIAL INTELLIGENCE FOR THE INDIAN JUSTICE SYSTEM: A STRATEGY PAPER 12 (2021), <https://vidhilegalpolicy.in/wp-content/uploads/2021/04/Responsible-AI-in-the-Indian-Justice-System-A-Strategy-Paper.pdf>.

⁵⁷ *In re Delhi Laws Act*, AIR 1951 SC 332 (India).

⁵⁸ *S.P. Gupta v. Union of India*, AIR 1982 SC 149 (India).

⁵⁹ *E.P. Royappa v. State of Tamil Nadu*, (1974) 4 SCC 3 (India).

arbitrariness is antithetical to equality and, later, in *Navtej Singh Johar v. Union of India*,⁶⁰ reiterated that the guarantee of equality evolves but retains dignity and non-discrimination at its heart. To meet this constitutional requirement, judicial AI must incorporate fairness metrics, regular bias audits and retraining procedures to avoid arbitrariness and discriminatory effects. Human oversight is, in the same vein, central to the European regime for high-risk AI, which requires such systems to be designed so that natural persons can effectively oversee them while they are in use.⁶¹

Lastly, AI systems must be designed to facilitate full recourse to constitutional remedies and judicial review under Articles 32 and 226 of the Constitution, which the Supreme Court in *L. Chandra Kumar v. Union of India*⁶² recognised as part of the basic structure of the Constitution.

The integrity of the judiciary in an age of algorithmic governance depends, too, on the capacity for audit and review, which is why disclosure and external audits of judicial AI have been urged.⁶³ By building human accountability, transparency and reviewability into its framework, an ANCHOR-compliant AI system helps turn constitutional remedies from theoretical guarantees into tangible safeguards, fortifying the rule of law in an increasingly digital world.

VI. FINDINGS AND ANALYSIS

A. The explainability paradox

Finding 1: Current AI models tend to prioritise either explainability and transparency or scalability and efficiency; no existing tool strikes an ideal balance between the two.

The analysis suggests that explainability in AI-driven legal systems should not be treated as a solvable technical challenge but as a paradox inherent in advanced AI applied to legal reasoning. This paper traces a historical cycle: early AI (RBR and CBR) offered transparency but struggled with nuance and scale, while modern AI (deep learning) handles complexity but sacrifices explainability. The transition from RBR and CBR to deep learning was a historical movement along the explainability-performance axis, in which gains on one side meant losses on the other, as the contrast between HYPO's explicit reasoning and COMPAS's opaque scores shows. This trade-off is extremely difficult to avoid in legal AI.

⁶⁰ *Navtej Singh Johar v. Union of India*, (2018) 10 SCC 1 (India).

⁶¹ Regulation (EU) 2024/1689, *supra* note 51, art. 14(1).

⁶² *L. Chandra Kumar v. Union of India*, (1997) 3 SCC 261 (India).

⁶³ VIDHI CTR. FOR LEGAL POL'Y, *supra* note 56, at 12 (calling for 'the twin requirements of ensuring disclosure and allowing external audits'); *see also* EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE, *supra* note 49 (principle 4: 'authorise external audits').

The ‘black box’ problem is not merely a technical challenge but a fundamental feature of the computational approach of AI when it is applied in messy, high-risk settings such as adjudication. Optimising AI systems for predictive precision tends to conflict with the legal system’s need for transparent and interpretable reasoning. The opacity of COMPAS lay not only in the algorithm itself but also in the inherent difficulty of explaining complicated statistical correlations in terms the legal system can understand.

Society must therefore decide how far it is prepared to sacrifice explainability and transparency for the sake of efficiency and scalability, since having the best of both worlds does not, with today’s technology, appear possible.

B. AI as a reflection

Finding 2: AI cannot be neutral when its training data reflects systemic bias.

