Artificial Intelligence in Malaysian Criminal Sentencing: Reassessing Due Process and Judicial Accountability

Mohamad Hilman Nabil bin Fakrul Nizam[1], Izuan Izzaidi bin Azmi[2], Nur Aleya Aisyah binti Mat Khalid[3], Danish Hakim bin Kamarul Anuar[4], Nuha Al-Hani binti Azharar[5]

[1-5]Faculty of Law, Universiti Teknologi MARA, Malaysia
✉ Corresponding email: uitmlegalresearch@gmail.com

History of Article
Submitted : July 18, 2026
Revised : July 25, 2026
Accepted : August 10, 2026
Published : September 22, 2026
DOI : https://doi.org/10.37253/conescintech.v6i1.12701

Copyright© 2026 by Author(s). This work is licensed under a Creative Commons Attribution-Non Commercial-Share Alike 4.0 International License. All writings published in this journal are personal views of the authors and do not represent the views of this journal and the author's affiliated institutions.

Abstract
The integration of artificial intelligence (AI) into criminal sentencing has increasingly influenced decision-making within contemporary criminal justice systems. In Malaysia, the implementation of the Artificial Intelligence Sentencing System (AISS) in Sabah and Sarawak courts has intensified debates concerning the compatibility of AI-assisted sentencing with the rule of law, judicial discretion and constitutional safeguards. Although AI-powered sentencing tools are promoted as mechanisms capable of improving efficiency, consistency and predictive assessment, their application raises serious concerns relating to transparency, explainability, procedural fairness and algorithmic bias. The absence of a specific Malaysian legal framework regulating AI-assisted sentencing further creates uncertainty regarding the extent to which courts may rely on algorithmic recommendations without compromising the accused person's constitutional rights under Articles 5 and 8 of the Federal Constitution. Accordingly, this study examines the legal implications of AI-assisted criminal sentencing in Malaysia through a doctrinal legal research methodology. The research analyses relevant statutory provisions, including the Penal Code (Act 574), Criminal Procedure Code (Act 593), Subordinate Courts Act 1948. The study finds that excessive reliance on opaque algorithmic systems may undermine natural justice, weaken judicial reasoning and challenge the principles underpinning the rule of law. This article argues that AI should operate solely as a supplementary judicial aid subject to strict human oversight, transparency requirements and explainability safeguards within Malaysia's criminal justice framework.

Keywords: Artificial Intelligence; Criminal Sentencing; Algorithmic Accountability.

Introduction

Artificial intelligence (AI) has emerged as one of the most transformative technological developments of the twenty-first century, reshaping various sectors including healthcare, finance, education, transportation, and public administration. In recent years, AI has increasingly been integrated into legal and judicial processes, giving rise to what is commonly referred to as "AI-assisted justice." Courts and legal institutions worldwide have begun exploring the potential of AI technologies to improve efficiency, consistency, and accessibility within the administration of justice (Završnik, 2020). Among the most significant developments in this area is the use of AI-assisted sentencing systems, which utilise algorithmic models to analyse previous judicial decisions and generate sentencing recommendations for judges. Such systems are often promoted as tools capable of reducing sentencing disparities, enhancing judicial consistency, and improving the overall efficiency of criminal justice systems (Putera et al., 2022).

The growing adoption of AI-assisted sentencing reflects broader efforts by governments and judicial institutions to modernise legal processes through technological innovation. Jurisdictions such as the United States, Canada, the United Kingdom, and several European countries have experimented with algorithmic decision-making tools in criminal justice. One of the most widely discussed examples is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system in the United States, which assists judges by assessing the likelihood of recidivism among offenders (Angwin et al., 2016). Supporters of such technologies argue that AI systems are capable of processing vast amounts of information more efficiently than humans, thereby producing more consistent and data-driven outcomes. From this perspective, AI-assisted sentencing has the potential to strengthen public confidence in the criminal justice system by minimising arbitrary decision-making and reducing disparities between similar cases (Mullainathan & Spiess, 2017).

