**Indonesian Journal *of* Digital Crime *and* Criminal Justice**
Vol. 2 No. 1. January 2026. pp. 49-78
E-ISSN: XXXX-XXXX| P-ISSN: XXXX-XXX
**Original Research Article**

# Artificial Intelligence in Sentencing Decisions: Legal, Ethical, and Human Rights Considerations

**Wan Ahmad Syarifuddin** [1]*, **Adi Baskara** [2]
[1] Singapore Management University, Singapore
[2] Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
* Corresponding author: a.syarifuddin@smu.edu.sg

## ABSTRACT

The use of artificial intelligence (AI) in sentencing decisions is increasingly discussed as part of modern criminal justice reform. In Indonesia and other jurisdictions, AI-based tools are proposed to assist judges in predicting recidivism, assessing risk, and improving sentencing consistency. However, such developments raise significant legal, ethical, and human rights concerns. This article examines the implications of AI-assisted sentencing within the Indonesian legal system, focusing on compatibility with principles of judicial independence, fairness, and due process. Using normative legal research and comparative analysis, the study evaluates the risks of algorithmic bias, lack of transparency, and over-reliance on automated decision-making. The findings suggest that AI systems may reinforce existing social inequalities embedded in training data, potentially undermining individualized justice. Furthermore, the absence of clear regulatory standards in Indonesia increases uncertainty regarding the admissibility and legitimacy of AI-generated recommendations in sentencing. The article argues that while AI may enhance efficiency, sentencing authority must remain fundamentally human-centered. It proposes strict regulatory safeguards, including explainability requirements, judicial oversight, and ethical guidelines for AI use in courts. The study concludes that AI in sentencing should be limited to advisory functions to preserve constitutional and human rights principles.

**KEYWORDS:** Artificial Intelligence, Sentencing, Criminal Justice, Indonesia, Human Rights, Judicial Independence, Algorithmic Bias

Available online at https://journal.criminallawinstitute.org/JDCCJ/article/view/22, since January 15, 2026
Submitted: October 21, 2025 ■ Reviewed: November 7, 2025 ■ Revised: December 9, 2025; January 8, 2026 ■ Accepted: January 10, 2026

## INTRODUCTION

The integration of artificial intelligence (AI) into public governance represents one of the most significant transformations in contemporary legal and institutional systems. Rather than functioning as a purely technical innovation, AI is increasingly embedded within broader structures of algorithmic governance, where computational systems influence, support, and in some cases partially replace human decision-making processes in the public sector (Kitchin 2017; Mittelstadt et al. 2016). Within criminal justice systems, this transformation is reflected in the adoption of data-driven tools for predictive policing, risk assessment, pretrial evaluation, sentencing guidance, and correctional management (Brayne 2020; Hao 2019).

Among these applications, the use of AI in sentencing decisions has generated particular academic and policy attention. Courts in several jurisdictions have experimented with algorithmic risk assessment tools designed to assist judges in evaluating offender risk profiles, recidivism probabilities, and sentencing recommendations. Advocates argue that such systems may enhance consistency, reduce sentencing disparities, and improve judicial efficiency by providing data-driven insights to support judicial reasoning (Dressel and Farid 2018). From this perspective, AI is framed as a tool for strengthening rationality and uniformity in sentencing practices, particularly in jurisdictions facing heavy caseloads and resource constraints.

However, the increasing reliance on algorithmic systems in judicial contexts raises fundamental concerns regarding legality, ethics, and human rights. Empirical studies demonstrate that algorithmic tools may reproduce structural biases embedded in training data, leading to discriminatory outcomes in criminal justice decision-making (Angwin et al. 2016; Benjamin 2019). Furthermore, scholars highlight the risk that opaque computational processes may undermine judicial transparency and weaken accountability in sentencing decisions (Burrell 2016; Selbst and Barocas 2018). These concerns become particularly significant when AI systems influence decisions that directly affect individual liberty.

In developing countries such as Indonesia, these debates are especially relevant in light of ongoing judicial modernization and digital transformation initiatives. While technological adoption is increasingly promoted as a means of improving efficiency and reducing procedural delays, the absence of clear regulatory frameworks governing judicial AI creates uncertainty regarding the legitimacy, accountability, and constitutional compatibility of algorithmic sentencing tools.

AI-assisted sentencing systems are often justified on the basis of improving consistency, efficiency, and evidence-based judicial decision-making. Nevertheless, significant legal and ethical concerns persist regarding the reliability and fairness of such systems. Studies show that algorithmic models may embed racial, socioeconomic, or institutional biases present in historical datasets, thereby producing unequal outcomes in sentencing processes (Benjamin 2019; Angwin et al. 2016).

In addition, the "black box" nature of many AI systems raises serious concerns regarding transparency and explainability. When judicial decisions are influenced by systems that cannot be meaningfully interpreted by judges, defendants, or legal practitioners, the principles of due process and fair trial may be compromised (Burrell 2016). Furthermore, questions remain regarding accountability: when AI systems influence sentencing outcomes, it is often unclear whether responsibility lies with judges, developers, institutions, or the algorithm itself.

This study addresses four key questions: (1) what role AI plays in sentencing decisions within modern criminal justice systems; (2) what legal and ethical challenges arise from AI-assisted sentencing; (3) how AI systems affect judicial independence, due process, and human rights guarantees; and (4) what regulatory mechanisms are required to ensure responsible implementation, particularly in Indonesia. The objective of this study is to critically examine the role of artificial intelligence in sentencing decisions and to analyze its legal, ethical, and human rights implications. The study also aims to evaluate comparative international approaches to judicial AI governance and to formulate normative recommendations for regulatory development in Indonesia.

From a theoretical perspective, this study contributes to interdisciplinary scholarship on algorithmic governance, legal automation, and digital criminal justice (Kitchin 2017). It also expands human-centered approaches to AI regulation by engaging with questions of fairness, accountability, and institutional legitimacy. From a practical perspective, this research provides policy-relevant insights for judicial institutions, legislators, and regulators concerning the responsible integration of AI into sentencing processes. It also offers normative guidance for the development of regulatory frameworks that ensure compliance with constitutional principles and international human rights standards.

