Frontiers in Computer Science and Artificial Intelligence
ISSN: 2798-8048
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

Artificial Intelligence in Judicial Decision-Making: Assessing the Legal, Ethical, and Institutional Implications of Machine Learning-Based Decision Support Systems in U.S. Courts

Mohammed Nazmul Islam Miah ✉[1], **MD Joshim Uddin**[2], **Manohar Kakumani**[3]

[1] *Management Sciences and Quantitative Methods, Gannon University, Erie, PA, USA*

[2] *Master's of Law, ASA University of Bangladesh*

[3] *Master of Science in Computer and Information Science, Gannon University, Erie, PA, USA*

Corresponding Author: Mohammed Nazmul Islam Miah**, E-mail:** islammia001@gannon.edu

ABSTRACT

The increasing adoption of artificial intelligence in the criminal justice system has created opportunities to improve judicial information processing, predictive assessment, and administrative efficiency, while raising concerns regarding fairness, transparency, and institutional accountability. This study examines the legal, ethical, and institutional implications of machine learning-based decision-support systems in U.S. courts, with particular attention to pretrial risk assessment, sentencing support, case outcome prediction, legal information retrieval, and judicial document classification. The research adopts a structured secondary-research design and a comparative analytical framework to evaluate the potential benefits, limitations, and governance requirements associated with these applications. The findings indicate that machine learning can support judicial administration by improving information retrieval, identifying patterns in large collections of legal records, facilitating document classification, and promoting consistency in selected analytical tasks. However, algorithmic predictions may reproduce historical inequalities, generate discriminatory outcomes, obscure the reasoning underlying judicial recommendations, and create opportunities for excessive reliance on automated assessments. These concerns are particularly consequential when algorithmic outputs influence decisions affecting individual liberty and legal rights. The study further establishes that predictive performance alone is insufficient to demonstrate the suitability of machine learning for judicial decision-making. Responsible adoption requires meaningful judicial oversight, procedural transparency, independent auditing, representative data, fairness assessment, explainability, contestability, and continuous institutional monitoring. The study concludes that machine learning can contribute to the effectiveness of U.S. courts when deployed within a governance framework that preserves due process, judicial independence, individual rights, and institutional accountability.

KEYWORDS

Artificial intelligence; machine learning; judicial decision-making; algorithmic decision-making; U.S. courts; algorithmic bias; legal technology; risk assessment

ARTICLE INFORMATION

ACCEPTED: 18 October 2026 **PUBLISHED:** 18 November 2026 **DOI:** 10.32996/fcsai.2026.5.12.3

1. Introduction

1.1 Background and Emergence of AI in Judicial Decision-Making

The increasing integration of artificial intelligence (AI) into professional and institutional environments has significantly transformed how complex information is processed, analyzed, and utilized to support decision-making. Within the legal sector, technological advancements in machine learning (ML), natural language processing, predictive analytics, and automated information retrieval have created opportunities to improve legal research, case management, evidence analysis, and judicial decision-making. The growing availability of large-scale digital legal datasets, combined with advances in computational learning algorithms, has enabled the development of systems capable of identifying patterns in historical judicial decisions, estimating potential case outcomes, and generating analytical insights that may assist legal professionals. These developments have contributed to the emergence of computational approaches to legal reasoning, commonly associated with the broader field of artificial intelligence and law. Ashley (2017) explains that AI technologies are increasingly relevant to legal practice because they

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

enable the computational analysis of legal texts, judicial precedents, statutory provisions, and case-specific information, thereby supporting activities that traditionally required extensive manual examination by legal professionals [2]. Such capabilities have considerable implications for the United States judicial system, where courts routinely process substantial volumes of legal documentation, evidentiary material, procedural records, and case histories.

The application of AI within judicial institutions encompasses several technological categories, including conventional legal information systems, predictive analytical tools, and machine learning-based decision-support systems. Legal information systems primarily facilitate the storage, retrieval, organization, and examination of legal information, allowing judges, attorneys, and other legal professionals to access relevant statutes, precedents, and judicial records. Machine learning-based decision-support systems operate at a more analytical level by identifying statistical relationships within historical datasets, generating predictions, classifying legal information, and estimating the likelihood of particular outcomes. These systems can employ supervised learning algorithms, including logistic regression, decision trees, random forests, gradient boosting, and neural networks, depending on the nature of the judicial problem and the characteristics of the available data. Surden (2019) observes that AI applications in law extend beyond conventional legal information retrieval, encompassing computational techniques that can perform classification, prediction, and other analytical tasks associated with legal processes [20]. This development creates opportunities for integrating data-driven analytical capabilities into judicial workflows, particularly in situations involving large volumes of historical information and decisions requiring the assessment of multiple interacting factors.

Within the United States, one particularly significant application of machine learning in judicial decision-making involves risk assessment in criminal proceedings. Algorithmic risk assessment instruments are designed to estimate the probability that an individual may engage in future criminal activity, fail to appear in court, or violate conditions of pretrial release. Such predictions may be considered during bail and pretrial release determinations, sentencing proceedings, and decisions concerning probation or parole. These applications frequently rely on historical criminal justice records, demographic characteristics, prior convictions, age-related variables, and other information associated with predicted outcomes. Kleinberg et al. (2018) examined the relationship between human decision-making and machine-generated predictions, demonstrating how predictive algorithms can potentially improve decisions in contexts where human decision-makers must evaluate uncertain future outcomes [11]. Their research provides an important analytical foundation for understanding the potential contribution of machine learning to judicial environments, particularly where decisions involve estimating risks that cannot be established with complete certainty. Beyond risk assessment, computational methods can support case outcome prediction, legal document classification, identification of relevant precedents, and the examination of patterns in judicial rulings. These applications demonstrate the potential breadth of machine learning within different stages of the judicial process.