Despite these perceived benefits, the incorporation of AI into criminal sentencing has generated significant legal, ethical, and constitutional concerns. Critics argue that sentencing decisions involve more than statistical predictions and require the careful exercise of judicial discretion, moral reasoning, and consideration of individual circumstances. Unlike administrative decisions, criminal sentencing directly affects fundamental rights and liberties, particularly the liberty of the accused person. Consequently, concerns have emerged regarding whether algorithmic recommendations may undermine due process, weaken judicial accountability, and compromise the transparency that is expected of judicial decision-making (Citron, 2008; Završnik, 2020). The increasing reliance on AI within criminal justice systems has therefore prompted important debates concerning the compatibility of algorithmic decision-making with the rule of law and established principles of natural justice.

Within the Malaysian context, discussions concerning AI and criminal justice became increasingly relevant following the introduction of the Artificial Intelligence Sentencing System (AISS) in Sabah and Sarawak. The AISS was developed through collaboration between the judiciary and technology experts as an initiative to assist judges in determining appropriate sentences for selected offences (Khong & Ho, 2022; Putera et

al., 2022). The system analyses previous sentencing patterns and generates recommendations that may be considered by judges during sentencing proceedings. Although the AISS was introduced primarily to improve consistency and efficiency, its implementation has raised important questions concerning the permissible role of AI in judicial decision-making. While judicial officers remain responsible for the final sentencing decision, concerns persist regarding the extent to which algorithmic recommendations may influence judicial reasoning and affect the exercise of independent judicial discretion (Khong & Ho, 2022).

These concerns are particularly significant in light of Malaysia's constitutional framework. Criminal sentencing involves the exercise of judicial power and directly impacts an individual's liberty. Article 5(1) of the Federal Constitution guarantees that no person shall be deprived of life or personal liberty save in accordance with law, while Article 8(1) guarantees equality before the law and equal protection of the law. These constitutional provisions embody fundamental principles of due process, fairness, and equality that must be respected throughout criminal proceedings (Federal Constitution). The use of AI-assisted sentencing therefore raises the question of whether reliance on algorithmic recommendations may inadvertently compromise these constitutional safeguards. For instance, if the reasoning underlying an AI-generated recommendation is not fully transparent or explainable, an accused person may be unable to effectively challenge the basis of the recommendation. Such circumstances may undermine procedural fairness and potentially conflict with the constitutional guarantees afforded by Articles 5 and 8 of the Federal Constitution (Citron, 2008; Goodman & Flaxman, 2017). The legal concerns surrounding AI-assisted sentencing are not merely theoretical. Existing scholarship has identified numerous risks associated with algorithmic decision-making in criminal justice. A substantial body of literature highlights the problem of algorithmic bias, whereby AI systems may reproduce or amplify historical inequalities embedded within training datasets (Mayson, 2019). Scholars have observed that AI systems are only as reliable as the data upon which they are trained. Consequently, where historical sentencing patterns reflect social, economic, or institutional biases, AI-generated recommendations may perpetuate those biases under the guise of technological objectivity. This concern has become a central issue within international discussions on AI governance and criminal justice reform (Angwin et al., 2016; Mayson, 2019).

In addition to concerns regarding bias, existing studies have emphasised the problem of transparency and explainability. Many advanced AI systems operate through complex computational processes that are not easily understood by judges, lawyers, or defendants. This phenomenon is commonly referred to as the "black box" problem (Goodman & Flaxman, 2017; Malgieri & Comandé, 2017). Where sentencing recommendations are generated through opaque algorithms, it becomes difficult to determine how specific conclusions were reached or whether legally relevant factors were appropriately considered. Scholars have argued that such opacity is fundamentally inconsistent with the principles of judicial accountability and reasoned decision-making that underpin modern legal systems (Citron, 2008; Oswald, 2018). Since courts are generally required to provide reasons for their decisions, the increasing use of AI-

generated recommendations raises concerns regarding whether judges can adequately explain decisions that are influenced by algorithmic outputs.

Furthermore, previous research has questioned the impact of AI-assisted sentencing on judicial discretion. Sentencing has traditionally been regarded as a judicial function requiring judges to balance multiple factors, including the seriousness of the offence, mitigating circumstances, aggravating factors, public interest considerations, and the offender's personal circumstances. While AI systems may assist by identifying sentencing patterns, excessive reliance on algorithmic recommendations risks transforming judicial decision-making into a mechanical exercise (Putera et al., 2022). Several scholars therefore argue that AI should function only as a supplementary tool rather than a substitute for human judgment. The preservation of meaningful judicial discretion remains essential to ensuring that sentencing decisions remain individualised, proportionate, and consistent with the principles of justice (Khong & Ho, 2022; Završnik, 2020).