Artificial intelligence in legal decision-making is widely conceptualized as part of broader algorithmic governance systems that rely on computational models to structure public decision-making processes (Mittelstadt et al. 2016). In criminal justice, these systems are often justified as tools that enhance efficiency and reduce human error; however, they simultaneously raise concerns about bias, transparency, and accountability.

In sentencing theory, classical models emphasize retribution, deterrence, rehabilitation, and proportionality as normative foundations of punishment. However, the introduction of AI-based risk assessment tools introduces actuarial logic into sentencing processes, shifting the focus from individualized moral judgment toward statistical prediction (Harcourt 2015). This transformation has been described as a movement from "judgment-based justice" to "data-driven justice," which raises concerns regarding the erosion of judicial discretion.

From a governance perspective, scholars argue that algorithmic systems reshape power relations between the state and individuals by embedding decision-making authority within technical infrastructures that are often inaccessible to public scrutiny (Kitchin 2017). Burrell (2016) further emphasizes that algorithmic opacity creates structural barriers to accountability, particularly in high-stakes domains such as criminal justice.

Human rights scholarship similarly highlights that AI in sentencing may threaten fundamental guarantees of fair trial, equality before the law, and non-discrimination. Empirical research demonstrates that algorithmic risk

assessment tools may produce disparate impacts on minority and disadvantaged groups, thereby reinforcing existing social inequalities (Benjamin 2019; Angwin et al. 2016). These concerns are increasingly reflected in international regulatory discussions, including emerging AI governance frameworks in the European Union and OECD. Despite growing global literature, significant gaps remain in relation to developing countries, particularly regarding regulatory preparedness and institutional capacity to govern AI in judicial sentencing systems.

This study adopts a normative legal research methodology supported by doctrinal, comparative, and human rights-based approaches. The statutory approach is used to analyze relevant legal frameworks governing criminal justice and judicial decision-making. The comparative approach examines regulatory developments in jurisdictions that have adopted AI-assisted sentencing tools. The human rights approach evaluates compliance with fair trial guarantees and equality principles under international law. Primary legal materials include constitutional provisions, criminal statutes, procedural laws, judicial regulations, and international human rights instruments. Secondary materials consist of peer-reviewed journal articles, policy reports, and academic books on artificial intelligence, algorithmic governance, and criminal justice reform. The analysis is conducted through qualitative doctrinal interpretation combined with comparative legal reasoning to identify legal gaps, governance challenges, and normative implications of AI-assisted sentencing systems.

## ARTIFICIAL INTELLIGENCE AND SENTENCING DECISIONS

Artificial intelligence (AI) has emerged as a transformative force in contemporary criminal justice systems, particularly in the context of judicial decision-making processes. Rather than operating as a fully autonomous adjudicatory mechanism, AI in criminal justice is predominantly deployed as a set of computational tools designed to support human decision-making through machine learning, predictive analytics, and algorithmic modeling. Scholars conceptualize these developments as part of a broader shift toward

"data-driven justice," in which statistical inference increasingly informs institutional practices that were traditionally grounded in judicial discretion and legal reasoning (Kitchin 2017; Mittelstadt et al. 2016).

From a functional perspective, AI systems in criminal justice are primarily used to process large datasets in order to identify patterns, generate predictions, and produce decision-support outputs. Machine learning models, in particular, are capable of learning from historical judicial data to produce probabilistic assessments related to offender behavior, sentencing outcomes, and risk levels. Predictive analytics further extends this capacity by enabling institutions to forecast future criminal behavior or system-level trends, thereby influencing sentencing policy and judicial administration (Brayne 2020). In this context, AI does not replace judicial authority but reconfigures the informational environment within which sentencing decisions are made.

Within criminal justice systems, AI is commonly understood as a constellation of technologies that includes machine learning systems, predictive analytics, and decision-support algorithms. Machine learning systems operate by identifying statistical relationships within large datasets, enabling them to generate predictions or classifications based on historical patterns. In sentencing contexts, these systems may analyze variables such as prior convictions, demographic factors, and offense characteristics to estimate risk profiles.

Predictive analytics expands this logic by applying statistical modeling techniques to anticipate future outcomes, including recidivism risks and potential threats to public safety. Decision-support algorithms, in turn, translate these predictions into structured recommendations that may assist judges in determining appropriate sentencing ranges or supervisory measures. However, as scholars have emphasized, these systems are not neutral instruments; rather, they reflect the assumptions, biases, and limitations embedded in their training data and design architectures (O'Neil 2016; Benjamin 2019).

One of the most prominent applications of AI in sentencing is the use of risk assessment systems designed to evaluate the likelihood of recidivism.

These systems are increasingly used in pretrial and sentencing contexts to inform judicial decisions regarding detention, probation, or custodial sentences. In the United States, for example, tools such as COMPAS have been widely discussed in both academic and judicial debates due to concerns regarding transparency and potential racial bias (Angwin et al. 2016; Dressel and Farid 2018).

Another significant application involves sentencing recommendation systems, which aim to enhance consistency in judicial outcomes by providing guideline-based suggestions derived from historical sentencing data. Proponents argue that such systems may reduce unwarranted disparities and improve predictability in sentencing practices. However, critics caution that standardization through algorithmic tools may inadvertently constrain judicial discretion and reduce the individualized assessment required in proportional sentencing.

In addition, judicial analytics tools are increasingly employed to support data-driven sentencing by aggregating and analyzing large-scale judicial datasets. These tools provide insights into sentencing patterns, case durations, and institutional performance, thereby contributing to broader judicial efficiency reforms. Nevertheless, their increasing influence raises normative questions regarding the extent to which quantitative outputs should shape inherently qualitative judicial determinations.

At the global level, the use of AI in judicial systems reflects a broader trend toward digital transformation of justice institutions. Many jurisdictions have begun experimenting with algorithmic tools in sentencing and related judicial processes as part of wider e-justice reforms (Susskind 2019). This experimentation is particularly visible in countries pursuing "smart court" or "digital justice" initiatives, where AI is integrated into case management, legal research, and judicial decision-support systems.