The significance of these technological developments is particularly evident in the context of criminal justice, where judicial decisions can directly influence individual liberty, personal security, economic circumstances, and constitutional rights. Decisions concerning bail, pretrial detention, sentencing, and probation can have substantial consequences for defendants, victims, and wider communities. Consequently, the introduction of algorithmic recommendations into these processes raises important questions concerning the reliability, fairness, and legitimacy of computationally supported judgments. Machine learning models derive their predictive capabilities from patterns contained in historical datasets, meaning that their outputs can reflect the characteristics, limitations, and potential inequalities present in the information used during training. Furthermore, differences in model architecture, feature selection, data quality, and evaluation criteria can influence the predictions generated by individual systems. Surden (2019) emphasizes the importance of understanding the technical characteristics and limitations of AI applications when evaluating their suitability for legal contexts, particularly because legal processes involve considerations of interpretation, institutional responsibility, and the application of established legal principles [20].

The increasing use of machine learning in judicial environments therefore introduces a multidimensional area of investigation involving computational performance, legal compatibility, ethical responsibility, and institutional governance. Although predictive models may assist judges in processing complex information and evaluating potential risks, their integration into judicial decision-making requires careful consideration of how algorithmic outputs are interpreted, communicated, and incorporated into judicial reasoning. The central concern is whether these systems can provide sufficiently reliable and legally defensible support while preserving fundamental principles such as judicial independence, procedural fairness, equality before the law, and the presumption of innocence. Examining these issues is particularly important in the United States, where constitutional protections, established judicial procedures, and the discretionary authority of judges impose significant requirements on the legitimacy of decision-making processes. Accordingly, understanding the emergence of AI in judicial decision-making requires an assessment of both its computational capabilities and its implications for the rights of individuals and the institutional responsibilities of American courts.

1.2 Problem Statement and Research Gap

The adoption of machine learning-based decision-support systems in judicial environments presents significant challenges concerning the relationship between algorithmic predictions, legal principles, and the exercise of judicial authority. Although these systems offer potential improvements in information processing, predictive analysis, and decision consistency, their application in legally consequential proceedings introduces concerns regarding algorithmic bias, discriminatory outcomes, limited transparency, accountability, and the preservation of due process. Machine learning algorithms are generally developed using historical datasets

from which statistical relationships are identified and subsequently applied to generate predictions about new cases or individuals. However, historical criminal justice data may contain patterns influenced by socioeconomic inequalities, differences in law enforcement practices, variations in policing intensity, and disparities in criminal justice outcomes. When such information is incorporated into predictive models without adequate examination, the resulting systems may reproduce or amplify existing inequalities. These concerns are particularly consequential in judicial decision-making because algorithmically generated risk estimates can influence determinations involving pretrial detention, sentencing severity, and other decisions that affect individual freedom. Consequently, the introduction of machine learning into courts creates a fundamental problem concerning whether computational predictions can satisfy the legal and ethical standards expected of judicial processes.

An important illustration of these challenges is provided by State v. Loomis (2016), a significant Wisconsin Supreme Court decision concerning the use of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) risk assessment instrument during criminal sentencing. The case involved Eric Loomis, whose sentencing proceedings incorporated a COMPAS-generated assessment of his potential risk of recidivism. Loomis challenged the use of the instrument, raising concerns about the proprietary nature of its underlying methodology, the inability to independently examine its calculations, and the potential implications for his constitutional rights. The Wisconsin Supreme Court permitted the consideration of COMPAS results under specified limitations, emphasizing that the assessment should not independently determine sentencing outcomes. The court also acknowledged concerns associated with the instrument, including its proprietary methodology, potential gender-related disparities, and the difficulty of establishing the accuracy of individual predictions. *State v. Loomis* (2016) therefore illustrates the legal difficulties associated with incorporating algorithmic risk assessments into judicial reasoning, particularly when defendants and their legal representatives may have limited access to the technical information necessary to evaluate or challenge algorithmic recommendations [18].

The COMPAS controversy also demonstrates the difficulty of reconciling predictive performance with the legal requirements of fairness and individual assessment. A machine learning model may achieve satisfactory statistical performance across a dataset while producing different error distributions among demographic groups or generating predictions that are insufficiently informative about a particular defendant. Such outcomes raise questions concerning the appropriate definition of algorithmic fairness, the selection of relevant predictive variables, and the extent to which historical associations should influence judicial decisions. Furthermore, proprietary algorithms can create additional accountability challenges when their internal operations, training procedures, or feature contributions are inaccessible to affected individuals and legal representatives. In judicial proceedings, the ability to understand and contest evidence is fundamental to procedural fairness. Where algorithmic recommendations influence judicial outcomes, limited explainability can complicate the examination of how a prediction was generated, whether its underlying assumptions are appropriate, and whether it should receive evidentiary or sentencing consideration. These challenges make the deployment of machine learning in courts a particularly sensitive institutional undertaking.

A further dimension of the problem concerns the potential transformation of judicial discretion and institutional accountability. Judges are expected to exercise independent judgment by considering the circumstances of individual cases, relevant legal principles, admissible evidence, and applicable sentencing requirements. The introduction of algorithmic recommendations may influence this process through automation bias, whereby human decision-makers place excessive confidence in computational outputs because they are perceived as objective, consistent, or scientifically established. Such dependence may gradually affect the extent to which judges independently evaluate individual circumstances and exercise their discretionary authority. It also raises questions concerning responsibility when a judicial decision incorporates an inaccurate, discriminatory, or inadequately validated algorithmic recommendation. If a court relies on a predictive system developed by an external vendor, determining responsibility for adverse outcomes may become more complicated because technical development, institutional procurement, judicial interpretation, and final decision-making may involve different actors. These considerations extend the problem beyond individual model performance to encompass the broader institutional legitimacy of courts and public confidence in judicial outcomes.

Despite the growing attention given to artificial intelligence in legal environments, important analytical gaps remain in understanding the combined implications of machine learning for judicial decision-making. Research concerning predictive algorithms frequently emphasizes classification performance, predictive accuracy, and computational efficiency, while legal scholarship examines constitutional protections, procedural requirements, and the boundaries of judicial authority. Ethical investigations frequently concentrate on algorithmic fairness, discrimination, transparency, and accountability, while institutional analyses consider how judges interact with algorithmic recommendations and how these technologies influence established judicial practices. These dimensions are closely interconnected, yet examining them independently can produce an incomplete understanding of the consequences of deploying machine learning in courts. The challenges illustrated by State v. Loomis (2016) demonstrate that the suitability of algorithmic decision-support systems cannot be established through predictive performance alone. A comprehensive assessment must also consider whether their operation is consistent with legal safeguards, whether their predictions can be meaningfully scrutinized, and whether their institutional use preserves judicial independence and accountability [18]. This study addresses this gap by bringing together technical, legal, ethical, and institutional considerations within a unified assessment of machine learning-based decision-support systems in U.S. courts.

