Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 2, 2026

JUSTICE BEHIND THE BLACK BOX: ALGORITHMIC ACCOUNTABILITY AND THE LIMITS OF PAKISTAN'S LEGAL FRAMEWORKS

Amna Hammad[*1], Sakhawat Ali[2], Taj Ali[3], Sunh Abbasi[4]

[*1,3,4]*LLB (Hons.) Batch 2k25, Department of Law, University of Sindh Jamshoro Campus*
[2]*LLB (Hons.) Batch 2k24, Department of Law, University of Sindh Jamshoro Campus*

[*1]annalodhi1123@gmail.com, [2]sakhawat.khk2005123@gmail.com, [3]alilagharitaj@gmail.com, [4]sunhhabbasi@gmail.com

DOI:https://doi.org/10.5281/zenodo.21772795

Keywords
Artificial Intelligence (AI), Algorithmic accountability, Legal Transparency, Judicial Automation

Article History
Received: 21 December 2025
Accepted: 05 February 2026
Published: 20 February 2026

Copyright @Author
Corresponding Author: *
Amna Hammad

Abstract
The adoption of AI in the judiciary of Pakistan has shown numerous efforts to increase the efficiency of the judicial process while retaining the human monopoly on judicial decision-making. However, the increasing use of AI introduces major risks, such as biased programming, black-box reasoning, and technical hallucinations. Existing laws are completely unequipped to handle these issues because there is currently no legal framework that governs how AI is used in judicial matters or daily life. Even existing laws like the Prevention of Electronic Crimes Act (PECA) 2016, which deals with online and electronic crimes, lack specific requirements for algorithmic accountability, creating a regulatory void regarding biased programming and errors caused by using AI. Establishing robust AI accountability is vital for maintaining the moral authority of the court, judicial fairness, and public trust, ensuring that justice remains a product of human conscience rather than opaque machine logic. This study examines Supreme Court Precedents and national policies to see if they are effective in addressing these challenges. To help strengthen legal control, this study proposes a "Judicial AI Accountability Framework" which identifies liabilities, requires transparency regarding the use of AI in courts, and introduces judicial audit trails—unchangeable forensic digital records that document all judge-AI interactions to increase traceability. This helps to turn policy promises into robust legal protections and keep judicial decisions credible, even as automation increases. This study uses doctrinal and qualitative methodology that relies on statutes, precedents, policy documents, news articles, and academic literature. Ultimately, this research contributes to creating effective regulatory mechanisms for ethical and responsible use of AI in the Pakistani Judiciary. This research finds that existing statutes, policy, and precedents provide no operational accountability for AI-assisted adjudication. It recommends codifying the proposed Judicial AI Accountability Framework through a dedicated statute and an independent oversight authority to translate promises into binding legal protections.

1. INTRODUCTION

The judiciary across the globe is currently in the midst of a major shift as Artificial Intelligence becomes a staple of the judicial system. This technological shift is driven by the promise of enhanced efficiency, precision, and cost-effectiveness in legal processes, moving beyond traditional methods (Munir, 2024). Developers are creating AI tools to support every phase of legal work, ranging from basic court

administration to the complex evaluation of evidence, case outcome forecasting, and even assisting with final judgment (Munir, 2024). While these innovations offer faster results, they also bring forward difficult issues such as privacy invasion, bias, and hallucinations in AI outputs, necessitating robust regulatory control (Munir, 2024).

In Pakistan, the legal system is increasingly looking towards AI as a way to solve long-standing problems like systematic delays, massive case backlogs, and limited access to justice (Soomro et al, 2024). Implementing this technology is viewed as a key step in modernising courts and improving the overall functioning of the entire legal system (Qazi, 2023). A major part of this strategy is the proposed Smart Court System (SCS), which is intended to completely change existing methods through digital case management, e-filing capabilities, and automatic generation of court orders to lift the heavy administrative burden on the staff (Soomro et al., 2024). Beyond administrative improvements, AI is set to take a more active role in judicial tasks through AI-assisted workflow and Decision Support Systems built to provide predictive analysis, automate the drafting of legal papers, and ensure resources are used more effectively across different regions, ultimately helping to clear backlogs in both civil and criminal proceedings (Soomro et al., 2024). Additionally, AI-based sentiment analysis is being suggested as a way to enhance fairness by detecting biases present in court transcripts.