Despite these developments, global scholarship highlights that the adoption of judicial AI remains uneven and highly contested. While some jurisdictions emphasize efficiency and innovation, others express caution due to concerns about fairness, accountability, and human rights compliance. As a result, the global trajectory of judicial AI is characterized not by uniform

adoption, but by a complex and evolving balance between technological innovation and normative constraints.

## POTENTIAL BENEFITS OF AI-ASSISTED SENTENCING

The integration of artificial intelligence (AI) into sentencing processes has been widely discussed in contemporary legal and governance scholarship as part of a broader shift toward data-driven judicial administration. Within this framework, AI is not conceptualized as a replacement for judicial reasoning, but rather as a decision-support mechanism capable of enhancing the quality, consistency, and efficiency of sentencing outcomes. Proponents of algorithmic governance argue that such technologies may contribute to more rationalized and standardized decision-making processes in criminal justice systems, particularly in jurisdictions facing high caseloads and institutional resource constraints (Susskind 2019; Yeung 2018).

From a theoretical standpoint, these claimed benefits are closely associated with the logic of computational governance, where large-scale data analysis and machine learning techniques are used to improve institutional performance and reduce human variability in decision-making. However, as scholars caution, these benefits must be understood within a broader socio-technical context in which algorithmic outputs are shaped by data quality, design assumptions, and institutional practices (Mittelstadt et al. 2016; Kitchin 2017).

One of the most frequently cited advantages of AI-assisted sentencing is its potential to enhance consistency in judicial outcomes. Sentencing disparities have long been identified as a structural concern in criminal justice systems, where similar cases may produce divergent outcomes due to variations in judicial discretion, institutional context, or local practice. AI systems, particularly those trained on large datasets of prior sentencing decisions, are often presented as tools capable of reducing such inconsistencies by identifying patterns and proposing standardized sentencing ranges.

In this sense, algorithmic systems may contribute to the development of

more uniform sentencing practices by minimizing subjective variation and promoting greater predictability in judicial outcomes. Nevertheless, scholars emphasize that consistency achieved through algorithmic standardization does not automatically equate to substantive justice, particularly when underlying datasets reflect historical inequalities or systemic biases (O'Neil 2016).

AI-assisted sentencing systems are also frequently justified on the basis of improving procedural efficiency within criminal justice institutions. By automating certain aspects of case analysis and information processing, AI tools may reduce the time required for judges and court administrators to review case files, evaluate prior records, and assess sentencing guidelines. This may be particularly significant in jurisdictions experiencing case backlogs and limited judicial capacity. From an institutional perspective, efficiency gains are often associated with the broader objectives of digital-era governance, which emphasize the modernization of public administration through technological innovation and process optimization (Dunleavy et al. 2006). In sentencing contexts, AI systems can accelerate information retrieval and synthesis, thereby enabling judges to focus more on legal reasoning and individualized assessment rather than administrative processing.

Another important benefit attributed to AI in sentencing is its capacity to support evidence-based decision-making. AI systems are designed to analyze large volumes of historical data, identify statistical patterns, and generate predictive insights that may assist judicial actors in evaluating sentencing options. This data-driven approach is often framed as a means of enhancing the rationality and empirical grounding of judicial decisions.

In this context, AI may assist judges by providing structured risk assessments, recidivism probabilities, and comparative sentencing benchmarks derived from prior cases. Such tools can contribute to more informed decision-making processes, particularly in complex cases where large amounts of information must be synthesized. However, scholars caution that reliance on statistical inference must be carefully balanced with normative legal judgment, as judicial decision-making is ultimately grounded in principles of proportionality, fairness, and individualized assessment

(Harcourt 2015). In addition to improving decision quality and efficiency, AI systems may also reduce administrative burdens within judicial institutions. Many routine tasks in sentencing processes—such as data entry, case classification, document retrieval, and basic statistical analysis—can be partially automated through algorithmic systems. This automation can free judicial and administrative personnel to focus on more complex legal tasks requiring human interpretation and discretion.

From an institutional governance perspective, this redistribution of workload may enhance overall judicial resource allocation and improve institutional responsiveness. However, as critical scholarship highlights, the benefits of automation must be carefully evaluated in relation to potential risks of over-reliance on technological systems, particularly where automation may obscure accountability or reduce transparency in decision-making processes (Burrell 2016).

## LEGAL CHALLENGES OF AI-ASSISTED SENTENCING

The incorporation of artificial intelligence (AI) into sentencing processes raises profound legal challenges that extend beyond technical optimization and directly engage with foundational principles of criminal justice. While AI systems are often presented as tools that enhance consistency and efficiency, their integration into judicial decision-making inevitably intersects with constitutional doctrines concerning judicial independence, due process guarantees, and accountability structures. As algorithmic governance becomes increasingly embedded in legal institutions, scholars caution that such developments may subtly reconfigure the role of the judge from an autonomous decision-maker to a user of probabilistic recommendations generated by opaque computational systems (Yeung 2018; Burrell 2016). This shift raises normative concerns about whether sentencing remains a fundamentally human-centered judicial function or becomes partially delegated to algorithmic infrastructures.

Judicial independence is a cornerstone of the rule of law, requiring that judges exercise impartial and autonomous judgment free from external

pressures. In sentencing contexts, this principle is closely linked to judicial discretion, which enables judges to tailor sanctions based on individualized assessments of culpability, context, and proportionality. Traditional sentencing theory emphasizes that such discretion is essential for achieving substantive justice, as rigid standardization may fail to capture the complexity of human behavior and situational nuance (Harcourt 2015).

However, the introduction of AI-based sentencing tools introduces a new form of epistemic influence that may indirectly constrain judicial autonomy. Even when algorithmic outputs are formally advisory, empirical research suggests that judges may exhibit automation bias, whereby they disproportionately rely on algorithmic recommendations, particularly when such outputs are presented as objective or data-driven (Cave and Dignum 2019). This phenomenon raises concerns that judicial reasoning may gradually become anchored to computational outputs, thereby narrowing the scope of independent legal interpretation.

Scholars have further argued that the normalization of algorithmic recommendations may contribute to a structural shift in sentencing practices, where judicial discretion is not formally eliminated but functionally weakened through institutional dependence on AI systems (Kleinberg et al. 2018). This dynamic challenges the traditional understanding of judicial independence as both a legal and cognitive safeguard against external influence.