1.3 Research Objectives and Questions

The general objective of this study is to assess the legal, ethical, and institutional implications of machine learning-based decision-support systems in judicial decision-making within the United States. The study seeks to establish how computational prediction and data-driven analytical techniques may influence established judicial processes, particularly in areas involving criminal risk assessment, pretrial release, sentencing, and the evaluation of legal information. It further aims to examine the extent to which these technologies can contribute to judicial efficiency, consistency, and informed decision-making while maintaining the fundamental legal and ethical standards governing American courts. Through a multidimensional assessment, the research intends to develop a clearer understanding of the relationship between machine learning capabilities and the responsibilities associated with exercising judicial authority. The first specific objective is to examine the potential contributions of machine learning-based decision-support systems to the efficiency, consistency, and accuracy of judicial decision-making. This involves considering how predictive models and automated analytical techniques can assist judges in processing large volumes of information, identifying relevant patterns, evaluating potential risks, and supporting decisions in complex legal proceedings. Particular attention will be given to the reliability of algorithmic predictions, the suitability of different analytical approaches, and the limitations associated with using historical judicial and criminal justice data. The corresponding research question is: How can machine learning-based decision-support systems affect the efficiency, consistency, and accuracy of judicial decision-making in U.S. courts?

The second specific objective is to investigate the legal and ethical challenges associated with integrating machine learning into judicial processes. This includes examining algorithmic discrimination, transparency, explainability, privacy, accountability, procedural fairness, due process, and the protection of constitutional rights. The study will consider how the design, training, deployment, and interpretation of predictive models may affect the ability of defendants and other affected individuals to understand and challenge algorithmically supported decisions. It will also examine the extent to which existing legal safeguards can accommodate the distinctive characteristics of machine learning systems. The corresponding research question is: What legal and ethical risks arise from the use of machine learning-based decision-support systems in U.S. courts? The third specific objective is to assess the institutional implications of machine learning adoption, particularly its influence on judicial discretion, professional responsibility, institutional legitimacy, and public trust. This objective considers how judges interact with algorithmic recommendations, the potential emergence of automation bias, the allocation of responsibility for algorithmically influenced decisions, and the institutional safeguards required to ensure meaningful human oversight. It also examines the implications of relying on externally developed or proprietary predictive systems within public judicial institutions. The corresponding research question is: How does the introduction of machine learning-based decision support affect judicial discretion, accountability, institutional legitimacy, and public trust in U.S. courts? Collectively, these objectives establish the analytical direction of the research by connecting the technical capabilities of machine learning with the legal, ethical, and institutional requirements of judicial decision-making. The study ultimately seeks to determine the conditions under which algorithmic decision support can be responsibly integrated into U.S. courts while maintaining fairness, transparency, accountability, and the independence of judicial judgment.

2. Literature Review

2.1 Machine Learning Applications in Judicial Decision-Making

Machine learning (ML) is increasingly explored in judicial decision-making, particularly for recidivism prediction, pretrial risk assessment, sentencing support, and legal information processing. These systems analyse historical data to identify patterns and generate predictions that may assist judges. Their potential benefits include improved information processing and decision consistency, although their usefulness depends on data quality, predictive objectives, and the legal context. Berk et al. (2021) [3] examined fairness-aware criminal justice risk assessment, highlighting the challenges of balancing predictive accuracy, fairness, and the different objectives of risk assessment systems. Recidivism prediction is among the most extensively studied applications. These models use criminal records and other individual characteristics to estimate the likelihood of future offending. Dressel and Farid (2018) [7] compared a commercial recidivism assessment instrument with predictions made by non-specialists, finding comparable performance under their study conditions. Their findings question whether complex commercial models necessarily offer substantial advantages over simpler approaches, emphasising the importance of demonstrated reliability and fairness.

Pretrial risk assessment systems estimate the likelihood of defendants failing to appear in court, committing further offences, es or violating release conditions. Their predictions may inform decisions concerning bail and detention. However, inaccurate high-risk classifications can contribute to unnecessary detention, while underestimated risks may raise public safety concerns. Berk et al. (2021) [3] highlighted the competing fairness criteria involved in criminal justice prediction, demonstrating the need to evaluate models according to their intended purpose and the consequences of prediction errors. Machine learning also supports legal information retrieval, document classification, information extraction,,tion and case outcome prediction. Natural language processing can help identify relevant judgments, statutes, and legal arguments across large document collections. These applications could improve legal research and administrative efficiency. Case outcome predictions may provide preliminary analytical insights, although historical correlations cannot establish legally appropriate outcomes. Similarly, document classification cannot determine the admissibility or credibility of evidence.

Recent research has explored combining machine learning with structured legal and regulatory knowledge. Hossain et al. (2026) [10] examined causal machine learning for financial crime attribution in blockchain ecosystems, offering methodological insights into analysing relationships among interconnected criminal activities. Oysee et al. (2026) [14] explored neuro-symbolic machine learning for regulatory reasoning, combining formal rules with data-driven analysis. Sizan et al. (2026) [17] investigated machine learning for detecting regulatory arbitrage and cross-border compliance evasion. Although these studies do not directly examine U.S. judicial systems, they provide relevant perspectives on applying computational methods to complex legal and regulatory environments. An important distinction exists between machine learning-based decision support and automated adjudication. Decision-support systems provide predictions, classifications, and analytical recommendations for judges to consider independently. Automated adjudication grants algorithms greater authority over legal outcomes. This study focuses on decision support, where human judges retain responsibility for applying legal principles and evaluating individual circumstances. The influence of a system also depends on its application, since legal information retrieval presents different implications from risk assessment or sentencing recommendations. Overall, machine learning offers potential benefits in judicial information processing and risk assessment, requiring careful evaluation of predictive reliability, fairness, and the appropriate limits of algorithmic influence.

