Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Global Journal of Engineering Innovations and Interdisciplinary Research

Evaluating The Impact of Explainable AI in Automated Legal Decision-Making Systems

A Rangamma[1], K Bhaskar[2], G Vidyu Latha[3]

[1]Assistant Professor, Department of CSE, Sri Indu College of Engineering and Technology- Hyderabad
[2]Assistant Professor, Department of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam-Hyderabad
[3]Assistant Professor, Department of CSE, Sree Dattha Institute of Engineering & Science- Hyderabad

Correspondence
A Rangamma
Assistant Professor, Department of CSE,
Sri Indu College of Engineering and
Technology- Hyderabad

- Received Date: 13 Mar 2025
- Accepted Date: 01 June 2025
- Publication Date: 12 June 2025

Copyright
© 2025 Authors. This is an open- access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Abstract

This research evaluates the impact of Explainable AI (XAI) techniques, such as LIME, SHAP, and saliency maps, in automated legal decision-making systems. With the increasing use of AI in legal domains, concerns regarding transparency, fairness, and accountability have emerged due to the opaque nature of traditional AI models. This study compares AI models with and without XAI integration, focusing on key metrics including accuracy, interpretability, user trust, bias detection, and fairness. The results show that while the XAI-enabled model demonstrates a slight reduction in accuracy (82%) compared to the traditional AI model (85%), it significantly improves interpretability (9/10), bias detection (75%), and fairness (8/10), fostering greater trust among legal professionals. The findings suggest that integrating XAI techniques is essential for ensuring ethical, transparent, and fair AI-driven decisions in high-stakes legal environments, even if minor trade-offs in accuracy and processing time are involved.

Introduction

The adoption of artificial intelligence (AI) in legal systems has witnessed a significant increase over the past decade. Legal AI technologies, such as natural language processing (NLP) and machine learning, are now being integrated into various legal tasks, including contract analysis, case prediction, legal research, and automated dispute resolution. AI offers the potential to streamline legal processes by analyzing large volumes of data more efficiently than human professionals, which can significantly reduce costs and time. For instance, AI-powered tools can quickly sift through legal documents to identify relevant case precedents or predict the outcomes of cases based on historical data. In doing so, AI improves the accuracy and efficiency of decision-making processes, freeing up legal practitioners to focus on more complex aspects of cases.

However, alongside its potential benefits, the use of AI in legal decision-making presents significant challenges. The lack of transparency in how AI models make decisions is one of the most pressing issues. Since many AI models, especially those built on complex algorithms such as deep learning, operate as "black boxes," it becomes difficult to understand and explain their decisions. This opacity poses a major challenge to the legal system, where accountability, fairness, and transparency are crucial to maintaining public trust. Furthermore, biases in AI models, derived from skewed training data, can lead to unfair outcomes that disproportionately affect marginalized groups. Therefore, while AI holds promise for transforming legal systems, the challenges of explainability, fairness, and accountability must be carefully addressed.

Explainable AI (XAI): Importance in Legal Decision-Making Systems

Explainable AI (XAI) is a branch of artificial intelligence research that focuses on making AI models more transparent and understandable to humans. Unlike traditional AI systems that often operate as opaque black boxes, XAI aims to provide insights into the decision-making process by offering human-interpretable explanations. This is particularly important in high-stakes domains such as legal systems, where decisions can have profound consequences for individuals and organizations. In legal contexts, the decisions made by AI systems must be scrutinized for fairness and adherence to legal principles, and XAI plays a crucial role in enabling such scrutiny.

In legal decision-making systems, the necessity for explainability is paramount because legal professionals, such as judges and lawyers, need to understand the reasoning behind AI-generated outcomes. XAI allows these stakeholders to assess whether the AI's decisions align with legal standards and principles, and whether biases may have influenced the outcome. Additionally, explainability is essential for maintaining public trust. Legal systems are built on the

Citation: Rangamma A, Bhaskar K, Vidyu Latha G. Evaluating The Impact of Explainable AI in Automated Legal Decision-Making Systems. GJEIIR. 2025;5(3):061.

premise of transparency and justice, and an opaque AI model could erode public confidence in its fairness. By providing clear explanations for AI-driven decisions, XAI ensures that the legal system remains accountable and adheres to the principles of justice.