The principle of due process requires that judicial proceedings be fair, transparent, and subject to meaningful contestation. In the context of AI-assisted sentencing, this principle is challenged by the opacity of algorithmic systems, which often operate as "black boxes" that generate outputs without providing intelligible explanations for their reasoning processes (Burrell 2016; Wachter, Mittelstadt, and Floridi 2017).

A key concern is whether defendants possess an effective right to a fair trial when sentencing decisions are influenced by algorithmic assessments that cannot be fully explained or interrogated. Procedural fairness requires that parties understand the basis of judicial decisions and have the opportunity to challenge adverse evidence. However, when sentencing

recommendations are derived from proprietary models or complex machine learning systems, the ability to contest such outputs becomes severely limited.

This problem is compounded by the asymmetry of technical knowledge between state institutions and defendants. As a result, algorithmic sentencing systems may create structural barriers to contestability, undermining the adversarial nature of criminal proceedings. Scholars argue that without robust mechanisms for explainability and transparency, AI-assisted sentencing risks violating core due process guarantees embedded in both constitutional law and international human rights frameworks (Citron 2008; Selbst and Barocas 2018).

The use of AI in sentencing also raises complex questions regarding legal and institutional accountability. In traditional judicial systems, responsibility for sentencing decisions is clearly attributed to judges, who are bound by legal reasoning and subject to appellate review. However, when AI systems contribute to or influence sentencing outcomes, accountability becomes fragmented across multiple actors, including software developers, data providers, judicial institutions, and end users.

This fragmentation creates what scholars describe as "accountability gaps," where it becomes difficult to determine who bears legal responsibility for erroneous or discriminatory outcomes produced by algorithmic systems (Wachter and Mittelstadt 2019). If an AI system produces biased recommendations that influence a sentencing decision, it remains unclear whether liability lies with the judge who relied on the system, the institution that deployed it, or the private entity that developed the algorithm.

Such ambiguity challenges foundational principles of public law, particularly the requirement that exercises of state power be attributable and reviewable. Without clear accountability structures, AI-assisted sentencing risks undermining trust in judicial institutions and weakening mechanisms of legal redress for affected individuals.

A further unresolved issue concerns the legal status of AI-generated sentencing recommendations. In many jurisdictions, AI tools are formally designated as advisory systems intended to assist judicial reasoning rather than replace it. However, the practical influence of such systems may blur the

distinction between advisory input and de facto decision-making authority.

From an evidentiary perspective, algorithmic outputs raise complex questions regarding admissibility, reliability, and probative value. Unlike traditional forms of evidence, AI-generated recommendations are often based on probabilistic models rather than verifiable factual claims, making it difficult to evaluate their legal weight within sentencing proceedings (Kerr 2010). This creates uncertainty regarding whether such outputs should be treated as expert evidence, procedural guidance, or informal judicial input.

Moreover, the lack of standardized regulatory frameworks governing AI in sentencing contributes to inconsistent practices across jurisdictions. Some legal systems permit limited use of algorithmic tools as supplementary guidance, while others remain cautious due to unresolved constitutional concerns. This regulatory ambiguity reflects broader global uncertainty regarding the proper institutional role of AI in judicial decision-making processes (Završnik 2019; Susskind 2019).

## ETHICAL CHALLENGES IN ALGORITHMIC SENTENCING

The deployment of artificial intelligence (AI) in sentencing processes raises profound ethical concerns that extend beyond technical performance and directly implicate foundational principles of justice, including fairness, transparency, human dignity, and moral responsibility. While algorithmic systems are frequently promoted as neutral and objective instruments for improving judicial consistency, critical scholarship has increasingly demonstrated that such systems are embedded within socio-technical environments that inevitably reproduce and sometimes amplify existing inequalities (O'Neil 2016; Benjamin 2019). In this sense, algorithmic sentencing cannot be understood as a purely technical innovation, but rather as an ethical intervention into the distribution of punishment and state coercive power.

One of the most widely discussed ethical concerns in algorithmic sentencing relates to the problem of bias. Algorithmic bias does not arise in isolation but is typically produced through complex interactions between

data selection, model design, and institutional implementation. At the level of data, historical datasets used to train predictive systems often reflect long-standing patterns of racial, socioeconomic, and institutional discrimination. When such datasets are used to inform sentencing algorithms, they risk encoding past inequalities into future decision-making processes, thereby reinforcing systemic disadvantage (Benjamin 2019; Angwin et al. 2016).

Beyond data-related issues, model bias emerges from the structural design of algorithms themselves. Decisions regarding variable selection, weighting mechanisms, and classification thresholds can produce distortions that systematically disadvantage certain groups. These biases are not always intentional but are often the result of technical simplifications that fail to capture the complexity of human behavior and social context.

At the outcome level, algorithmic sentencing tools may generate disparate impacts on vulnerable populations, particularly marginalized racial or socioeconomic groups. Empirical studies have shown that risk assessment tools may disproportionately classify minority defendants as higher risk, thereby influencing sentencing severity and contributing to unequal punishment outcomes (Dressel and Farid 2018). This raises significant ethical concerns regarding distributive justice and equality before the law.

A second major ethical challenge concerns transparency and explainability in algorithmic sentencing systems. Many advanced machine learning models operate as "black box" systems, meaning that their internal decision-making processes are not easily interpretable by human users, including judges and legal practitioners (Burrell 2016). This opacity creates a fundamental tension between computational complexity and the legal requirement for reasoned justification in judicial decisions.

The ethical principle of explainable artificial intelligence (XAI) has therefore emerged as a key normative response to this problem. Explainability requires that algorithmic outputs be sufficiently transparent to allow judicial actors to understand, evaluate, and critically assess the basis of AI-generated recommendations. In sentencing contexts, this is particularly important because judicial reasoning must remain intelligible, reviewable, and contestable to satisfy procedural justice standards (Wachter, Mittelstadt,

and Floridi 2017).

Without adequate explainability, there is a risk that judges may defer to algorithmic outputs without fully understanding their epistemic foundations, thereby weakening the quality of judicial reasoning and undermining accountability. From an ethical perspective, this raises concerns about the legitimacy of decisions that are influenced by systems whose internal logic cannot be meaningfully scrutinized.