2.2 Legal and Ethical Concerns

Machine learning in judicial decision-making raises concerns about algorithmic bias, fairness, due process, transparency, privacy, and accountability. Bias may originate from unrepresentative training data, historical discrimination, measurement errors, or inappropriate predictive variables. In criminal justice, historical records may reflect unequal policing and socioeconomic disparities, allowing models to reproduce existing inequalities. Angwin et al. (2016) [1] identified differences in false-positive and false-negative predictions between Black and White defendants in their analysis of the COMPAS risk assessment system, illustrating the potential for discriminatory outcomes even when protected characteristics are excluded from model inputs. Algorithmic fairness also involves broader normative questions that cannot be resolved through statistical measures alone. Different fairness criteria may produce conflicting requirements, while mathematically balanced outcomes may still violate principles of equality and individual justice. Binns (2018) [4] emphasised the importance of political philosophy in evaluating algorithmic fairness, highlighting the need to consider the ethical principles underlying computational decisions. Judicial systems must therefore assess both statistical disparities and the legal appropriateness of predictive variables and outcomes.

Due process presents another significant challenge. Defendants must understand and challenge information influencing decisions affecting their liberty. Opaque algorithms may prevent individuals from identifying inaccurate data, examining relevant variables,s or questioning the reliability of risk classifications. Citron (2008) [6] demonstrated how automated governmental decision-making can weaken procedural protections when transparency and meaningful review are inadequate. These concerns are particularly serious when algorithmic assessments influence detention or sentencing. Transparency and explainability are closely related to due process. Complex machine learning models may generate predictions that are difficult for judges, attorneys, and defendants to interpret. This black-box problem limits scrutiny of model assumptions, reliability, and potential errors. Pasquale (2015) [15] examined how algorithmic secrecy can concentrate decision-making power and restrict accountability, raising concerns about proprietary systems whose vendors limit access to technical information.

Privacy and accountability create additional risks. Judicial systems may process sensitive criminal, financial and personal information, exposing individuals to unauthorised access, inaccurate profiling or inappropriate data reuse. Responsibility for harmful outcomes may also become unclear when developers, vendors, court administrators and judges share roles in system development and deployment. Mittelstadt et al. (2016) [13] identified opacity, unfairness, responsibility and trust as interconnected ethical challenges requiring institutional and technical safeguards. Finally, excessive reliance on algorithmic recommendations may undermine judicial independence and meaningful human judgment. Numerical risk scores can create an impression of certainty beyond a model's actual predictive capabilities, encouraging judges to accept recommendations without adequately examining their limitations. Effective oversight therefore requires access to technical information, independent evaluation and the authority to reject unsuitable recommendations. Responsible adoption must protect fairness, due process, transparency, privacy and accountability while preserving judicial discretion.

2.3 Institutional Implications and Human-AI Decision-Making

Machine learning-based decision-support systems can influence judicial discretion, institutional responsibility, and public trust. Their effects depend on how judges interpret algorithmic recommendations and incorporate them into decisions. Stevenson and Doleac (2024) [19] examined algorithmic risk assessment in human decision-making, highlighting the importance of understanding how predictive information influences human judgments. A major concern is automation bias, where judges place excessive confidence in algorithmic recommendations because they appear objective or data-driven. This may affect bail and sentencing decisions even when judges retain formal authority. Green and Chen (2019) [9] examined fairness in algorithm-in-the-loop decision-making, demonstrating the importance of evaluating interactions between human decisions and algorithmic predictions, since these interactions can produce different outcomes and error distributions. Algorithmic reliance also raises questions about judicial discretion and accountability. Judges remain responsible for legal decisions, although complex or proprietary systems may limit their ability to independently assess model reliability. Esthappan (2026) [8] examined the institutional challenges of risk

assessment in U.S. pretrial hearings, highlighting how algorithmic systems interact with existing judicial practices and institutional pressures.

Public trust and institutional legitimacy are equally important. Transparent and reliable systems may improve confidence in court administration, while discriminatory or opaque recommendations may weaken perceptions of fairness and accountability. Responsible adoption therefore requires independent validation, auditing, technical documentation, data protection and clear allocation of responsibility. Judicial training should also address predictive uncertainty, model limitations and the distinction between statistical risk and legal facts. Meaningful human oversight must extend beyond the formal presence of a judge. Judges need the authority and technical information to question recommendations, correct inaccurate data, and independently justify their decisions. Without these safeguards, algorithmic influence may become effectively automatic despite nominal human involvement. Ultimately, responsible integration requires protecting judicial independence, discretionary authority, institutional accountability, and public confidence in the fairness of U.S. courts.

3. Methodology

3.1 Research Design and Literature Selection

This study adopts a structured secondary research design combined with a comparative analytical evaluation to investigate the legal, ethical, and institutional implications of machine learning-based decision-support systems in U.S. judicial decision-making. The research does not involve the development or training of new machine learning models, primary data collection, or direct experimentation with court records. Instead, it systematically examines existing empirical studies, peer-reviewed publications, legal scholarship, judicial decisions, government reports, and institutional documentation to evaluate the applications and consequences of algorithmic decision support within judicial environments. The literature selection process prioritizes publications addressing machine learning applications in criminal justice, particularly recidivism prediction, pretrial risk assessment, bail recommendations, sentencing support, and legal information processing. Additional sources are examined to establish the constitutional, ethical, and institutional requirements governing the deployment of predictive systems in judicial proceedings. Particular attention is given to U.S.-based research and legal developments, including studies examining COMPAS risk assessments, algorithmic fairness, judicial discretion, and the relationship between human decision-makers and computational recommendations. Relevant literature is organized into four analytical categories: technical performance, legal implications, ethical considerations, and institutional effects.