Although extensive international scholarship exists regarding AI-assisted sentencing, significant gaps remain within Malaysian legal literature. Most existing studies focus on the technical capabilities, ethical implications, or comparative experiences of foreign jurisdictions. Relatively little attention has been devoted to examining the constitutional implications of AI-assisted sentencing within Malaysia's legal framework. Existing Malaysian discussions frequently emphasise the operational aspects of the Artificial Intelligence Sentencing System but seldom undertake a comprehensive analysis of its compatibility with constitutional safeguards, judicial accountability, and the rule of law (Khong & Ho, 2022; Putera et al., 2022). Consequently, there remains a lack of scholarly consensus regarding the appropriate legal framework that should govern AI-assisted sentencing in Malaysia.

This research addresses that gap by examining AI-assisted criminal sentencing through the lens of Malaysian constitutional and criminal law. Unlike previous studies that primarily focus on technological efficiency or operational implementation, this study critically evaluates the compatibility of AI-assisted sentencing with constitutional guarantees, judicial discretion, and principles of natural justice. The novelty of this study lies in its doctrinal examination of AI-assisted sentencing within the specific context of Malaysian constitutional law, particularly Articles 5 and 8 of the Federal Constitution. By integrating constitutional analysis with contemporary debates on AI governance, the study provides a legal perspective that has thus far received limited attention within Malaysian scholarship.

Accordingly, this study seeks to achieve four principal objectives. First, it aims to examine the legal implications of AI-assisted sentencing within the Malaysian criminal justice system. Second, it seeks to analyse the compatibility of AI-assisted sentencing with constitutional safeguards under Articles 5 and 8 of the Federal Constitution. Third, it evaluates the impact of AI-generated recommendations on judicial discretion and accountability. Finally, the study proposes appropriate legal and regulatory safeguards that may guide the future development and implementation of AI-assisted sentencing within Malaysia.

The significance of this study is both theoretical and practical. From a theoretical perspective, the research contributes to the growing body of scholarship concerning the intersection between law, technology, and constitutional governance. It expands existing discussions regarding AI-assisted justice by examining the implications of algorithmic decision-making within a constitutional framework. From a practical perspective, the findings may assist policymakers, judicial institutions, legal practitioners, and technology developers in understanding the legal risks associated with AI-assisted sentencing and identifying safeguards necessary to protect constitutional rights. Ultimately, the study seeks to ensure that future technological innovation within the Malaysian criminal justice system remains consistent with the principles of due process, judicial accountability, and the rule of law. This research is limited to the use of AI in criminal sentencing and does not examine the application of AI in other areas of criminal justice, such as criminal investigation, prosecution, or law enforcement activities.

Research Method

This study employs a qualitative doctrinal legal research methodology to examine the legal implications of artificial intelligence (AI)-assisted sentencing within the Malaysian criminal justice system. Doctrinal legal research is particularly appropriate for this study because it focuses on the systematic analysis of legal rules, principles, doctrines, and frameworks governing a particular legal issue. Since the primary objective of this research is to evaluate the compatibility of AI-assisted sentencing with constitutional safeguards, judicial discretion, and principles of accountability, a doctrinal approach provides the most suitable method for analysing the existing legal position and identifying potential regulatory gaps.

The study primarily relies on library-based research involving the collection and examination of both primary and secondary legal sources. Primary legal materials include the Federal Constitution, particularly Articles 5 and 8 relating to personal liberty and equality before the law, the Penal Code (Act 574), the Criminal Procedure Code (Act 593), the Subordinate Courts Act 1948, and other relevant statutes governing criminal justice administration in Malaysia. The study also examines policy documents and guidelines relevant to artificial intelligence governance, including Malaysia's National Guidelines on Artificial Intelligence Governance and Ethics, to assess the extent to which existing regulatory instruments address concerns arising from AI-assisted judicial decision-making.