Algorithmic sentencing also raises fundamental ethical questions concerning human dignity and individual autonomy. Sentencing is not merely a technical calculation of risk or efficiency but a deeply normative act that reflects societal values regarding responsibility, culpability, and moral agency. Human-centered theories of justice emphasize that individuals must be treated as autonomous moral agents rather than as statistical objects or predictive data points (Harcourt 2015).

The use of algorithmic tools in sentencing risks shifting the focus from individualized moral judgment to aggregated behavioral prediction. This transformation may reduce the space for contextual interpretation and moral deliberation in judicial decision-making. As a result, defendants may be perceived primarily through algorithmic risk profiles rather than as unique persons with individual histories and circumstances. From an ethical standpoint, this raises concerns about the potential dehumanization of justice processes, where individuals are reduced to numerical scores or probabilistic classifications. Such developments may conflict with foundational principles of human dignity embedded in both constitutional law and international human rights frameworks.

A further ethical issue concerns the appropriate boundaries of automation in sentencing decisions. While AI systems may be suitable for supporting administrative efficiency and providing informational assistance, their role in core judicial functions remains highly contested. Scholars argue that there must be clear ethical limits to the delegation of decision-making authority in domains that involve fundamental rights and liberties (Yeung 2018).

In this context, the principle of human oversight becomes central.

Human judges must retain ultimate responsibility for sentencing decisions, even when assisted by algorithmic tools. This requirement reflects the broader ethical commitment to maintaining human control over coercive state power, particularly in criminal justice systems where decisions directly affect liberty, security, and social status. Moreover, ethical governance frameworks increasingly emphasize the need for "meaningful human control," ensuring that human decision-makers are not merely symbolic actors validating algorithmic outputs, but actively engaged in critical evaluation and justification of sentencing outcomes. Without such safeguards, there is a risk that automation may gradually displace human judgment, leading to what some scholars describe as "functionally automated justice" systems.

## HUMAN RIGHTS IMPLICATIONS

The integration of artificial intelligence (AI) into sentencing processes raises significant human rights implications, particularly in relation to equality before the law, due process guarantees, privacy protection, and the requirement of meaningful human oversight. From a normative perspective, these concerns reflect broader tensions between technological efficiency and the protection of fundamental rights as enshrined in international human rights instruments, including the International Covenant on Civil and Political Rights (ICCPR). Scholars in digital governance argue that algorithmic systems must therefore be evaluated not only in terms of their technical performance, but also in relation to their compatibility with foundational principles of human dignity and rights-based governance (Zuboff 2019; Mittelstadt et al. 2016).

The principle of equality before the law constitutes a foundational norm in both constitutional and international human rights law, requiring that similarly situated individuals be treated in a consistent and non-discriminatory manner. In the context of AI-assisted sentencing, however, this principle is increasingly challenged by the risk of algorithmic discrimination. Empirical research has demonstrated that predictive systems

used in criminal justice may reproduce or amplify existing social inequalities embedded in historical datasets, thereby generating disparate impacts on racial, ethnic, or socioeconomic groups (Benjamin 2019; Angwin et al. 2016).

While AI systems are often justified on the basis of enhancing consistency in sentencing, such consistency does not necessarily guarantee substantive equality. On the contrary, uniform application of biased models may institutionalize inequality under the appearance of neutrality. This raises critical concerns regarding whether algorithmic sentencing tools can be reconciled with the principle of equal protection under the law, particularly in contexts where data-driven models reflect structural disadvantages rather than objective legal criteria.

The right to due process is central to fair trial guarantees and requires that judicial proceedings be conducted in a transparent, participatory, and reviewable manner. In AI-assisted sentencing, due process concerns arise primarily from the opacity of algorithmic systems, which may limit the ability of defendants to understand or challenge the basis of sentencing recommendations (Burrell 2016).

Transparency is a key component of procedural fairness, as individuals must be able to comprehend the reasoning behind judicial decisions that affect their liberty. However, when sentencing outcomes are influenced by complex machine learning models, the underlying logic may be inaccessible not only to defendants but also to legal professionals and judges themselves. This creates significant challenges for procedural participation, particularly in relation to the ability to contest algorithmic evidence or challenge its validity in court. Furthermore, effective remedies become difficult to secure when harm results from algorithmic decision-making processes that are not fully explainable. Without clear avenues for review and contestation, the procedural integrity of sentencing decisions may be weakened, raising concerns about compliance with due process standards under both domestic constitutional law and international human rights frameworks.

AI-assisted sentencing systems frequently rely on the processing of large volumes of personal, behavioral, and sometimes sensitive data, including criminal records, digital footprints, and socio-economic indicators. The

aggregation and analysis of such data raise significant privacy concerns, particularly in relation to the scope, purpose, and proportionality of data collection (Solove 2021).

From a human rights perspective, the right to privacy requires that any interference with personal data be lawful, necessary, and proportionate. However, predictive sentencing systems may rely on extensive data profiling that extends beyond traditional legal boundaries, potentially creating risks of surveillance expansion and data overreach. These concerns are further intensified in jurisdictions where data protection frameworks remain underdeveloped or inconsistently enforced. As a result, the use of AI in sentencing must be accompanied by robust data governance mechanisms that ensure compliance with privacy standards, limit unnecessary data collection, and protect individuals from excessive state intrusion.

A critical emerging principle in AI governance is the requirement for meaningful human involvement in automated or semi-automated decision-making processes. In the context of sentencing, this principle is often articulated as the "human-in-the-loop" requirement, which ensures that final decisions affecting fundamental rights are made or substantively reviewed by human judges rather than delegated entirely to algorithmic systems (Yeung 2018).

The right to human review is closely linked to concerns about accountability, legitimacy, and moral agency in judicial decision-making. While AI systems may provide probabilistic recommendations, they lack the capacity for moral reasoning, contextual interpretation, and normative judgment that are essential in sentencing determinations. Therefore, maintaining human oversight is not only a procedural safeguard but also a substantive requirement for preserving the ethical integrity of criminal justice systems. Without meaningful human intervention, there is a risk that sentencing decisions may become overly dependent on algorithmic outputs, thereby undermining the relational and interpretive dimensions of judicial reasoning.