The rapid introduction of AI into Pakistan's courts brings to the forefront a critical and complex set of challenges, specifically regarding accountability and the balance between human oversight and automated speed. Pakistan currently lacks a legal and regulatory framework to manage AI, which leaves people and institutions vulnerable to privacy breaches, cyber threats, and even social manipulation (Ahn & Chen, 2020). A primary concern is that AI adjudication could be biased, either because of flaws in the data used to train it or in the design of the algorithms. If this is not addressed, biased AI systems could make existing social inequalities worse and negatively affect marginalised communities, undermining the pursuit of social justice (Javed & Li, 2025). This leads to a fundamental question: Who is responsible when an AI system makes a flawed decision? This debate intensifies when discussing whether AI should actually replace human decision-making, as highlighted by conflicting views on the functions and ethical implications of using these tools in the legal profession (Munir, 2024).

To navigate these issues, this research recommends the creation of "The AI Judicial Accountability Framework". This framework would enforce human-centric oversight by treating AI as a rebuttable witness. It mandates constitutional compliance, ensures developer liability, and prohibits automated sentencing. It also requires forensic digital audit trails to ensure all algorithmic interactions are transparent, traceable, and subject to human oversight.

2. RESEARCH GAP

In Pakistan digitisation has prioritised efficiency over the legal and ethical challenges posed by AI. This has led to a situation where AI systems are being considered or implemented without specific requirements for algorithmic accountability, especially regarding decision-making processes, potential for biased programming, and errors (Farid & Ghazanfar, 2025). The existing legal frameworks, primarily the Prevention of Electronic Crimes Act 2016 and the Draft Data Protection Bill, fall short of directly addressing the challenges of AI accountability in judicial decision-making (Farid & Ghazanfar, 2025). The PECA 2016 provides a general framework for digital technologies but lacks specific provisions for algorithmic decision systems within legal adjudication. Similarly, the Draft Data Protection Bill does not include sector-specific rules for judicial bodies, thereby omitting crucial procedural safeguards for automated legal reasoning that could ensure accountability (Farid & Ghazanfar, 2025). Furthermore, the Information Technology Ordinance 2002 lacks procedures that could be extended or adapted to cover AI-generated insights in legal proceedings (Afzal & Ghawaas, 2025). This lack of a clear legal framework creates uncertainties about the admissibility of AI-generated evidence, the extent of judicial reliance on AI recommendations, and the

procedural rights of litigants in an AI-assisted environment (Farrukh et al., 2024; Gondal et al., 2025).

3. METHODOLOGY

The methodology for this paper follows two main paths to ensure a comprehensive examination of both the legal principles and the practical usage of technology into the courts.

This study uses a doctrinal approach to analyse the existing legal frameworks in Pakistan. This includes Article 10-A of the Constitution of Pakistan 1973, the Prevention of Electronic Crimes Act 2016, and the Draft Data Protection Bill. The central component of this study is the critical analysis of National Artificial Intelligence (AI) Policy 2025 and the landmark Supreme Court precedent "Ishfaq Ahmed vs Mushtaq Ahmed". It identifies whether these statutes, precedents, and policy are sufficient to address the accountability vaccum in the existing legal system.

This research also involves a qualitative assessment of our institutional readiness for digital change. It examines how AI is being used in practice, such as the use of Judge-GPT in the District Courts of Punjab. By reviewing academic papers and policy documents, the study explores the practical challenges of AI, including hidden biases, the lack of transparency in how algorithms make decisions, and who is liable for such outputs. Finally, the paper suggests a better framework for the future.

4. LITERATURE REVIEW

Pakistan's legal landscape contains certain statutory principles that demand accountability in AI integration, such as the right to a fair trial under Article 10-A and the protection of privacy, which requires that technological interventions influencing judicial outcomes be transparent and subject to human oversight (Farid & Ghazanfar, 2025). Pakistan's alignment with international guidelines such as UNESCO Recommendations on the Ethics of Artificial Intelligence underscore this commitment (Farid & Ghazanfar, 2025).