This paper challenges the assumption that AI will be a neutral agent that frees us from human prejudice. The COMPAS system is an example of how AI can replicate, and potentially exacerbate, societal biases embedded in historical data. The debate about definitions of ‘fairness’ is not a mathematical abstraction; it is a serious philosophical and social argument about what ‘justice’ means once it is put into code. In retrospect the conclusion is obvious: if the data used to train an AI system reflect significant societal prejudices, the system is likely to exhibit those prejudices too, perhaps more strongly.

The tensions in algorithmic fairness noted earlier show that ‘fairness’ is a contested social construct rather than a technical parameter, and that human design choices about what makes a system ‘fair’ are subjective and often biased. Expecting an AI system to be completely neutral is therefore unrealistic when its training data reflect historical prejudice.

This analysis indicates that AI is an exacting, if uncomfortable, mirror of deeply embedded social injustice and of the inherent subjectivity of human ideas of ‘fairness’. Achieving ‘unbiased AI’ depends less on perfecting algorithms than on confronting flawed data inputs that carry the inherited prejudices of human institutions. Bias in AI systems is therefore typically the reproduction and magnification of entrenched societal prejudice rather than a purely technical deficiency.

C. Hybrid models as the future

Finding 3: Hybrid models appear more desirable and practical than fully autonomous AI systems.

This paper has already highlighted the significance of hybrid solutions such as SUPACE, arguably the most ethically sound and sustainable model for AI in the judiciary, and of the European Union's emphasis on human control. Both acknowledge the inherent limitations of AI and the enduring importance of human judgment in a rights-based legal system, which shows that the hybrid approach is not a mere pragmatic compromise but a matter of principle.

Full automation of complex judicial processes is, it is argued, undesirable and hostile to fundamental principles of justice, including empathy, moral judgment, contextual sensitivity and human responsibility. These attributes are inextricably bound up with human nature and cannot be adequately replicated by, or responsibly outsourced to, even the most advanced algorithms.

Apart from these ethical incompatibilities, existing autonomous AI systems lack the capacity for meaningful understanding, flexibility and contextual judgment needed in novel legal contexts or in those that require nuanced interpretation beyond pattern matching. Crossing this ethical threshold would mean delegating decisions with profound human consequences, including findings of culpability, judgments of intent and the imposition of sanctions, to systems that lack genuine moral reasoning, experiential knowledge and empathy. Such automation risks the cold application of rules regardless of human suffering, makes moral responsibility impossible without a conscious decision-maker, and creates an accountability vacuum in which no human being is ultimately answerable.

These tensions call for a hybrid, human-centred model, which guides the development of the ANCHOR Guidelines set out below.

VII. THE ANCHOR GUIDELINES

The ANCHOR (Accountability, Non-delegation, Constitutional Conformity, Human Oversight, and Review) Guidelines are a theoretical blueprint intended to enhance the legitimacy of the judiciary while harnessing the benefits of AI. The underlying design principle is 'AI as co-judge, not over the judge'. Holding together the twin imperatives of technological proficiency and judicial legitimacy, the model seeks to bridge the gap that currently exists between human judgment and algorithmic precision. Unlike fully autonomous models that risk undermining due process, ANCHOR aims to preserve human accountability while drawing on the speed and scale

of AI. The model draws inspiration from SUPACE, which this paper regards as the most promising integration of AI into the Indian judiciary to date, and builds on it by widening the scope of the AI system's involvement in cases.

A. Design philosophy and model structure

Under the ANCHOR model, the AI system operates in an assistive capacity, with humans monitoring its performance and available to intervene when necessary. In such a system, the ultimate decision-maker remains the human judge. The AI system would work as an efficient assistant that automates time-consuming tasks, such as:

- extensive legal research;
- identifying conflicting precedents across different jurisdictions;
- summarising lengthy depositions or witness testimony;
- cross-referencing specific legal codes or regulations; and
- identifying patterns in past judgments in similar cases, among others.

In such a framework, the AI system would have no legal authority to deliver binding decisions, since judges are intended to remain the final decision-makers.