The concept of "human rights by design" has emerged as a normative framework advocating the integration of human rights principles into the

development and deployment of AI systems from the earliest stages of design. Rather than treating human rights as external constraints applied after system implementation, this approach emphasizes proactive embedding of rights protections into technical architectures, data governance structures, and institutional workflows (Zuboff 2019).

In sentencing contexts, this implies that AI systems should be designed with built-in safeguards to ensure fairness, transparency, accountability, and privacy protection. Such safeguards may include bias detection mechanisms, explainability features, auditability functions, and human oversight requirements. By embedding human rights considerations into system design, policymakers and developers can reduce the risk of systemic rights violations and enhance the legitimacy of AI-assisted judicial processes. Ultimately, the human rights by design approach reflects a broader normative shift toward rights-centered digital governance, where technological innovation is guided not solely by efficiency or performance metrics, but by foundational commitments to dignity, equality, and justice.

## COMPARATIVE ANALYSIS OF INTERNATIONAL EXPERIENCES

The global deployment of artificial intelligence (AI) in criminal justice systems reveals divergent regulatory philosophies and governance models shaped by differing legal traditions, institutional capacities, and normative commitments. Comparative analysis demonstrates that while some jurisdictions prioritize efficiency and predictive accuracy, others emphasize fundamental rights, transparency, and accountability as governing constraints on algorithmic decision-making. This tension reflects a broader global debate on whether AI in justice systems should be optimized primarily as a managerial tool or regulated as a constitutional and human rights-sensitive technology (Calo 2015; Wachter, Mittelstadt, and Floridi 2017).

In the United States, one of the most widely discussed applications of algorithmic decision-making in criminal justice is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) risk assessment system. COMPAS is designed to estimate the likelihood of recidivism and

assist judges in making sentencing, bail, and parole decisions by generating risk scores based on offender characteristics and historical data patterns (Angwin et al. 2016).

The system was initially developed to improve consistency and efficiency in judicial decision-making by providing empirically grounded risk assessments. However, its deployment has generated significant controversy, particularly following empirical studies suggesting that the system may produce racially disparate outcomes, disproportionately labeling African American defendants as higher risk compared to white defendants with similar profiles (Angwin et al. 2016; Dressel and Farid 2018). These findings have triggered extensive legal and ethical debates regarding algorithmic fairness, due process, and transparency.

Judicial scrutiny of COMPAS reached a landmark moment in *State v. Loomis* (2016), where the Wisconsin Supreme Court upheld the use of COMPAS in sentencing but emphasized that it should not be the sole determinant of judicial decisions. The court also highlighted concerns regarding proprietary secrecy and the inability of defendants to fully challenge algorithmic outputs, thereby underscoring due process limitations inherent in black-box systems.

The European Union has adopted a fundamentally rights-based approach to AI governance, emphasizing the protection of fundamental rights, transparency, and accountability in algorithmic systems. This approach is reflected in the European Commission's *Ethics Guidelines for Trustworthy AI*, which establish that AI systems should be lawful, ethical, and robust, with particular attention to human oversight and explainability (European Commission 2019).

More recently, the EU AI Act has introduced a risk-based regulatory framework that classifies AI systems according to their potential impact on fundamental rights. Within this framework, AI applications in criminal justice—particularly those involving risk assessment and sentencing—are generally considered "high-risk systems," subject to stringent requirements such as transparency obligations, data governance standards, human oversight, and conformity assessments (European Parliament 2024).

The EU model reflects a precautionary regulatory philosophy, where technological innovation is permitted but strictly conditioned by constitutional principles, particularly the Charter of Fundamental Rights of the European Union. This rights-centered regulatory architecture positions human dignity, fairness, and accountability as non-negotiable constraints on algorithmic governance.

Canada has developed one of the most structured governmental approaches to algorithmic accountability through its Algorithmic Impact Assessment (AIA) framework. Introduced by the Treasury Board of Canada Secretariat, the AIA is designed to evaluate the potential risks and impacts of automated decision-making systems used in public administration, including criminal justice-related applications (Government of Canada 2019).

The AIA model classifies AI systems according to levels of impact severity and requires corresponding mitigation strategies, including transparency obligations, human oversight requirements, and periodic audits. This approach reflects a governance philosophy that seeks to balance innovation with accountability by embedding risk assessment mechanisms into the lifecycle of algorithmic systems. Canada's framework is particularly significant because it operationalizes accountability through procedural governance rather than purely substantive legal prohibitions. However, scholars have noted that the effectiveness of the AIA depends heavily on institutional compliance and the quality of implementation across federal agencies (Crawford and Schultz 2019).

Singapore represents a leading example of proactive state-led AI governance and judicial digital innovation. Through initiatives such as the Model AI Governance Framework developed by the Personal Data Protection Commission (PDPC), Singapore has emphasized principles of explainability, fairness, and human-centric decision-making in AI deployment (PDPC Singapore 2020). In the judicial domain, Singapore has integrated digital technologies into court administration and case management systems as part of its broader Smart Nation strategy. These initiatives focus on enhancing efficiency, reducing administrative burdens, and improving access to justice through digital platforms. Importantly, Singapore's governance model

emphasizes "human-in-the-loop" principles, ensuring that AI systems remain subordinate to judicial discretion. While Singapore's approach is often praised for its pragmatic balance between innovation and control, critics argue that its relatively centralized governance structure may limit external scrutiny and independent oversight compared to more adversarial legal systems (Lee 2021).

The comparative analysis of international experiences highlights several critical lessons for Indonesia in developing a governance framework for AI in criminal justice and sentencing systems. First, across all jurisdictions examined, transparency emerges as a foundational requirement for legitimizing algorithmic decision-making. Without sufficient explainability and access to information, AI systems risk undermining procedural fairness and public trust.

Second, regulatory safeguards are essential to ensure that AI systems do not reproduce or amplify existing structural inequalities. The experiences of the United States demonstrate the risks of algorithmic bias when systems are deployed without adequate oversight or transparency mechanisms. In contrast, the EU and Canadian models illustrate the importance of embedding rights-based safeguards and risk assessment frameworks into regulatory design.