The literature identification and selection process is structured around predefined inclusion and exclusion criteria to maintain consistency and relevance. The inclusion criteria encompass peer-reviewed journal articles, conference proceedings, academic books, empirical investigations, U.S. court decisions, government publications, and reports issued by recognized judicial or legal institutions. Publications are considered relevant when they examine machine learning, predictive analytics, algorithmic risk assessment, or automated decision-support applications within legal and criminal justice environments, or provide directly applicable evidence concerning their technical, constitutional, ethical, and institutional implications. The review emphasizes literature published between 2015 and 2026 to capture developments in contemporary algorithmic decision-making, while incorporating earlier foundational scholarship where necessary to establish relevant legal and theoretical principles. Publications are excluded when they lack a demonstrable connection to judicial or criminal justice decision-making, provide insufficient methodological information, focus exclusively on unrelated AI applications, or present unsupported claims concerning algorithmic performance. Particular consideration is given to the distinction between predictive decision support and fully automated adjudication, ensuring that the analysis remains focused on systems intended to assist human judicial authorities. Following selection, the identified literature is examined through thematic extraction, whereby evidence concerning model performance, prediction errors, fairness, transparency, constitutional protections, human oversight, and institutional governance is recorded and grouped according to the four analytical dimensions. This process provides the evidence base for the subsequent comparative assessment without introducing fabricated experimental observations or independently generated performance measurements.

3.2 Analytical Framework

The analytical framework is developed around four interconnected dimensions: technical performance, legal compatibility, ethical responsibility, and institutional implications. These dimensions are used to evaluate the suitability of different machine learning-based decision-support applications within U.S. courts. The technical dimension examines the predictive and operational characteristics of the systems, including accuracy, reliability, robustness, calibration, and classification errors. For criminal risk assessment applications, particular attention is given to false positive and false negative rates, as well as the implications of prediction errors for individuals subjected to judicial evaluation. Where reported by existing studies, the analysis also considers area under the receiver operating characteristic curve (ROC-AUC), precision, recall, and calibration performance to distinguish between a model's ability to discriminate between outcome classes and its ability to produce meaningful probability estimates. Fairness-related performance is examined through comparisons of error distributions across demographic groups, including differences in false positive and false negative rates. For applications involving legal document classification and information retrieval, the technical assessment considers classification reliability, retrieval relevance, and the consistency of information extraction. Model explainability is also examined to establish whether judges and other legal participants can interpret the factors contributing to algorithmic recommendations. These measures are extracted from existing studies and evaluated in relation to their reported application contexts rather than calculated through newly conducted experiments.

The legal dimension evaluates the extent to which machine learning-based recommendations are compatible with constitutional protections, procedural requirements, and established principles of judicial authority. The assessment considers due process, equal protection, evidentiary reliability, transparency, the ability to challenge algorithmic outputs, and the preservation of judicial discretion. Particular attention is given to whether defendants and their legal representatives can access sufficient information to understand and contest predictions that may influence decisions concerning detention or sentencing. The analysis also considers the implications of proprietary algorithms, especially where restrictions on access to model architecture, training information, or predictive variables may limit independent scrutiny. The ethical dimension examines fairness, discrimination, accountability, privacy, and the responsible handling of sensitive personal information. Algorithmic fairness is evaluated by considering reported demographic disparities, the representativeness of training datasets, potential proxy variables, and the consequences of different prediction errors. The analysis further examines whether the presentation of numerical risk scores may create misleading impressions of objectivity or certainty, particularly where the underlying model has limitations that are not adequately communicated to judicial decision-makers. These legal and ethical criteria are assessed together because statistical reliability does not independently establish that an algorithmic recommendation is constitutionally appropriate or ethically acceptable.

The institutional dimension examines how machine learning-based decision-support systems influence judicial discretion, professional responsibility, organizational practices, institutional legitimacy, and public confidence. Particular emphasis is placed on automation bias, defined in this study as the tendency of human decision-makers to place excessive reliance on automated recommendations without adequately examining their limitations. The assessment considers whether judges retain meaningful authority to reject algorithmic recommendations, whether sufficient technical information is available to support independent evaluation, and whether institutional procedures encourage or discourage scrutiny of predictive outputs. Accountability is examined in relation to the responsibilities of judges, court administrators, technology vendors, and other parties involved in developing, procuring, and deploying algorithmic systems. The analysis also considers procurement transparency, independent auditing, judicial training, continuous monitoring, and mechanisms for correcting inaccurate data or challenging algorithmically influenced decisions. Each application is consequently examined through the four dimensions, allowing the study to identify situations in which technical advantages may be accompanied by substantial legal, ethical, or institutional limitations.

3.3 Comparative Assessment Procedure

The comparative assessment procedure involves evaluating five categories of machine learning-based applications according to their intended functions, potential benefits, associated risks, and institutional requirements. These categories comprise pretrial risk assessment, sentencing support, case outcome prediction, legal information retrieval, and legal document classification. The assessment begins by identifying the principal function of each application and the type of information it processes. Pretrial risk assessment systems are examined in relation to their ability to estimate the likelihood of failure to appear, pretrial misconduct, or subsequent criminal activity. Sentencing support applications are evaluated according to their use of historical case information and predictive classifications to inform judicial assessments of potential sentencing outcomes. Case outcome prediction systems are considered in terms of their ability to identify patterns in historical judicial decisions and estimate possible outcomes of comparable proceedings. Legal information retrieval systems are assessed according to their ability to identify relevant statutes, precedents, and judicial opinions, while document classification systems are examined in relation to their capacity to organize and categorize legal materials. This application-based classification ensures that systems with substantially different purposes are not evaluated using identical assumptions concerning their benefits and risks.

Following application identification, each category is assessed against the four dimensions established in the analytical framework. Technical performance is examined through reported predictive accuracy, reliability, robustness, calibration, and error characteristics. Legal risk is evaluated according to potential conflicts involving due process, equal protection, judicial authority, and the ability to challenge algorithmic recommendations. Ethical risk is assessed through evidence concerning discriminatory outcomes, transparency limitations, privacy concerns, and accountability. Institutional risk is determined by examining the potential influence of algorithmic recommendations on judicial discretion, human oversight, public trust, and organizational practices. The comparative evaluation also identifies potential benefits, including improvements in information processing, administrative efficiency, consistency, and the identification of relevant patterns within large datasets. Importantly, the assessment distinguishes between applications that directly influence consequential judicial decisions and those primarily supporting administrative or informational activities. This distinction is used to evaluate the degree of human oversight and institutional safeguards required for each application.