The AI system would also be designed to make recommendations based on all the available information; such a recommendation would not bind the judge and would have only persuasive value, to assist the judge where needed. When making a recommendation, the system must provide all relevant information, its supporting research and the reasons for its conclusion, so that the judge can carry out an independent analysis before deciding the matter. Where a judge departs from the AI's recommendation, a concise written rationale must be recorded. This will help fine-tune the AI model and discourage arbitrary exercise of judicial discretion.

It is also proposed that a dedicated team be constituted to oversee the performance and operation of the AI systems. The team would function as a watchdog, setting parameters, auditing outputs and stepping in only when the AI's performance goes awry or an anomaly is detected. It would track the frequency and nature of judicial departures from the AI's recommendations in order to calibrate and continually refine the model. The model would also incorporate the human-rights principles that underpin the European approach to AI and be grounded in constitutional morality, due process and the rule of law.

B. Addressing core challenges

One of the most significant challenges for AI in legal systems is establishing precise accountability mechanisms. To address it, this paper draws on Bathaee,⁶⁴ who proposed that in mission-critical or high-risk settings the user or developer of an AI system should bear broader liability for harm, akin to a strict liability regime, whereas in less risky settings a negligence rule is more suitable.

Building on Bathaee's risk-based scheme, ANCHOR proposes a two-tier standard: a statutory strict liability regime covering developers and deployers for harm caused by AI used in mission-critical or high-risk judicial functions (for example, sentencing guidance or bail recommendations), and a negligence standard for lower-risk contexts (for example, scheduling or research support). This distinction ensures that rights-sensitive functions are subject to the highest duty of care, while the standard of reasonableness still governs less critical modules.

The problem of explainability is difficult to address because of the explainability paradox. To tackle it, a regime should be established under which developers are liable if their AI system fails to meet standards of transparency that reasonable care and caution would require. As Bathaee argues, where the AI is transparent, knowledge of how its decision-making works can show whether reasonable care was taken, whereas where it is a black box, 'the deployment of the AI in the face of a lack of transparency may be sufficient to establish a lack of reasonable care'; causation should then turn 'not on whether the particular harm caused by the AI was reasonably foreseeable, but whether the harm was a foreseeable consequence of deploying black-box AI autonomously'.⁶⁵

In addition, for every piece of AI-recommended material, insight or data point presented to a judge, the system should offer a clear, traceable path back to the source data, precedents and weighted factors from which the output was derived. Judges, lawyers, auditors and even members of the public should be able to ask the system for a human-readable explanation of why a particular piece of information was considered relevant or how a pattern was identified, and so gain a clearer understanding of the AI's contribution to the judicial process.

AI reflects the societal biases present in its training data. Where deep systemic biases exist in AI systems, the problem becomes harder to solve, which makes the judge's oversight critical. If, moreover, the system is trained on a large, global dataset, such as international case law and case studies, rather than on small, localised data, it may approach greater neutrality. This

⁶⁴ Bathaee, *supra* note 14, at 935.

⁶⁵ Bathaee, *supra* note 14, at 935, 938.

reasoning draws on the ‘law of large numbers’,⁶⁶ which has a counterpart in the legal field: a demographic group that is a minority in one jurisdiction may be a majority in another, so that a larger dataset can offset such disparities. The hypothesis has its limits, however, since some groups, women in particular, have faced discrimination across the world. This places a special burden on judges to correct such biases in their rulings, because a larger dataset does not guarantee neutrality; it merely increases the likelihood of eliminating or limiting localised biases.

Under the ANCHOR blueprint, every AI model must undergo rigorous, mandatory testing for impartiality against pre-set standards spanning multiple demographic groups before it is deployed. The blueprint also envisages systematic, automated testing of the efficacy and outputs of AI models to identify newly emerging biases, supported by well-defined protocols for retraining and recalibration whenever bias is detected.

A core principle of the ANCHOR Guidelines is ‘AI as co-judge, not over the judge’, which places human judgment above the machine in complex judicial functions. Human judgment is the foundation on which fundamental rights are protected. The system guarantees human discretion: final decisions, particularly those affecting liberty and rights, always rest with the human judge, who has complete discretion to accept, reject or amend AI-provided inputs.