In addition, judicial decisions from Malaysian courts are analysed to identify the legal principles governing sentencing, judicial discretion, procedural fairness, and the rule of law. Relevant case law, including *PP v Denis P Modili*, is examined to determine whether existing legal doctrines are capable of addressing issues associated with algorithmic decision-making in criminal sentencing (Khong & Ho, 2022). Particular attention is given to judicial principles concerning transparency, reasoned decision-making, and the exercise of independent judicial judgment, as these principles are central to the legitimacy of AI-assisted sentencing systems.

The study further utilises secondary legal materials, including academic journal articles, textbooks, conference papers, law reform reports, policy papers, and scholarly commentaries relating to artificial intelligence, criminal justice, constitutional law, and judicial accountability. These materials provide theoretical and comparative perspectives that assist in evaluating both the benefits and risks of AI-assisted sentencing. Existing literature is reviewed to identify prevailing scholarly views regarding algorithmic bias, explainability, due process, and the preservation of judicial discretion in the age of artificial intelligence (Citron, 2008; Goodman & Flaxman, 2017; Mayson, 2019; Završnik, 2020).

A comparative legal analysis is also adopted to examine regulatory approaches implemented in selected foreign jurisdictions, particularly the United States and the European Union. These jurisdictions have developed extensive legal and policy discussions concerning the use of AI in criminal justice and therefore provide valuable insights into potential safeguards that may be adapted within the Malaysian context. Comparative analysis enables the identification of best practices, regulatory challenges, and accountability mechanisms that may assist Malaysia in formulating an appropriate legal framework for AI-assisted sentencing (Goodman & Flaxman, 2017; Malgieri & Comandé, 2017; Oswald, 2018).

The collected materials are analysed using descriptive, analytical, and critical methods. The descriptive approach is used to explain the operation of AI-assisted sentencing systems and the existing legal framework governing criminal sentencing in Malaysia. The analytical approach is employed to evaluate the compatibility of AI-assisted sentencing with constitutional safeguards and established legal principles. Finally, the critical approach is utilised to assess the adequacy of existing laws and policies in addressing emerging challenges posed by algorithmic decision-making and to formulate recommendations for future legal reform.

Through this methodology, the study seeks to provide a comprehensive legal assessment of AI-assisted sentencing in Malaysia and contribute to the development of a regulatory framework that balances technological innovation with constitutional rights, judicial accountability, and the rule of law.

Results and Discussion

AI-Assisted Sentencing and Due Process Rights

The increasing adoption of artificial intelligence within criminal sentencing has generated significant concerns regarding the protection of due process rights. Due process constitutes a fundamental principle of criminal justice and requires legal proceedings to be conducted fairly, transparently, and in accordance with established legal procedures. Within the Malaysian context, the principle is reflected in Article 5(1) of the Federal Constitution, which provides that no person shall be deprived of life or personal liberty save in accordance with law. Although Article 5 does not expressly mention procedural fairness, Malaysian courts have consistently interpreted the provision as encompassing broader principles of natural justice and fair treatment (Federal Constitution; Citron, 2008).

The introduction of the Artificial Intelligence Sentencing System (AISS) raises important questions regarding whether algorithmic recommendations can be reconciled with these constitutional guarantees. Criminal sentencing directly affects the liberty of an accused person and therefore requires a high degree of transparency and accountability. Traditionally, sentencing decisions are accompanied by judicial reasoning that explains how relevant factors were considered and why a particular sentence was imposed. Such reasoning enables parties to understand the basis of the decision and facilitates appellate review where necessary (Khong & Ho, 2022).

However, AI-assisted sentencing systems may challenge these principles because algorithmic recommendations are often generated through complex computational processes that are not readily understandable to judges, lawyers, or defendants. This lack of explainability creates what scholars commonly describe as the "black box" problem (Goodman & Flaxman, 2017; Malgieri & Comandé, 2017). Where an AI system generates a sentencing recommendation without providing a clear explanation of the factors that influenced the outcome, an accused person may be unable to effectively challenge the recommendation or identify potential errors. Consequently, concerns arise regarding whether reliance on such recommendations satisfies the requirements of procedural fairness and due process (Citron, 2008).

The first reported Malaysian application of AI-assisted sentencing can be observed in *PP v Denis P Modili*, where the Sabah and Sarawak AI Sentencing System generated a sentencing recommendation for consideration by the court. Importantly, the court did not regard the AI-generated recommendation as binding and ultimately exercised independent judicial judgment in determining the appropriate sentence (Khong & Ho, 2022). This demonstrates that the Malaysian judiciary presently regards AI as an advisory mechanism rather than an autonomous decision-maker.