A critical issue associated with the usage of AI is the accountability vacuum when AI tools influence crucial judicial decisions, such as sentencing, bail, or case predictions. The absence of clear guidelines on who is responsible for errors, biases, or unfair outcomes becomes a major concern (Farid & Ghazanfar, 2025) (Mangi et al., 2025). Pakistan currently lacks specific court rules or statutes that define the permissible scope of using AI in judicial support. This regulatory void creates ambiguity regarding liability when AI contributes to an adverse legal outcome. The Law and Justice Commission of Pakistan has yet to issue clear policy or guidance addressing AI accountability in courtrooms, leading to regulatory inertia (Farid & Ghazanfar, 2025). Furthermore, infrastructural irregularities present a formidable barrier. While some superior courts possess advanced systems, many lower courts lack even basic digital infrastructure and rely on traditional paper-based systems (Aziz & Abid, 2024; Farid & Ghazanfar, 2025). This uneven landscape poses serious challenges, as legal professionals express concerns about job displacement, skill devaluation, and a lack of understanding regarding AI functionalities exacerbating existing access-to-justice divides. Moreover, courts often face financial limitations and shortage of personnel with adequate technological expertise, which hinder the effective adoption and maintenance of AI systems (Aziz & Abid, 2024). There is an urgent need for specialised training programmes for judicial officers, lawyers, and support staff to effectively interact with and oversee AI tools (Afzal & Ghawaas, 2025).

The conservative nature of Pakistani courts poses a constitutional hurdle for the deployment of fully autonomous AI judges, suggesting that AI-assisted tools that support, rather than replace human adjudication, are the constitutionally compliant path forward (Farrukh et al., 2024). Even though accountability requires that human judges retain ultimate decision-making authority and that AI recommendations are not merely rubber-stamped, the protocols ensuring that judges critically evaluate and override AI inputs do not exist in Pakistan (Farid & Ghazanfar, 2025).

Finally, the absence of a comprehensive national strategy for AI in justice means that policy commitments fail to translate into

tangible legal safeguards (Farid & Ghazanfar, 2025). The need for digital audit trails—which refers to chronological, automated records that track and document every step, data input, and decision-making process within a software system—to increase transparency and traceability of AI's role in judicial processes is also a significant unaddressed area (Farid & Ghazanfar, 2025). Also, a proactive regulatory stance is essential to ensure that AI serves to strengthen the integrity and fairness of the justice system (Afzal & Ghawaas, 2025; Qureshi et al., 2025).

5. CRITICAL EVALUATION OF PAKISTAN'S NATIONAL ARTIFICIAL INTELLIGENCE (AI) POLICY 2025

Structural Misalignment Between The Policy And Judicial Accountability

The National Artificial Intelligence Policy 2025 serves as a blueprint for transitioning Pakistan into a knowledge-based economy, created by the Ministry of IT and Telecommunications, is structured around six strategic pillars, yet its treatment of the judiciary remains superficial. While the policy identifies the justice sector as a priority for quicker decision-making and the rule of law, it currently approaches judicial governance through the lens of general digitisation rather than specific algorithmic accountability. This broad, horizontal approach to AI governance is fundamentally incompatible with the unique constitutional and procedural demands of judicial decision-making. The policy groups equal access to justice alongside policing, disaster management, agricultural development under its sector-specific roadmap, creating a one-size-fits-all strategy that overlooks the unique moral and constitutional weight of judicial decisions. Unlike administrative sectors such as irrigation or land record digitisation, judicial decision-making directly implicates human rights and personal liberty, requiring far more than general efficiency. The policy therefore fails to provide the deep, technical standards necessary to interrogate algorithmic bias, generative hallucinations, or biased risk assessments in legal drafting, all of which are critical to maintaining the accountability and integrity of the court (Dahl et al., 2024; Blair-Stanek & Van Durme, 2025).

The critical analysis reveals that the policy favours broad-scale capacity building over the specific, high-risk requirements of judicial AI accountability. Target-driven initiatives such as training 200,000 individuals annually and upskilling 70 per cent of the public sector workforce are foundational for national readiness, yet they treat AI literacy as a uniform skill set rather than addressing the specialised forensic and ethical needs of the judiciary. The most profound regulatory gap lies in the absence of accountability measures for judicial tools themselves. Although the policy's Secure AI Ecosystem pillar mandates transparency and human oversight, these remain abstract ethical objectives rather than precise legal obligations integrated into the judicial process. The policy establishes that human oversight is essential for high-risk scenarios but fails to provide a technical or legal definition of what constitutes acceptable human intervention when a judge utilises tools like Judge-GPT (Mökander, 2023; Verma et al., 2025). Without specific rules describing permissible scope, there is a significant risk of cognitive offloading, where judges may unintentionally rubber-stamp AI-generated risk assessments due to pragmatic pressure and high caseloads (Pruss, 2021).