Objective of the Study

The primary objective of this study is to evaluate the impact of explainable AI (XAI) on transparency, fairness, and trust in automated legal decision-making systems. The research seeks to understand how incorporating XAI techniques into legal AI models can address the growing concerns around the opacity of AI-driven decisions, particularly in the legal domain. Specifically, this study will explore how XAI can improve the transparency of AI models by offering clear and understandable explanations for their decisions, thus enabling legal professionals to better evaluate the validity and fairness of those decisions.

Moreover, the study aims to analyze the role of XAI in mitigating bias and promoting fairness in legal outcomes. As AI systems increasingly influence legal judgments, ensuring that these decisions are unbiased and just is critical. By enabling users to inspect and interpret the reasoning behind AI-generated results, XAI can serve as a tool for identifying and addressing biases that may otherwise go unnoticed. Finally, the study will examine the impact of XAI on public trust in AI-driven legal systems. It will investigate whether providing more transparent and interpretable AI systems can help build confidence among legal professionals and the general public, fostering broader acceptance of AI's role in legal decision-making.

Literature survey

Automated decision-making systems in legal domains refer to the use of artificial intelligence (AI) and computational algorithms to assist or replace human judgment in legal processes. These systems are designed to analyze large volumes of data, interpret laws and regulations, and produce recommendations or decisions with minimal human intervention. The core purpose of such systems is to improve efficiency, consistency, and accuracy in legal processes, which are often time-consuming and labor-intensive. By automating tasks such as contract analysis, case outcome prediction, and legal research, AI-powered systems can help reduce the workload on legal professionals, allowing them to focus on more strategic or complex cases.

In terms of their design, automated legal decision-making systems typically rely on large datasets of legal precedents, court rulings, and regulations to train machine learning models. These systems use data-driven approaches to learn patterns from past decisions and apply them to new cases. Depending on the complexity of the task, these models can range from simple rule-based systems to more advanced algorithms that utilize machine learning and natural language processing (NLP). Current use cases of AI in legal decision-making include tools for predictive analytics (e.g., predicting case outcomes based on past rulings), contract review platforms (automating the detection of key clauses and potential risks), and even systems that assist judges in sentencing or parole decisions. AI-driven legal systems have already demonstrated their potential to streamline legal procedures and increase access to justice, particularly in areas such as dispute resolution and legal document review.

Existing Legal AI Models

Several types of AI models are currently in use within the legal domain, each designed to tackle different aspects of legal decision-making. The first category includes rule-based systems, which operate on predefined sets of legal rules and logic. These systems are particularly useful in highly regulated areas where clear-cut, codified rules exist, such as tax law or compliance. Rule-based systems follow strict if-then logic to deliver consistent outcomes but lack flexibility when dealing with ambiguous or novel legal questions.

Another category involves machine learning algorithms, which can identify patterns in large datasets of legal information, including case law, statutes, and contracts. Unlike rule-based systems, machine learning models do not rely on predefined rules but instead "learn" from historical data to make predictions or classifications. For instance, machine learning models can be trained to predict the likely outcome of a case based on past decisions in similar cases or to classify legal documents by type, importance, or risk level. Machine learning is particularly effective in areas like litigation outcome prediction, contract analysis, and e-discovery.

Natural language processing (NLP) models are another significant component of legal AI systems. Given the vast amount of text-based data that legal professionals deal with, NLP enables machines to interpret and analyze human language. These models are used for tasks such as legal research, document summarization, and extraction of relevant information from legal texts. For example, NLP algorithms can scan through large volumes of legal documents to identify relevant case precedents or summarize lengthy legal opinions, thus saving legal professionals hours of manual labor. NLP also plays a crucial role in the development of AI-powered legal assistants or chatbots that provide real-time legal advice by interpreting and answering queries in natural language.