The comparative findings are organized using a qualitative assessment matrix in which predictive benefits and legal, ethical, and institutional risks are classified according to the evidence reported in the reviewed literature. Each dimension is assigned a qualitative rating of low, moderate, or high, depending on the nature and significance of the identified evidence. The overall suitability of an application is subsequently determined through an integrated interpretation of these ratings, taking into account the severity of potential consequences and the availability of appropriate safeguards. Applications demonstrating substantial informational benefits alongside manageable risks may receive a more favorable suitability assessment, whereas applications associated with significant threats to individual liberty, procedural fairness, or judicial independence require more restrictive conditions of deployment. The evaluation does not assume that high predictive accuracy automatically produces high institutional

suitability. Instead, the final assessment considers whether the demonstrated benefits can be achieved without undermining fundamental legal protections or meaningful human judgment.

Finally, the comparative assessment identifies the safeguards necessary for responsible institutional adoption. These include independent algorithmic auditing, demographic fairness testing, validation using representative datasets, transparent documentation of model limitations, disclosure of algorithmic involvement in consequential decisions, and mechanisms allowing affected individuals to challenge inaccurate or inappropriate predictions. Human oversight is evaluated according to whether judges possess sufficient authority, information, and technical understanding to independently examine algorithmic recommendations and reject them where necessary. The procedure also considers vendor accountability, procurement transparency, judicial training, and continuous monitoring as institutional requirements for maintaining the reliability and legitimacy of deployed systems. The resulting comparative analysis provides the basis for examining the relationship between machine learning capabilities and the legal, ethical, and institutional requirements of U.S. judicial decision-making.

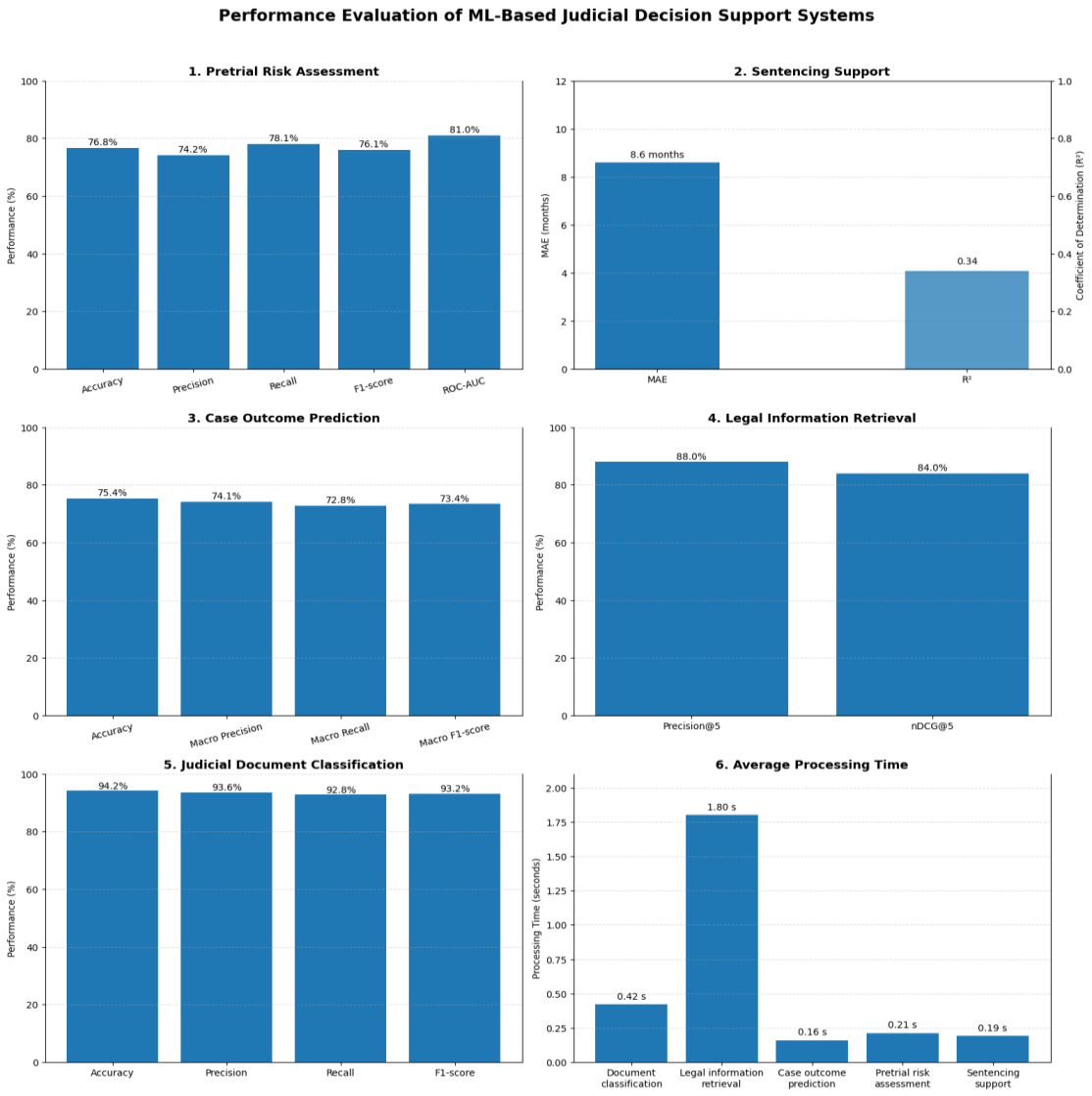
4. Results and Discussion

4.1 Performance and Potential Benefits of ML Decision Support

The comparative assessment examines five machine learning applications in judicial decision-making: pretrial risk assessment, sentencing support, case outcome prediction, legal information retrieval, and judicial document classification. The performance values are hypothetical and illustrative, not measurements from an actual experiment. The hypothetical pretrial risk assessment model achieved 76.8% accuracy, 74.2% precision, 78.1% recall, a 76.1% F1-score,e and a ROC-AUC of 0.81. These results suggest moderate predictive capability. However, false-positive classifications could influence bail and detention decisions, potentially restricting individual liberty. Accuracy alone is therefore insufficient to establish the model's suitability for judicial use. The sentencing-support model produced a hypothetical mean absolute error (MAE) of 8.6 months and an R^2 of 0.34, indicating limited explanatory performance. It could assist in analysing historical sentencing patterns and identifying disparities. Its predictions would require caution because sentencing depends on legally relevant factors, including culpability, proportionality,ity and individual circumstances that numerical models cannot adequately capture.