This gap is compounded by the absence of mandatory reproducibility standards or detailed digital audit trails, mechanisms essential for allowing appellate courts to review the logic behind machine outputs of lower courts. The policy provides no technical bridge to hold developers or judicial officers accountable for specific algorithmic logic failures, and although it references third-party auditing, the system is designed for general user trust rather than the forensic auditing required for tools that impact personal liberty. Consequently, the policy does not grant litigants a right to an explanation for the data points that influence a judicial outcome, and even if legally responsible, judges may be technically incapable of identifying deep-seated data biases or hallucinations, rendering accountability purely symbolic.

The Gap Between Aspirational Commitments And Enforceable Protection

Specific regulatory structure for the judiciary remain in nascent stages, with the AI Directorate only slated for 2027. By treating

judicial tools under the broad umbrella of high-impact systems, the policy ignores the explicit accountability measures required for the justice system. Further, the policy relies on flexible rules and experimental sandboxes that often lack the rigid procedural protections required by Article 10-A of the Constitution regarding the right to a fair trial, and provides no clarity on how litigant rights are protected if subjected to an experimental program during a pilot phase. Policy language also prioritises future intentions over current established legal instruments: measures such as AI-integrated security guides, human oversight processes, and AI lifecycle evaluations are described as aspirations rather than realities. While the policy promises to match international standards like the UNESCO Recommendations on the Ethics of Artificial Intelligence (AI), it fails to translate them into enforceable legal principles that allow judicial auditing or contesting of algorithms in court. The fact that the AI Directorate must still work with legal specialists to create governance frameworks indicates that concrete regulatory infrastructure is largely a future task. Without codification into specific laws, the policy's commitments remain insufficient to safeguard the integrity and moral authority of the Pakistani justice system, underscoring the urgent need for a dedicated Judicial AI Accountability Framework that identifies required liabilities, mandates transparency, introduces digital audit trails, and preserves the legitimacy of judicial decision-making in an automated environment.

6. DOCTRINAL ANALYSIS OF ISHFAQ AHMED V. MUSHTAQ AHMED JUDGMENT (2025 SCP 112)

The landmark judgment in Ishfaq Ahmed v. Mushtaq Ahmed case (2025 SCP 112), delivered by Mr. Justice Syed Mansoor Ali Shah and Mr. Justice Aqeel Ahmed Abbasi, evolved from a standard property dispute into a pivotal discussion on the usage of Artificial Intelligence within the Pakistani judiciary.

The facts are that the petitioner entered into an oral tenancy agreement with his brother, Mushtaq Ahmed, in 2016 for a residential property in Moza Rajghar, Lahore, for a monthly rent payments of Rs 35,000. Following the default in the payment, eviction proceedings were initiated. While the initial petition was dismissed, an appellate forum ruled in petitioner's favour in 2021. This was subsequently overturned by the Lahore High Court (LHC) in 2022 through the constitutional petition, leaving the petitioner to seek justice from the Supreme Court. The Supreme Court observed that this routine matter took several years to conclude despite the summary nature of rent proceedings. The Court noted that such undue delay erodes constitutional rights and raises broader economic and social justice concerns, prompting the judiciary to explore AI as a tool to address these systemic inefficiencies.

Stance On AI-Integration, Transparency, And Judicial Accountability

The Precedent highlights the global adoption of AI and large language models, recognising their ability to uncover insights. AI tools like Westlaw, LexisNexis, Casetext, Bloomberg Law, and Vlex are integrated into judicial workflow to support legal research and drafting. In this context, Judge Amir Munir in Pakistan revealed using ChatGPT to adjudge a case both in civil and criminal matter. The judgment discusses how the Federal Judicial Academy (FJA) in Islamabad has taken a crucial step in judicial innovation by introducing Judge-GPT, developed in collaboration with ETH Zurich, which is presently used by 1,500 judges in the District Judiciary. The judgment highlights that these tools enhance clarity and precision by improving textual quality, detecting errors, and refining structure. They assist in organising complex case files, identifying legal questions, and summarising records, leading to more informed and timely decisions. However, the judgment highlights a significant black-box problem, where the internal decision-making logic of AI remains hidden or is overly complex despite visible inputs and outputs. This transparency gap prevents users from understanding the reasoning as to how and why AI has reached a specific result. The court notes that AI tools used for sentencing, bail, and comparing judicial decisions often perpetuate historical biases found in their training data. The court asserts that judicial integrity is compromised if AI systems are allowed to influence court rulings through the integration

of hidden demographic or ideological biases. Furthermore, the judgment argues that without explainability, judges and advocates cannot justify AI-assisted rulings undermining legal defensibility and moral authority of court's rulings. Ultimately, the absence of explainability weakens legal accountability, as outcomes stripped of meaningful justification erode transparency and public confidence in the judiciary.