Furthermore, transparency is a fundamental component of the rule of law. Courts are expected to provide reasoned decisions to demonstrate that judicial power has been exercised fairly and rationally. If judges rely heavily on AI-generated recommendations without fully understanding the reasoning behind them, the legitimacy of sentencing decisions may be questioned. Although the AISS currently functions as an advisory tool rather than an autonomous decision-maker, the increasing integration of AI into sentencing processes necessitates safeguards to ensure that transparency and procedural fairness remain protected (Oswald, 2018; Završnik, 2020).

Therefore, while AI-assisted sentencing may enhance efficiency and consistency, its implementation must not compromise the accused person's right to understand, challenge, and scrutinise the basis of sentencing decisions. Any future expansion of AI-assisted sentencing in Malaysia should therefore incorporate explainability requirements and mechanisms that enable meaningful review of algorithmic recommendations (Goodman & Flaxman, 2017).

Judicial Discretion and the Rule of Law

Judicial discretion has long been recognised as a central feature of criminal sentencing. Sentencing involves more than the mechanical application of legal rules; it

requires judges to evaluate the facts of each case, consider mitigating and aggravating circumstances, assess the offender's background, and determine an appropriate punishment that reflects both justice and proportionality. The exercise of judicial discretion therefore ensures that sentencing remains responsive to the unique circumstances of individual cases.

The emergence of AI-assisted sentencing has prompted concerns regarding the extent to which algorithmic recommendations may influence judicial decision-making. While AI systems are capable of identifying patterns and trends from previous cases, they are unable to fully appreciate human experiences, moral considerations, and contextual factors that frequently influence sentencing outcomes. Excessive reliance on AI-generated recommendations may therefore risk reducing sentencing to a data-driven exercise rather than a process grounded in legal judgment and judicial reasoning (Završnik, 2020).

This concern is particularly relevant within Malaysia's constitutional framework, where judicial power is vested in the courts. Judges are entrusted with the responsibility of independently determining appropriate sentences based on the facts and circumstances of each case. The legitimacy of the judicial process depends upon the ability of judges to exercise independent judgment rather than merely adopting recommendations generated by technological systems.

The importance of preserving judicial discretion can again be observed in *PP v Denis P Modili*. Although the AI system generated a sentencing recommendation, the court ultimately retained authority over the final sentencing decision. The recommendation served merely as a reference point and did not replace judicial reasoning or discretion (Khong & Ho, 2022).

Supporters of AI-assisted sentencing argue that the technology does not replace judicial discretion but merely assists judges by providing reference points based on previous cases. In theory, judges remain free to accept, modify, or reject AI-generated recommendations. Nevertheless, critics argue that algorithmic recommendations may exert a subtle influence on judicial reasoning, creating what is sometimes referred to as "automation bias." This phenomenon occurs when individuals place excessive trust in technological outputs despite the existence of alternative conclusions or contradictory evidence (Oswald, 2018).

The rule of law further requires that judicial decisions be based on legal reasoning rather than technological convenience. Sentencing decisions must remain attributable to judges rather than algorithms. Consequently, any use of AI within sentencing must preserve meaningful human oversight and ensure that judges retain ultimate responsibility for sentencing outcomes. AI should therefore be regarded as a supplementary tool designed to support judicial reasoning rather than a substitute for judicial decision-making (Khong & Ho, 2022; Završnik, 2020).

The preservation of judicial discretion is essential to maintaining public confidence in the administration of justice. Accordingly, any future regulatory framework governing AI-assisted sentencing in Malaysia should explicitly affirm that sentencing

authority remains vested exclusively in judicial officers and that AI recommendations are advisory rather than determinative.

Algorithmic Bias and Accountability

One of the most significant criticisms of AI-assisted sentencing concerns the risk of algorithmic bias. Although AI systems are frequently portrayed as objective and impartial, their outputs are heavily dependent upon the data used to train them. Where historical data contains social, economic, or institutional biases, AI systems may inadvertently replicate or reinforce those biases through their recommendations (Mayson, 2019).