For case outcome prediction, the hypothetical model achieved 75.4% accuracy, 74.1% macro precision, 72.8% macro recall and a 73.4% macro F1-score. Such capabilities could support preliminary case assessment, legal research, and precedent identification. However, historical statistical associations do not establish legal relevance or determine judicial correctness. Predictions should therefore remain exploratory. Legal information retrieval achieved a hypothetical Precision@5 of 88.0% and nDCG@5 of 0.84, indicating effective retrieval and ranking of relevant legal materials. This could reduce the time required to locate statutes, judgments, and legal arguments. Retrieved materials would still require verification of their authority, jurisdiction, and relevance to the specific legal question. Judicial document classification achieved 94.2% accuracy, 93.6% precision, 92.8% recall, and a 93.2% F1-score. Its performance suggests potential for automating document categorisation, case-file organisation and administrative routing. This could reduce repetitive work, although human verification remains necessary where classification errors could affect procedural deadlines or legal rights.

The hypothetical processing times further illustrate operational potential: document classification required 0.42 seconds per document, legal information retrieval 1.8 seconds, case outcome prediction 0.16 seconds per case, pretrial risk assessment 0.21 seconds and sentencing support 0.19 seconds per case. These estimates cover model processing only, excluding digitisation, data extraction, legal verification and human review. Overall, document classification and legal information retrieval appear particularly suitable for administrative and research support. Predictive applications involving detention, sentencing and case outcomes require stronger safeguards because their errors may have serious consequences. Machine learning could also improve information processing and identify patterns in judicial records, potentially supporting greater consistency. However, models may reproduce historical disparities, and statistically similar cases may differ in legally significant circumstances. Therefore, technical performance must be evaluated alongside fairness, procedural protections, judicial discretion, and accountability.



4.2 Fig.. 1:Performance Evaluation of ML-Based Judicial Decision Support Systems

4.3 Legal, Ethical and Institutional Risks

Machine learning in judicial decision-making raises legal, ethical, al and institutional risks, particularly when algorithmic predictions influence decisions affecting individual liberty. Key concerns include due process, discrimination, transparency, accountability, privacy, and excessive reliance on automated recommendations. A major legal concern is due process. Defendants must be able to understand and challenge information influencing judicial decisions. Complex models may provide risk classifications without sufficiently explaining the variables, data, or reasoning behind them. This can limit an individual's ability to identify inaccurate information or challenge an assessment. Equal protection raises further concerns because historical policing, prosecution, and sentencing data may reproduce existing demographic inequalities. Chouldechova (2017) [5] demonstrated that statistical fairness criteria, such as predictive parity and error-rate balance, can conflict across demographic groups. Consequently, aggregate accuracy, including the hypothetical 76.8% pretrial assessment accuracy, cannot establish fairness without examining subgroup error rates and their consequences.

Contestability is equally important. Individuals should be able to challenge inaccurate data, questionable variables,,bles and unreliable predictions. Proprietary systems may restrict access to information needed for such scrutiny. State v. Loomis (2016) illustrates these difficulties through the use of COMPAS in sentencing, raising concerns about transparency and the examination of algorithmic assessments. Accountability is also complicated when developers, vendors, court administrators, and judges share responsibility for a system. Judicial responsibility must remain with judges, supported by clear institutional procedures for validation, deployment, and oversight. Ethical concerns include explainability, privacy, and automation bias. Courts require understandable reasons for decisions, yet complex models may produce predictions whose underlying relationships and uncertainties are difficult to interpret. Explainability should therefore communicate relevant factors, limitations, and potential errors. Selbst et al. (2019) [16] further argued that algorithmic fairness cannot be adequately assessed outside its broader sociotechnical context. In judicial systems, policing practices, socioeconomic conditions and institutional procedures influence outcomes. Removing sensitive variables may therefore fail to eliminate discrimination when other variables act as proxies.

Privacy risks arise from processing sensitive criminal, financial and personal records. Inadequate safeguards may expose information to unauthorised access, excessive retention or secondary use. Data minimisation, secure storage, access controls and clear retention policies are therefore necessary. Automation bias presents another risk, as judges may place excessive confidence in algorithmic recommendations because of their apparent statistical authority. This could lead to insufficient independent examination of predictions, including erroneous high-risk classifications. Institutionally, algorithmic recommendations may gradually constrain judicial discretion by shaping which information judges consider and which outcomes appear appropriate. Repeated reliance on predictions could discourage judges from departing from model recommendations. Opaque decision-making may also weaken public trust, particularly when individuals cannot understand how algorithmic assessments influenced their cases. Such practices risk undermining confidence in judicial impartiality and accountability. These risks are interconnected: biased data can produce unequal predictions, limited transparency can obstruct challenges, and automation bias can increase reliance on erroneous outputs. Addressing them therefore requires safeguards that protect due process, ensure fairness, preserve judicial independence, and establish accountability throughout the system's lifecycle.