The judgment also emphasises that accountability is the backbone of judicial process, particularly when AI influences legal decisions. It mandates that human authorities maintain final responsibility, ensuring systems are in place to challenge or correct any questionable AI-generated outcome. Judges must retain exclusive accountability, utilising AI strictly as a supportive resource rather than a conclusive force in the administration of justice. It establishes that AI may assist court by calculating the risks or drafting legal documents, the presiding judge is responsible for the final authentication and carries full legal liability for the ruling. This framework underscores the vital role of human oversight, requiring judges to scrutinise and, if necessary, reject AI recommendations to uphold the dignity of the court. It ensures that while developers may be liable for technical flaws or algorithmic biases, the judiciary remains the ultimate authority.

Finally, the judgment emphasises that justice has a moral and emotional heart that artificial intelligence cannot copy. AI may identify patterns but it lacks compassion, ethical judgment, and understanding of human suffering. It maintains that judges bring empathy to their decisions, weighing fairness and the possibility of change against cold logic. Crucially, the legal authority of courts comes from the public belief that a human being has truly listened and acted with a conscience. Without these human parts, justice becomes a robotic exercise. The judgment declares that handing over core legal decisions to AI is the failure of judicial duty and a form of misconduct that harms fairness and proper legal procedure. The dignity of the court lies in the process of thinking and responding, a task described as purely and uniquely human. Unlike AI, which is limited by old data, human reasoning uses imagination to create new legal ideas and protect new rights. The courtroom must remain a place for careful thought rather than computer-led control. AI tools are strictly assistants and must never replace a judge's conscience or choice, as the judge alone is responsible for the accuracy and honesty of all court outcomes.

The Limits In The Current Case Law

The primary issue in the judicial stance on accountability lies in the lack of a technical bridge between legal responsibility and algorithmic transparency. Although the verdict holds a judge finally responsible, it fails to mandate the technical disclosures needed for a judge to actually oversee the AI. This creates a situation where a judge is legally responsible for a decision influenced by a process he can not fully understand or interrogate (Rodrigues, 2020). Without a requirement for algorithmic impact assessments or a right to explanation for the specific weights used by Judge-GPT, the accountability remains purely symbolic. The precedent assumes that judges can identify deep-seated data biases or hallucinations in LLM outputs, which overestimates human ability to spot errors in complex black-box systems.

Moreover, the current approach ignores the accountability gap caused by external software development. Since the Judge-GPT used by District Judiciary was developed in collaboration with outside partner ETH Zurich, the judgment does not clarify how a litigant can challenge the underlying code or data if they are found to be discriminatory. By focusing only on the judge's authority, the court overlooks the pre-decision phase where biased programming can subtly influence a judge's perception of a case (Vargas-Murillo et al., 2024). Without a way to audit developers, systemic errors could be replicated across thousands of cases before they are even detected by a single human reviewer.

The efficacy of this accountability model is further weakened by the lack of a reproducibility standard. In the traditional legal setting, accountability is maintained because a higher court can review the exact logic of a lower court. However, many AI systems are non-deterministic, meaning they may produce

different outputs for the same query (Blair-Stanek & Dumre, 2025). Without a mandatory audit trail to log every AI interaction, the appellate process is blinded (Wang, 2020). The judiciary is essentially adopting a trust-but-verify policy without providing the actual tools or rules needed to verify the machine's logic.

Additionally, the law fails to distinguish between human-made errors and algorithmic logic failures. If a judge relies on an AI summary of a precedent that omits a crucial piece of evidence, and that omission is due to a flaw in the Judge-GPT architecture, the current precedent simply put the liability onto the judge. This does not solve the problem of the flawed technology, it merely shifts the blame while allowing the faulty system to remain in use. To be truly effective, accountability must move from a general policy of human responsibility to a specific set of technical standards that include mandatory bias testing, disclosure of data sources, and the right for litigants to contest the use of specific algorithms.