Third, all jurisdictions emphasize the necessity of maintaining meaningful human oversight in AI-assisted decision-making. Whether through judicial discretion (United States), human rights safeguards (European Union), or procedural governance mechanisms (Canada and Singapore), human involvement remains central to ensuring legitimacy and accountability. For Indonesia, these comparative insights suggest that AI governance in criminal justice should not be limited to technological modernization but must be embedded within a broader normative framework that prioritizes transparency, accountability, human rights protection, and institutional capacity-building. Without such safeguards, the adoption of AI in sentencing risks creating governance gaps that may undermine the legitimacy of the criminal justice system.

## REGULATORY FRAMEWORK FOR AI-ASSISTED SENTENCING IN INDONESIA

The regulatory governance of artificial intelligence (AI) in sentencing processes in Indonesia must be situated within the broader transformation of judicial digitalization and public sector innovation. Over the past decade, Indonesia has progressively adopted e-court systems, electronic case management, and integrated digital judicial services as part of its modernization agenda, reflecting global trends in digital-era governance (Dunleavy et al. 2006; Susskind 2019). However, these reforms primarily concern administrative efficiency rather than algorithmic decision-making in judicial outcomes. Consequently, while Indonesia has developed a relatively advanced framework for digital court administration, it has not yet established a specific legal regime governing AI-assisted sentencing. This gap is normatively significant because sentencing decisions involve the direct limitation of fundamental rights, particularly liberty, and therefore require a higher threshold of legal regulation than ordinary administrative digitalization. As scholars of algorithmic governance emphasize, the transition from digital administration to algorithmic adjudication introduces new constitutional and human rights implications that cannot be addressed through traditional procedural reforms alone (Yeung 2018; Wachter, Mittelstadt, and Floridi 2017).

Indonesia's current legal landscape reflects a hybrid condition in which digital governance tools are increasingly embedded in judicial administration, yet remain largely detached from substantive adjudicatory functions. The Supreme Court's digital transformation initiatives, including electronic filing systems and online case management platforms, demonstrate a clear institutional commitment to efficiency, accessibility, and administrative modernization (Susskind 2019). These reforms align with broader national policies on electronic governance and digital public services, which emphasize bureaucratic simplification and technological integration. Nevertheless, existing legal frameworks—such as criminal procedure laws and judicial regulations—do not explicitly regulate the use of algorithmic

systems in sentencing decisions. This creates a structural distinction between procedural digitalization and algorithmic adjudication. From a comparative perspective, jurisdictions such as the European Union have already begun addressing this transition through risk-based regulatory models that classify judicial AI as high-risk systems requiring stringent safeguards (European Commission 2019). In contrast, Indonesia's legal framework remains predominantly silent on issues such as algorithmic transparency, explainability, and judicial accountability in AI-assisted sentencing, thereby generating normative uncertainty regarding the legality and admissibility of such systems.

The absence of a comprehensive regulatory framework for AI in sentencing exposes several critical governance gaps in Indonesia's criminal justice system. First, there is no dedicated legal standard governing the design, validation, or deployment of algorithmic decision-support systems in judicial contexts, which contrasts sharply with emerging international approaches such as Canada's Algorithmic Impact Assessment framework (Government of Canada 2019). Second, accountability structures remain underdeveloped, particularly regarding the allocation of responsibility when AI-generated recommendations influence judicial outcomes. This fragmentation of responsibility reflects what scholars describe as the "accountability gap" in algorithmic governance, where legal liability becomes diffused across judges, institutions, and technology providers (Wachter and Mittelstadt 2019). Third, transparency requirements are minimal, particularly concerning access to algorithmic logic, training data, and performance metrics. As Burrell (2016) argues, the "black box" nature of machine learning systems undermines interpretability and challenges procedural fairness. In the Indonesian context, these gaps are particularly significant because sentencing decisions are constitutionally bound by principles of due process and equality before the law, meaning that any opacity in decision-making processes may raise constitutional and human rights concerns under both domestic law and the ICCPR framework (UN Human Rights Committee 2007).

A normative framework for AI-assisted sentencing in Indonesia should

be grounded in a set of interrelated governance principles that integrate constitutional law, human rights standards, and algorithmic accountability theory. First, the principle of human-centered AI requires that judicial discretion remain the ultimate authority in sentencing decisions, ensuring that algorithmic systems function strictly as advisory tools rather than decision-making substitutes (Yeung 2018). Second, explainability must be institutionalized as a legal requirement, ensuring that algorithmic outputs are interpretable and capable of judicial scrutiny, consistent with procedural fairness standards (Wachter, Mittelstadt, and Floridi 2017). Third, accountability must be clearly allocated through legal mechanisms that define responsibility among judges, institutions, and developers, thereby preventing liability diffusion. Fourth, fairness must be operationalized through anti-discrimination safeguards that address algorithmic bias arising from historical datasets and structural inequalities, as documented in empirical studies of risk assessment tools in criminal justice systems (Angwin et al. 2016; Benjamin 2019). Finally, oversight mechanisms should include independent auditing institutions capable of evaluating algorithmic performance, bias, and compliance with legal norms. These principles collectively reflect a shift toward rights-based algorithmic governance, where technological innovation is subordinated to constitutional and ethical constraints.

A responsible governance model for AI-assisted sentencing in Indonesia should adopt a multi-layered institutional architecture that integrates advisory functionality, mandatory human oversight, and continuous regulatory evaluation. First, AI systems must be restricted to an advisory-only role, ensuring that algorithmic outputs do not possess binding legal force in sentencing decisions. This is consistent with international best practices emphasizing that high-stakes judicial decisions must remain human-centered to preserve legitimacy and moral accountability (Susskind 2019). Second, mandatory human review should be established as a procedural safeguard, requiring judges to critically assess and independently justify sentencing decisions even when algorithmic recommendations are used. This reflects the principle of "meaningful human control," which has become central in AI

ethics and governance discourse (Cave and Dignum 2019). Third, periodic auditing mechanisms should be institutionalized to assess algorithmic accuracy, bias, and systemic impact, drawing on models such as Canada's Algorithmic Impact Assessment framework (Government of Canada 2019). Finally, the entire system must be embedded within a rights-based governance structure that prioritizes constitutional guarantees of due process, equality, privacy, and fair trial rights, thereby ensuring that AI adoption enhances rather than undermines the legitimacy of Indonesia's criminal justice system.