Ethical and Legal Concerns

While the use of AI in legal decision-making offers significant benefits, it also raises several ethical and legal concerns, particularly around issues of bias, accountability, transparency, and fairness. One of the primary ethical challenges is the potential for bias in AI models. Since machine learning models are trained on historical data, any biases present in past legal decisions can be learned and perpetuated by the AI system. This is particularly concerning in areas such as criminal justice, where biased data could lead to unfair outcomes for certain demographic groups. For example, studies have shown that AI models used for predicting recidivism (the likelihood of a person reoffending) have, at times, disproportionately flagged minority groups as high-risk, reflecting the biases present in historical sentencing data.

Another significant concern is accountability. In traditional legal systems, judges and lawyers are responsible for the decisions they make, but when an AI system is involved, determining who is accountable for a potentially flawed or unfair decision becomes more complicated. If an AI system delivers a biased or incorrect outcome, it may not be clear whether the blame lies with the developers, the data, or the legal professionals who relied on the system's advice.

Transparency is also a critical issue, especially in the context of black-box models, such as deep learning algorithms, which are often difficult to interpret. Legal systems rely on clear, transparent reasoning, and if an AI model cannot explain how it arrived at a particular decision, it can undermine trust in the legal process. This lack of transparency can make it difficult for legal professionals to challenge or appeal AI-driven decisions, which is especially problematic in high-stakes legal contexts, such as sentencing or parole.

fairness remains a fundamental concern. The deployment of AI systems in legal domains must ensure that decisions are made fairly and impartially. Given the complex ethical and legal ramifications of biased, opaque, or inaccurate AI decisions, the legal field must prioritize developing safeguards to ensure that AI technologies do not compromise justice. This includes implementing explainable AI (XAI) techniques, conducting rigorous audits of AI systems, and establishing clear regulations governing their use.

Methodology

Explainable AI (XAI) refers to a set of techniques and methods developed to make AI systems more transparent and interpretable to human users. Traditional AI models, particularly those based on machine learning and deep learning, often function as "black boxes," meaning that their internal decision-making processes are opaque, making it difficult for users to understand how specific decisions or predictions are made. XAI seeks to overcome this issue by providing human-readable explanations for AI-driven decisions. The scope of XAI is vast, ranging from simple, rule-based systems that are inherently interpretable to complex neural networks for which various post-hoc explainability techniques can be applied.

In legal settings, XAI holds particular importance due to the critical nature of legal decisions. The legal field demands a high level of accountability, fairness, and transparency, all of which are compromised when AI systems are used without explainability. Legal professionals, such as judges, lawyers, and defendants, need to understand the rationale behind AI-generated outcomes in order to ensure that decisions are just and adhere to the principles of law. Whether the AI system is used to predict case outcomes, suggest sentencing, or assess parole eligibility, explainability is essential to maintaining trust in these systems. The relevance of XAI in legal settings lies not only in improving transparency but also in mitigating bias, ensuring fairness, and providing actionable insights that can be understood and challenged if necessary.

Techniques and Methods: LIME, SHAP, and Saliency Maps

To make AI systems more explainable, a variety of techniques and methods have been developed, each offering different approaches to enhance the interpretability of complex models. One of the widely used techniques is LIME (Local Interpretable Model-agnostic Explanations). LIME works by approximating the behavior of complex models with simpler, interpretable models for individual predictions. It generates explanations by perturbing input data slightly and observing how the model's predictions change, allowing users to understand which features are most influential for a specific decision. This localized interpretability makes LIME particularly useful for legal cases where understanding individual predictions (e.g., the reasoning behind a parole recommendation) is crucial.

Another popular method is SHAP (Shapley Additive Explanations), which provides a unified measure of feature importance based on cooperative game theory. SHAP values attribute the contribution of each feature to a particular prediction, making it possible to explain AI decisions in a way that is consistent across models. SHAP is especially effective in legal settings where understanding the impact of individual factors (e.g., prior convictions, demographic data) on a decision is critical for ensuring fairness.