Another key gap in this precedent is the lack of pre-use testing standards. Unlike the European Union's AI Act, which classifies judicial AI as high-risk and requires outside audits, Pakistan has no formal framework to certify tools like Judge-GPT (Mökander, 2023; Munir, 2024; Verma et al., 2025). This indicates a judge might use a tool that has never been tested for statistical drift or data poisoning, yet the judgment would still carry the weight of law. The precedent sets a duty of care for the judge but no duty of quality for the technology itself. The current approach also fails to separate assistive AI from interpretive AI. By treating all models the same, it ignores the unique risks of generative hallucinations in drafting (Dahl et al., 2024; Curran et al., 2023). Without an audit trail—a logged history of how AI was prompted and what the output was—or laws for algorithmic transparency, appeal courts cannot check if a lower court's judgment was skewed by a faulty AI. This makes the right to challenge evidence meaningless against opaque models.

Also, relying on human-in-the-loop as the main safeguard ignores cognitive offloading. As cases pile up in district courts, the pressure to clear backlogs will lead judges to accept AI advice without real review (Pruss, 2021). This judgment does not elaborate on when a judge's over-reliance on AI becomes misconduct. In the end, unclear guidelines for the use of AI in courts risks a black-box system that confuses the public. When people cannot see how a judge used AI, it weakens the right to a reasoned judgment. This opacity makes justice feel like it is being handled by cold logic instead of human conscience, which hurts the court's moral standing. Without mandatory disclosure of AI use, the public may stop trusting the judiciary, fearing that the system values speed over actual fairness.

7. PROPOSED JUDICIAL AI ACCOUNTABILITY FRAMEWORK

The use of AI in Pakistani courts must not compromise fairness or human control. To protect the rule of law, this paper proposes a framework for judiciary. It ensures that technology only assists judges instead of replacing them. The framework sets strict rules for transparency, human oversight, and developer liability. At its core, this approach changes how we view machine-generated work in legal processes.

Cross-Examination Of AI Insights

The framework treats AI as a rebuttable witness whose outputs must pass adversarial testing. Using common law rules of expert evidence, any AI output—whether a case summary or risk assessment—must meet three conditions: valid underlying data, a reliable model version, and freedom from algorithmic bias. The original query must be shared with all parties, and litigants may cross-examine the algorithm through a qualified technical representative or a court-appointed amicus curiae.

Substantive Review And Human Oversight

To prevent judges from simply rubber-stamping machine recommendations, the framework introduces the Substantive Review Standard. Any judge using AI in adjudication must record a Statement of Human Intervention, noting points of disagreement with the tool or the reasoning used to validate its output against traditional legal sources such as the Pakistan Law Digest. Oversight is meaningful only when the judge makes an independent evaluation, understands the machine's logic, and retains an

absolute right to override it. A rejection-by-default policy applies to any unverified AI outputs, keeping the decision's moral and intellectual weight with the human judge.

Transparency And The Reasoned Output Mandate

Litigants have a right to know when AI has influenced their case and how the algorithm reached its conclusion. Systems must disclose the decision parameters and formulas used in their outputs, and any tool deployed in the Pakistani Judiciary must be explainable-by-design (Arshad et al., 2025; Dancy & Žalnieriūtė, 2025). If a system cannot produce a step-by-step logical trace of its reasoning, its output is not admissible for a case's judgment. Judicial Technical Auditors translate technical algorithmic information into plain legal language, while developers must maintain response teams and disclose data logging and testing records, ensuring that the risk of black-box AI falls on companies and users, not on victims of errors. A human-on-appeal approach keeps nuanced reasoning at the appellate level, positioning AI as a support rather than a replacement for human expertise.

Required Liability And Transparency Measures

The framework adopts a human-centred approach to AI liability to prevent an accountability vaccum. A Multi-Tiered Liability Model separates judicial discretion from technical failure: a judge remains solely responsible for any judgment under a mandatory Duty of Verification, and failure to check an AI-generated summary or citation that causes a miscarriage of justice is treated as professional negligence rather than a software glitch. Developers and vendors bear strict liability for Product Integrity, so technical flaws such as fabricated precedents rest with the producer. Before deployment in the High Courts or District Courts, any tool must obtain a certificate of Veracity through a forensic audit by the AI Directorate confirming an error rate below a statutory safe limit, treating judicial AI as a high-risk product.