## CONCLUSION

The analysis demonstrates that artificial intelligence (AI) presents substantial opportunities for enhancing sentencing practices, particularly in improving consistency, efficiency, and data-informed decision-making within criminal justice systems. However, these potential benefits are accompanied by significant legal, ethical, and human rights risks. Issues such as algorithmic bias, opacity in decision-making processes, and fragmented accountability structures remain unresolved challenges that complicate the integration of AI into judicial sentencing. In particular, fully automated sentencing systems are difficult to reconcile with foundational principles of judicial independence and the requirement for individualized justice, as sentencing inherently involves normative judgment that cannot be fully captured by computational models (Yeung 2018; Burrell 2016).

From a legal and human rights perspective, the study underscores that AI governance in sentencing must remain firmly anchored in constitutional principles and international human rights standards. Judicial authority should not be delegated to automated systems, even in cases where algorithmic outputs are framed as advisory tools, as this may undermine procedural fairness and weaken the legitimacy of judicial decision-making (Wachter, Mittelstadt, and Floridi 2017). Accordingly, robust safeguards are necessary to ensure transparency, accountability, and meaningful human oversight, thereby preserving public trust in the criminal justice system.

Without such safeguards, the deployment of AI risks eroding core guarantees of due process, equality before the law, and fair trial rights.

In response to these challenges, this study recommends the development of comprehensive regulatory frameworks governing judicial AI, including mandatory transparency and explainability standards, as well as independent algorithmic audit mechanisms. At the institutional level, strengthening judicial digital literacy, establishing multidisciplinary oversight bodies, and enhancing ethical governance structures are essential to ensure responsible implementation. From a human rights perspective, safeguards must guarantee meaningful human review in all sentencing decisions, prevent discriminatory outcomes, and protect due process rights. Future research should focus on empirical evaluation of AI sentencing tools, comparative analysis of judicial AI governance across jurisdictions, and deeper examination of the constitutional and human rights implications of increasingly automated judicial systems.

## ACKNOWLEDGMENTS

None.

## FUNDING INFORMATION

None.

## CONFLICTING INTEREST AND ORGINALITY STATEMENT

The authors state that there is no conflict of interest in the publication of this article.

## GENERATIVE AI STATEMENT

None.

## REFERENCES

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

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1). https://doi.org/10.1177/2053951715622512

Calo, Ryan. 2015. "Robotics and the Lessons of Cyberlaw." *California Law Review* 103 (3): 513–563. https://doi.org/10.15779/Z38B56D

Cave, Stephen, and Virginia Dignum. 2019. "AI Ethics: A New Field for Law and Policy." *AI & Society* 34: 1–3. https://doi.org/10.1007/s00146-019-00913-9

Citron, Danielle Keats. 2008. "Technological Due Process." *Washington University Law Review* 85 (6): 1249–1313.

Crawford, Kate, and Jason Schultz. 2019. "AI Systems as State Actors." *Columbia Law Review* 119 (7): 1941–1972.

Dressel, Julia, and Hany Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Science Advances* 4 (1). https://doi.org/10.1126/sciadv.aao5580

Dunleavy, Patrick, Helen Margetts, Simon Bastow, and Jane Tinkler. 2006. "New Public Management Is Dead—Long Live Digital-Era Governance." *Journal of Public Administration Research and Theory* 16 (3): 467–494. https://doi.org/10.1093/jopart/mui057

European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels.

European Parliament. 2024. *Artificial Intelligence Act (AI Act)*. Brussels.

Government of Canada. 2019. *Directive on Automated Decision-Making*. Ottawa: Treasury Board Secretariat.

Hao, Karen. 2019. "AI Is Sending People to Jail—and Getting It Wrong." *MIT Technology Review*.

Harcourt, Bernard E. 2015. *Exposed: Desire and Disobedience in the Digital Age*.

Cambridge, MA: Harvard University Press.

Kerr, Orin S. 2010. "Digital Evidence and the New Criminal Procedure." *Columbia Law Review* 110 (4): 279–342.

Kitchin, Rob. 2017. "Thinking Critically About and Researching Algorithms." *Information, Communication & Society* 20 (1): 14–29. https://doi.org/10.1080/1369118X.2016.1154087

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein. 2018. "Discrimination in the Age of Algorithms." *Journal of Legal Analysis* 10 (1): 113–174. https://doi.org/10.1093/jla/laz001

Lee, Julian. 2021. "AI Governance in Singapore: Balancing Innovation and Control." *Asian Journal of Law and Society* 8 (2): 245–267.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* 3 (2). https://doi.org/10.1177/2053951716679679

O'Neil, Cathy. 2016. *Weapons of Math Destruction*. New York: Crown.

Personal Data Protection Commission (PDPC) Singapore. 2020. *Model AI Governance Framework*. Singapore.

Selbst, Andrew D., and Solon Barocas. 2018. "The Intuitive Appeal of Explainable Machines." *Fordham Law Review* 87 (3): 1085–1139.

Solove, Daniel J. 2021. *Understanding Privacy*. Cambridge, MA: Harvard University Press.

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

Susskind, Richard. 2019. *Online Courts and the Future of Justice*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198838364.001.0001

UN Human Rights Committee. 2007. *General Comment No. 32: Right to Equality Before Courts and Tribunals and to a Fair Trial (Article 14)*.

Wachter, Sandra, and Brent Mittelstadt. 2019. "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data." *Columbia Business Law Review* 2019 (2): 494–620.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR." *International Data Privacy Law* 7 (2): 76–99.

https://doi.org/10.1093/idpl/ipx005

Yeung, Karen. 2018. "Algorithmic Regulation: A Critical Interrogation." *Regulation & Governance* 12 (4): 505–523. https://doi.org/10.1111/rego.12158

Završnik, Aleš. 2019. "Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings." *European Journal of Criminology* 18 (5): 1–20. https://doi.org/10.1177/1477370819876762

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*. New York: PublicAffairs.