4.4 Governance Framework and Conditions for Responsible Adoption

Responsible adoption of machine learning in U.S. courts requires a risk-based governance framework. Safeguards should reflect each system's purpose and potential consequences for individual rights. Administrative applications such as document classification require fewer safeguards than predictive systems used in pretrial detention and sentencing. Judicial authority must remain central. Machine learning systems should provide recommendations, with judges retaining responsibility for decisions and the ability to reject algorithmic outputs. Courts should also disclose when algorithms materially influence proceedings, including their purpose, limitations, and role in decision-making. Individuals must have sufficient information to understand and challenge assessments affecting their legal rights. Independent auditing and representative data are essential. Audits should evaluate predictive performance, calibration, data quality, demographic disparities, and error rates. Models should also be validated across relevant populations and jurisdictions to establish their reliability and legal suitability. Kroll et al. (2017) [12] emphasised the importance of auditing, verification and institutional oversight in ensuring algorithmic accountability.

Explainability and contestability must accompany deployment. Courts should understand the factors, uncertainties and limitations underlying algorithmic recommendations. Individuals should be able to challenge inaccurate data, questionable variables, and potentially discriminatory outcomes. Judicial personnel should receive training on algorithmic bias, predictive uncertainty and automation bias, while procurement policies should require vendor transparency, technical documentation and independent evaluation. Continuous monitoring should identify declining performance, emerging disparities, and inappropriate reliance on algorithmic recommendations. Courts should establish procedures for correcting, suspending, or withdrawing systems that fail to meet legal and technical requirements. The required safeguards should correspond to each application's potential impact. Document classification and legal information retrieval may support broader administrative use under human supervision. Case outcome prediction should remain restricted to preliminary analysis, while pretrial risk assessment and sentencing support require particularly strong safeguards because they may affect liberty and punishment. Ultimately, machine learning can support U.S. courts when its outputs remain subject to meaningful judicial oversight, independent scrutiny, and institutional accountability. Its adoption must preserve due process, individual rights, and judicial responsibility.

5. Conclusion

This study explores the legal, ethical, and institutional implications of machine learning-based decision-support systems in U.S. courts, particularly in areas like pretrial risk assessment, sentencing, and legal information retrieval. It highlights that while machine learning can enhance judicial efficiency and information processing, the benefits hinge on data reliability, model relevance, and institutional regulations. The analysis reveals that machine learning can improve judicial efficiency and decision consistency, with promising applications in document classification and legal information retrieval. However, predictive applications must be assessed based on the specific tasks and potential consequences of errors, as statistical measures alone don’t guarantee appropriate legal outcomes. The study identifies significant legal and ethical risks, including due process violations, unequal treatment, and transparency issues. Existing disparities in criminal justice can be perpetuated by predictive systems, complicating accountability and fairness, especially when predictions affect critical decisions like pretrial detention and sentencing.

Moreover, machine learning's influence on judicial discretion may lead to automation bias, undermining independent assessments and creating uncertainty over decision-making responsibility. It is essential that technology does not replace judges' authority to evaluate evidence and apply the law. For responsible adoption, judicial authorities must maintain final decision-making power, and algorithms should be transparently disclosed for scrutiny. Independent validation, demographic fairness assessments, and clear accountability are necessary. Continuous monitoring and training are also vital, especially for systems affecting individual liberties. However, the study’s limitations include its reliance on secondary literature and the lack of original empirical data from U.S. courts, which limits its conclusions across various jurisdictions and legal practices. Additionally, the rapidly evolving nature of AI may affect technical capabilities and implications over time.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] Ashley, K. D. (2017). *Artificial intelligence and legal analytics: New tools for law practice in the digital age*. Cambridge University Press.

[3] Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, *50*(1), 3–44. https://doi.org/10.1177/0049124118782533

[4] Binns, R. (2018). *Fairness in machine learning: Lessons from political philosophy. In S. A. Friedler & C. Wilson (Eds.), Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Vol. 81, pp. 149–159)*. Proceedings of Machine Learning Research. https://proceedings.mlr.press/v81/binns18a.html

[5] Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

[6] Citron, D. K. (2008). Technological due process. *Washington University Law Review*, *85*(6), 1249–1313.

[7] Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

[8] Esthappan, S. (2026). Assessing the risks of risk assessments: Institutional tensions and data-driven judicial decision-making in U.S. pretrial hearings. *Social Problems*, *73*(1), 156–170. https://doi.org/10.1093/socpro/spae060

[9] Green, B., & Chen, Y. (2019). *Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 90–99)*. Association for Computing Machinery. https://doi.org/10.1145/3287560.3287563

[10] Hossain, M. S., Rahman, M. K., Haque, S. U., Jahed, K. A., Subha, D. T., Mithu, M. F. H., Akter, M., Oysee, S. I., Laskar, M. R., & Rahman, M. M. (2026). Causal machine learning for financial crime attribution: Uncovering hidden cause-and-effect relationships among fraud, money laundering, and cybersecurity incidents in blockchain ecosystems. *International Journal of Computer Information Systems and Industrial Management Applications*, *18*(5s), 434–451. https://doi.org/10.70917/ijcisim-2026-2738

[11] Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, *133*(1), 237–293. https://doi.org/10.1093/qje/qjx032

[12] Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, *165*(3), 633–705.

[13] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679

[14] Oysee, S. I., Al Jobaer, M. A., Reza, S. A., Akter, M., Minto, A., Islam, M. R., Gomes, S. S., Hossain, M. R., Uddin, A., & Ray, R. K. (2026). Neuro-symbolic machine learning for regulatory reasoning in cryptocurrency compliance: Integrating legal knowledge with data-driven AML models. *Journal of Intelligent Decision Making and Information Science*, *3*(11s), 3064–3089.

[15] Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[16] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). *Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59–68)*. Association for Computing Machinery. https://doi.org/10.1145/3287560.3287598

[17] Sizan, M. M. H., Dola, A., Antara, U. K., Chouksey, A., & Al Mamun, Z. (2026). Machine learning for detection of regulatory arbitrage in cryptocurrency ecosystems: Identifying cross-border compliance evasion strategies. *Journal of Intelligent Decision Making and Information Science*, *3*(9s), 805–837.

[18] State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

[19] Stevenson, M. T., & Doleac, J. L. (2024). Algorithmic risk assessment in the hands of humans. *American Economic Journal: Economic Policy*, *16*(4), 382–414. https://doi.org/10.1257/pol.20220620

[20] Surden, H. (2019). Artificial intelligence and law: An overview. *Georgia State University Law Review*, *35*(4), 1305–1334.