Digital Audit Trails And Traceability

The framework further mandates Forensic Digital Audit Trails through an unchangeable, time-stamped system recording every judge-AI interaction, including dataset versions and unprocessed machine output. These trails create permanent evidence of AI-assisted reasoning, allowing appellate courts to verify whether a lower court relied on a hallucinated precedent and bringing hidden algorithmic control to an end. Under the litigants' Right to Algorithmic Discovery, all parties may access the full audit trail to challenge the system's input-output integrity, and defence teams can trace which weightings and data points drove outcomes such as higher bail amounts tied to biased socio-economic information, demanding justification for those determinations in open court. Building on this, the Standard of Reproducibility requires that, for any AI-assisted decision to be legally binding, the machine's logic must be reproduced by an independent technical auditor, and non-deterministic systems are inadmissible for core judicial functions. Through these combined measures, the framework ensures that technology remains a fair extension of the legal process, that litigants can challenge every machine-driven step, and that judicial AI in Pakistan operates under verifiable, accountable, and transparent safeguards.

8. RESULTS

- Pakistani Law contains no sector-specific accountability rule for AI-assisted judicial decisions.
- The landmark judgment in the case of Ishfaq Ahmed v Mushtaq Ahmed assigns liability to the judge but creates no mechanism to interrogate machine reasoning.
- The National AI Policy 2025 treats judicial AI like any other high-impact sector, structurally mismatching it.
- A digital divide exists between superior and subordinate courts, risking codified access to justice gaps.
- Senior judicial officials have already used Chat-GPT in both civil and criminal matters, outpacing formal regulatory laws.

- The Smart Court System and National AI Policy 2025 both prioritise throughput over liability, audit, and oversight.
- Experimental sandboxes lack Article 10-A protection, so litigants have no defined remedy if pilot-side AI errors influence their case.

9. RECOMMENDATIONS

- Enact a dedicated Judicial AI Accountability Act giving statutory force to the proposed framework.
- Direct the Law and Justice Commission of Pakistan to issue binding Model Court Rules on AI-assisted adjudication.
- Constitute an independent judicial AI oversight authority under the Supreme Court.
- Mandate tiered AI certification through the Federal Judicial Academy for judges, staff, and auditors.
- Ring-fence a dedicated budget to bring subordinate courts to a minimum digital baseline.

10. CONCLUSION

This study set out to determine whether Pakistan's existing legal frameworks, national policy, and precedents are equipped to govern the rapid integration of AI into decision-making processes. The doctrinal analysis conducted across the Constitution, PECA 2016, Draft Data Protection Bill, and the IT Ordinance 2002 establishes that none of these were drafted with algorithmic adjudication in mind, while the National AI Policy 2025, and Supreme Court Landmark Judgment "Ishfaq Ahmed v Mushtaq Ahmed", though welcome, translate neither ambition into binding rules nor principles into operational safeguards. Therefore, the Judicial AI Accountability Framework proposed in this paper is offered as a constitutional necessity rather than a policy preference. It converts vague promises of human oversight into clear legal duties—sets liability standards, reproducibility, auditability, and disclosure. In this way technology in Pakistani courts remains an instrument of human justice not a substitute for it. As automation grows, the main question is no longer whether AI will change judicial outcomes in Pakistan but under whose rules it will do so. The framework given here provides one constitutionally grounded answer. It is meant to start the legal and constitutional changes that are needed now.

REFERENCES

Ahn, M. J., & Chen, Y.-C. (2020). Artificial intelligence in government: Potentials, challenges, and the future. In dg.o '20: Proceedings of the 21st Annual International Conference on Digital Government Research (pp. 243-252). Association for Computing Machinery. https://doi.org/10.1145/3396956.3398260

Arshad, S. A., Hashmi, M. A. I., Bano, N., & Saleem, H. A. R. (2025). Al judges and robotic justice: Are Pakistani courts ready for the future? ACADEMIA International Journal for Social Sciences, 4(3), 1039-1050. https://doi.org/10.63056/ACAD.004.01.0427

Afzal, M., & Ghawaas, H. G. (2025). Reforming civil justice in Pakistan through AI: A critical study of legal technologies and their impact on access to justice. Dialogue Social Science Review, 3(11), 1–40.

Aziz, A., & Abid, I. (2024). Balancing speed and justice: AI-powered court systems and legal ethics in Pakistan. Advance Social Science Archives Journal, 2(4), 1255–1268.