An ethical review board would oversee the development, deployment and ongoing use of all AI features to safeguard constitutional values and human-rights norms. Such formal evaluation ensures that AI advances justice while protecting fundamental principles such as empathy, accountability and fairness, so that no ‘ethical red line’ is crossed.

An appeal mechanism would also be built in, so that litigants, judges or developers can challenge AI-related decisions, or the findings of the oversight bodies, before the higher courts. The framework would also contain a sunset clause requiring the safeguards to be reviewed and updated periodically, so that they do not become obsolete as AI technology develops. Finally, while the framework is rooted in Indian constitutional jurisprudence, it does not ignore emerging global norms for AI regulation and can be aligned with best practices such as the OECD AI Principles and UNESCO’s principle of human oversight of AI.

This multi-layered design, comprising judicial norms, statutory duties, independent monitoring and periodic review, seeks to embed accountability in the architecture of AI-assisted adjudication. In doing so, it aims to turn liability from an aspirational protection into a legal

⁶⁶ JACOB BERNOULLI, *ARS CONJECTANDI* (Thurneysen Bros. 1713).

reality, so that the role of AI in the courts remains constitutionally sound, transparent and rights-protective.

Drawing on these diverse mechanisms, this article offers an integrated strategy for bringing AI into the judiciary, moving from theoretical strategies to concrete steps that address the fundamental ethical and practical concerns about the use of AI in the justice system. The approach demystifies the successful integration of AI, in contrast to the use of less regulated AI technologies.

C. Doctrinal anchor

The ANCHOR Guidelines are not free-standing design principles; they are built on the constitutional and administrative law principles that define the outer limits of judicial power in India. The requirement of fair, reasoned and non-arbitrary procedure under Article 21 finds expression in ANCHOR's insistence on the explainability and traceability of algorithmic outputs. The equality guarantee of Article 14, developed through the Court's rejection of arbitrariness in cases such as *E.P. Royappa* and *Navtej Singh Johar*, underlies the commitment to fairness audits and protection against discriminatory outcomes that lies at the core of the Constitutional Conformity pillar.

Equally, the non-delegation principle articulated for legislative power in *In re Delhi Laws Act* suggests, by analogy, that adjudicatory authority cannot be abdicated; this informs the Non-delegation pillar, which ensures that AI remains advisory rather than determinative. Finally, the right to remedies under Articles 32 and 226, affirmed as part of the basic structure in *L. Chandra Kumar*, grounds ANCHOR's emphasis on reviewability and audit trails. In this sense, ANCHOR is less an abstract framework than a constitutional translation: a means of giving practical, institutional expression to values already embedded in Indian law.

D. Comparative positioning

Although this paper proposes a human-centred AI model tailored to Indian judicial needs, it must be considered in the context of the wider global and domestic landscape of AI in law. Several existing systems, including COMPAS, SUPACE and XAI toolkits, offer both inspiration and valuable lessons that have informed the architecture of ANCHOR.

The COMPAS system, used in United States courts for recidivism risk assessment, exemplifies the dangers of opaque judicial AI. The ANCHOR Guidelines address these deficiencies through mandatory explainability with human-readable outputs, traceability logs for every AI-generated suggestion, and judicial override mechanisms built in by default. The Supreme Court's SUPACE is designed to assist with summarisation and research, but public details of its

transparency protocols and empirical testing remain limited. AI systems built to the ANCHOR blueprint would improve on SUPACE through a modular, testable design with empirical feedback loops, mandatory oversight boards and full audit trails, features that SUPACE's published design lacks.

Explainable AI toolkits such as LIME and SHAP offer post-hoc explanation techniques but are of limited use in legal contexts, since they produce abstract explanations that do not map onto legal reasoning. To address this, future AI tools under ANCHOR would use decision-tree-style justifications that map directly onto patterns of legal reasoning, such as the elements of a statutory provision and the analysis of precedent, and would link explanation mechanisms to accountability frameworks and rights of appeal. China's internet courts have been reported to use an AI-generated 'virtual judge' to question litigants in pre-trial proceedings,⁶⁷ a development that raises concerns about over-automation and legitimacy. This paper deliberately rejects autonomous adjudication, positioning AI as a co-judge confined to routine tasks and ensuring that every final decision is made and signed by a human judge.