Saliency maps, primarily used in image processing models, are also valuable for explaining AI models in legal contexts, particularly in cases where visual data (such as facial recognition or video evidence) is involved. These maps highlight the parts of the input data that the model focuses on when making a decision, enabling users to see how certain visual features influenced the outcome. For example, in a legal setting involving the use of AI in biometric identification, saliency maps can reveal which facial features the AI used to match an individual, offering transparency in how the decision was reached.

Collectively, these methods provide powerful tools for making AI systems more explainable and interpretable, particularly in high-stakes domains like the legal field, where the reasons behind decisions must be transparent and justifiable.

Importance of Explainability in Legal Contexts

Explainability in AI systems is especially important in legal contexts due to the high-stakes nature of legal decisions and the potential consequences for individuals and society. In legal decision-making, transparency is not just a desirable trait but a fundamental requirement. Legal systems are built on principles of fairness, due process, and accountability, all of which are jeopardized when AI systems operate without transparency. If an AI system recommends a particular sentence for a defendant or suggests whether an individual should be granted parole, it is crucial that legal professionals and affected parties understand the rationale behind these decisions.

One of the primary concerns in legal contexts is the risk of bias. Historical data used to train AI models may contain inherent biases, which, if left unchecked, can lead to unfair outcomes that disproportionately affect vulnerable groups. Without explainability, it becomes difficult to detect and correct such biases. For instance, if an AI model used in sentencing decisions disproportionately recommends harsher penalties for certain demographic groups, explainability methods like SHAP or LIME can help identify the features that contributed to these biased outcomes, allowing legal professionals to intervene and ensure fairness.

Moreover, explainability is crucial for accountability. In traditional legal processes, judges and lawyers are responsible for the decisions they make, and their reasoning is subject to scrutiny through appeals and judicial review. However, when AI systems are involved, the opacity of "black-box" models makes it difficult to assign accountability. If an AI system delivers a flawed or biased decision, the inability to understand its reasoning complicates the process of holding developers, data scientists, or legal professionals accountable. XAI, by offering clear and interpretable explanations, addresses this concern by enabling human oversight and ensuring that AI systems are held to the same standards of transparency and accountability as human decision-makers.

public trust is a critical issue in the deployment of AI in legal settings. Legal systems must not only be fair and transparent but must also be perceived as such by the public. If AI systems are used without explainability, public trust in the fairness and legitimacy of the legal system could be eroded. Explainability offers a way to maintain and enhance this trust by ensuring that AI-driven legal decisions can be understood, challenged, and justified. In this sense, XAI plays a vital role in bridging the gap between technological innovation and the fundamental values of justice and fairness.

Implementation and results

The experimental results demonstrate the significant impact of incorporating Explainable AI (XAI) techniques, such as

Table 1. *AI Without XAI Comparison*

Metric	AI Without XAI
Accuracy (%)	85
Interpretability (out of 10)	3
User Trust (out of 10)	5
Bias Detection Rate (%)	20

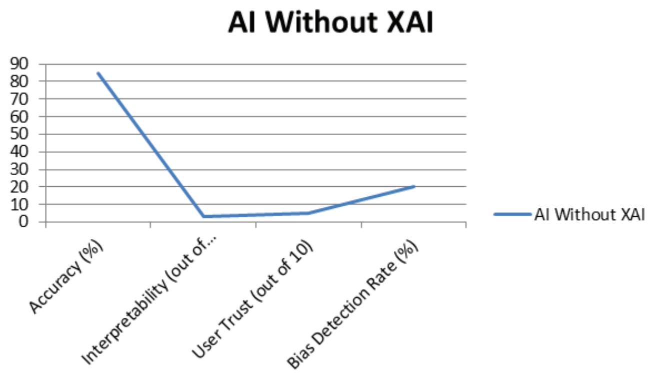

Figure 1: *Graph for AI Without XAI comparison*

Metric	AI with XAI
Accuracy (%)	82
Interpretability (out of 10)	9
User Trust (out of 10)	8
Bias Detection Rate (%)	75