Blair-Stanek, A., & Van Durme, B. (2025). LLMs provide unstable answers to legal questions. arXiv. https://doi.org/10.48550/arXiv.2502.05196

Cheong, B. C. (2024). Transparency and accountability in Al systems: safeguarding wellbeing in the age of algorithmic decision-making. Frontiers in Human Dynamics, 6, Article 1421273. https://doi.org/10.3389/fhumd.2024.1421273

Curran, S., Lansley, S., & Bethell, O. (2023). Hallucination is the last thing you need. arXiv. https://doi.org/10.48550/arXiv.2306.11520

Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024). Large legal fictions: Profiling legal hallucinations in large language models. Journal of Legal Analysis, 16(1), 64-93. https://doi.org/10.1093/jla/laae003

Dancy, T., & Zalnieriute, M. (2025). AI and transparency in judicial decision-making. Oxford Journal of Legal Studies. Advance online publication. https://doi.org/10.2139/ssrn.5331491

Farid, S., & Ghazanfar, M. (2025). AI, automation, and the Pakistan's judiciary: Mapping the ethical and legal boundaries of algorithmic justice. Advance Social Science Archive Journal, 4(1), 4176-4188. https://doi.org/10.5281/zenodo.17177821

Farrukh, T., Qureshi, F. N., & Abbasi, S. (2024). Artificial intelligence (AI) in legal system. Journal of Independent Studies and Research Computing, 22(1), 25-26. https://doi.org/10.31645/JISRC.24.22.1.2

Gondal, A. Q., Ahmad, M., & Hatta, Z. (2025). Revolutionizing criminal justice: The role of digitalization and AI in Pakistan's legal system. Al Khadim Research Journal of Islamic Culture and Civilisation, 6(1), 25–81.

Ishfaq Ahmed v. Mushtaq Ahmed, 2025 SCP 112 (Supreme Court of Pakistan 2025).

Javed, K., & Li, J. (2025). Bias in adjudication: Investigating the impact of artificial intelligence, media, financial and legal institutions in pursuit of social justice. PLOS ONE, 20(1), Article e0315270. https://doi.org/10.1371/journal.pone.0315270

Ministry of Information Technology and Telecommunication. (2025). National artificial intelligence policy 2025.

Muhammad Iqbal v. Zayad (Amir Munir, J., 28 March 2023). (https://courtingthelaw.com/wp-content/uploads/DOC-20230413-WA0052..pdf)

Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2024). Auditing large language models: a three-layered approach. AI and Ethics, 4(4), 1005-1115. https://doi.org/10.1007/s43681-023-00289-2

Munir, B., Ansari, A. A., & Arafat, Y. (2024). Artificial intelligence and legal decision-making in the USA and Pakistan: A critical appreciation of regulatory frameworks. Global Foreign Policies Review, VII(IV), 48-58. https://doi.org/10.31703/gfpr.2024(VII-IV).06

Qazi, M. (2023). Infusing digital technology in judicial operations in Pakistan: A critical analysis of global best practices and local initiatives. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4378378

Qureshi, G. M., Ahmed, F. A., & Chaudhary, F. (2025). Algorithmic justice and legal pluralism: Rethinking artificial intelligence regulation in Pakistan's hybrid legal system. Competitive Research Journal Archive, 3(2), 67–75.

Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. Journal of Responsible Technology, 4, Article 100005. https://doi.org/10.1016/j.jrt.2020.100005

Soomro, A. S., Baig, K., Khan, S. M., & Laghari, N. A. (2024). Revolutionizing justice: Strategic approaches to AI in Pakistan's courts. The Critical Review of Social Sciences Studies, 2(2), 508-526. https://doi.org/10.59075/aja7bm64

Vargas-Murillo, A. R., Pari-Bedoya, I. N. M. De la A., Turriate-Guzman, A. M., Delgado-Chávez, C. A., & Sanchez-Paucar, F. (2024). Transforming justice: Implications of artificial intelligence in legal systems. Academic Journal of Interdisciplinary Studies, 13(2), 433-442. https://doi.org/10.36941/ajis-2024-0059

Wang, N. (2020). "Black box justice": Robot judges and AI-based judgment processes in China's court system. In 2020 IEEE International Symposium on Technology and Society (ISTAS) (pp. 243-252). IEEE. https://doi.org/10.1109/ISTAS50296.2020.9462216

Mangi, D. B., Butro, I., & Memon, T. A. (2025). AI and criminal liability: Theoretical dilemmas in applying criminal law to artificial intelligence. The Critical Review of Social Sciences Studies, 3(2), 2174-2186. https://doi.org/10.59075/cxwpfa06

IEER
Institute for Excellence in Education & Research