Unlike existing models, systems built to ANCHOR would seek to bridge technological capability and democratic legitimacy. By building explainability, accountability, traceability and strong human override into every level of operation, they would strike a balance between innovation and the constitutional and ethical requirements intrinsic to Indian adjudication.

E. Imperatives and limitations

Deploying AI systems in line with the ANCHOR framework requires three essential prerequisites. First, the systems need access to high-quality, structured judicial datasets, which would call for extensive cooperation with judicial IT cells to overcome the fragmentation of India's legal data infrastructure through pre-processing and standardisation. Second, pilot implementation must include rigorous empirical evaluation that regularly monitors review periods, rates of judicial departure from AI recommendations and measures of litigant trust. Third, deployment must rest on open governance and broad public engagement, presenting AI explicitly as enabling infrastructure grounded in constitutional principles, not as an independent adjudicator.

A number of constraints, however, seriously test the practical feasibility of the framework. Deployment entails high financial and time costs beyond initial development, including regular maintenance, system upgrades and extensive training programmes. The human-capital

⁶⁷ Vasdani, *supra* note 47 (describing a 'virtual judge' at the Hangzhou Internet Court putting questions to the parties 'during a pretrial meeting'); *see also* Shi, Sourdin & Li, *supra* note 2 (concerns over 'the use of automated judgments').

challenges include the need for systematic technological training for judicial staff and possible institutional opposition to the displacement of lower-level positions. Judges may resist what they see as technological intrusion into established processes, which could lead to low adoption or even active resistance to implementation. The hypothetical nature of the framework calls for rigorous empirical testing before any broad adoption is considered.

These constraints underline the experimental character of AI tools and of the ANCHOR blueprint, and the substantial preparatory work needed before they can be incorporated into the Indian legal system.

VIII. CONCLUSION

This article has traced the path of AI in the courtroom, from early rule-based systems to current machine learning systems, and has made the case that human-centred governance is necessary to gain efficiency without undermining legitimacy. The ANCHOR Guidelines consolidate this approach, combining open algorithms with rigorous developer accountability and protection through human override.

Practical constraints nevertheless persist, such as concerns about data quality, institutional inertia and gaps in resources. Ethical questions, such as how to keep AI systems transparent without exposing private or proprietary information, also call for sharper normative guidance, and the dynamic nature of legal norms requires AI systems to be flexible and responsive to evolving statutes and precedents.

To ensure that ANCHOR is introduced in an informed and responsible way, this paper recommends a carefully scheduled pilot of 18 to 24 months in three selected courts: one metropolitan magistrate's court, one district civil court, and one rural, small-claims or family court, each expected to handle approximately 3,000 to 3,500 cases, for a total of around 9,000 to 10,000 cases. Together, these represent the wide variety of India's judicial contexts.

Each site would begin with a baseline period and then phase in the system gradually, allowing analysis over time. Assessment would cover measures of efficiency (such as case resolution times, judge-AI agreement and rates of appeal and reversal) as well as broader governance indicators such as consistency across similar cases, the trust of the parties, judicial workload, and perceptions of fairness among lawyers and litigants. Safeguards such as independent ethical approval, pre-registered outcomes and review triggers would ensure accountability, and structured feedback loops among judges, advocates and court authorities would allow the participating courts to learn and adjust. In that respect, the pilot is not merely a technical experiment but an exercise in governance: a test of whether AI can improve efficiency while

remaining faithful to constitutional commitments to fairness, transparency and judicial legitimacy.

By following this roadmap, it becomes possible to move beyond anecdotal pilots to rigorous evidence of the real-world impact of AI. With sustained collaboration and resources, judiciaries may by 2030 not only clear backlogs but also reinforce public trust through transparent, accountable and human-centred AI systems, affirming the rule of law in the digital age.