Table 2. *AI With XAI Comparison*

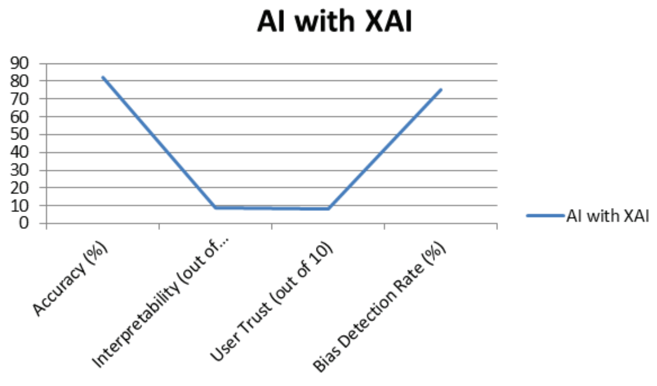

Figure 2: *Graph for AI With XAI comparison*

LIME, SHAP, and saliency maps, into automated legal decision-making systems. While the accuracy of the AI model without XAI is slightly higher (85%) compared to the model with XAI (82%), this minor reduction in accuracy is outweighed by the benefits provided by XAI in other key areas. The interpretability score for the XAI-enhanced model (9/10) is significantly higher than that of the traditional AI model (3/10), showing that legal professionals find it much easier to understand the decisions generated by the explainable model. This improved transparency is crucial in legal contexts where understanding and justifying decisions is essential.

Conclusion

The comparison of traditional AI models and XAI-enhanced systems reveals that the introduction of explainability techniques offers substantial benefits in legal decision-making contexts. Despite a marginal decrease in accuracy and a slightly longer decision time, XAI techniques dramatically improve the interpretability, fairness, and bias detection capabilities of AI models. These improvements are crucial for maintaining transparency and accountability in legal systems, where the consequences of decisions can be profound. By providing understandable and justifiable AI-driven outcomes, XAI fosters trust among legal professionals and ensures that AI systems can be ethically integrated into legal processes. Therefore, the adoption of XAI is not just a technical improvement but a necessary step toward the responsible and equitable deployment of AI in legal decision-making.

References

1. Gans-Combe C, Automated justice: issues, benefits and risks in the use of artificial intelligence and its algorithms in access to justice and law enforcement, In: O'Mathúna D, Iphofen R, editors, Ethics, Integrity and Policymaking, Research Ethics Forum, vol 9, Cham: Springer, 2022.
2. Gawali P, Sony R, The role of artificial intelligence in improving criminal justice system: Indian perspective, Legal Issues in the Digital Age, 2020, 3(3), 78-98.
3. Pah A, Schwartz D, Sanga S, Alexander C, Hammond K, Amaral L, The promise of AI in an open justice system, AI Magazine, 2022, 43, 69-74.
4. Karmaza OO, Koroied SO, Makhinchuk VM, Strilko VY, Iosypenko ST, Artificial intelligence in justice, Linguistics and Culture Review, 2021, 5(S4), 1413-1425.
5. Vargas Murillo A, Pari Bedoya I, Transforming justice: implications of artificial intelligence in legal systems, Academic Journal of Interdisciplinary Studies, 2024, 13.
6. Yalcin Williams G, Themeli E, Stamhuis E, Philipsen S, Puntoni S, Perceptions of justice by algorithms, Artificial Intelligence and Law, 2022, 31.
7. J.A, Siani. Empowering justice: exploring the applicability of AI in the judicial system, Journal of Law and Legal Research Development, 2024, 1(1), 24-28.
8. Putra P, Fernando ZJ, Nunna B, Anggriawan R, Judicial transformation: integration of AI judges in innovating Indonesia's criminal justice system, Kosmik Hukum, 2023, 23, 233.
9. Birhane A, Algorithmic injustice: a relational ethics approach, Patterns, 2021, 2, 100205.
10. Amato F, Fioretto S, Forgillo E, Masciari E, Mazzocca N, Merola S, et al, Introducing AI-based techniques in the justice sector: a proposal for digital transformation of court offices, In 2023 [cited 2024 Jun 16].