The issue of algorithmic bias has attracted considerable international attention. Critics argue that AI systems trained on historical sentencing patterns may reproduce existing disparities within criminal justice systems. For example, if historical sentencing data reflects unconscious bias against particular demographic groups, AI systems may incorporate those patterns into future recommendations. As a result, technology that is intended to improve consistency may instead perpetuate inequality under the appearance of neutrality (Angwin et al., 2016; Mayson, 2019).

Within Malaysia, concerns regarding algorithmic bias are particularly relevant in light of Article 8(1) of the Federal Constitution, which guarantees equality before the law and equal protection of the law. Any sentencing system that systematically disadvantages certain individuals or groups would be inconsistent with these constitutional principles. Therefore, ensuring fairness and equality must remain a central consideration in the development and implementation of AI-assisted sentencing technologies (Federal Constitution).

Closely related to concerns regarding bias is the issue of accountability. Traditional sentencing decisions are attributable to judges who can be questioned, criticised, or challenged through established legal processes. However, where sentencing recommendations are influenced by AI systems, questions arise regarding responsibility for erroneous or discriminatory outcomes. It may be unclear whether responsibility lies with judges, software developers, system administrators, or the institutions responsible for implementing the technology (Završnik, 2020).

This accountability gap presents a significant challenge for legal systems seeking to integrate AI into judicial processes. Without clear accountability mechanisms, individuals affected by inaccurate or unfair recommendations may face difficulties obtaining remedies. Accordingly, robust accountability structures are essential to ensure that AI-assisted sentencing remains subject to legal scrutiny and oversight (Citron, 2008). To address these concerns, future regulatory frameworks should require regular auditing of AI systems, independent assessments of algorithmic fairness, and continuous monitoring of sentencing outcomes. Such measures would help ensure that technological innovation does not undermine constitutional principles of equality and accountability (Binns, 2018; Verma & Rubin, 2018).

Comparative Regulatory Approaches

The challenges associated with AI-assisted sentencing have prompted various jurisdictions to develop legal and policy frameworks aimed at regulating the use of artificial intelligence within criminal justice systems. Examining these approaches provides valuable insights into the safeguards that may be adopted within the Malaysian context.

The United States remains one of the earliest jurisdictions to implement AI-assisted tools within criminal justice proceedings. Systems such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) have been utilised to assess the likelihood of recidivism and assist sentencing decisions. While supporters argue that such tools promote efficiency and consistency, their use has generated considerable controversy. Critics have highlighted concerns regarding algorithmic bias, particularly following allegations that COMPAS disproportionately classified certain groups as high-risk offenders (Angwin et al., 2016). The controversy demonstrated that algorithmic systems are not inherently neutral and may reproduce inequalities embedded within historical data.

The European Union has adopted a more cautious approach towards AI regulation through the development of comprehensive legislative frameworks. The European Union Artificial Intelligence Act classifies AI systems according to their level of risk and imposes stricter obligations on high-risk applications. AI systems utilised within law enforcement and judicial contexts are generally regarded as high-risk technologies due to their potential impact on fundamental rights. Under the European framework, providers of high-risk AI systems must comply with requirements relating to transparency, human oversight, data quality, risk management, and accountability (European Union, 2024; Goodman & Flaxman, 2017).

Similarly, the United Kingdom has emphasised the importance of maintaining human control over AI-assisted decision-making. Judicial officers are expected to retain responsibility for decisions and should not delegate decision-making authority entirely to algorithmic systems. This approach recognises that judicial accountability and public confidence depend upon the continued involvement of human decision-makers capable of exercising judgment and providing reasons for their decisions (Oswald, 2018).

The experiences of these jurisdictions reveal several common themes. First, transparency and explainability are regarded as essential safeguards. Second, meaningful human oversight remains a fundamental requirement for the lawful use of AI in judicial settings. Third, accountability mechanisms are necessary to ensure that individuals affected by AI-assisted decisions have access to remedies and review processes. Finally, regular monitoring and auditing of AI systems are considered crucial in identifying and correcting potential biases.

These comparative approaches demonstrate that successful AI governance within criminal justice systems requires a balance between technological efficiency and legal accountability. Malaysia may therefore draw valuable lessons from foreign jurisdictions when developing its own regulatory framework for AI-assisted sentencing.

Proposed Legal Framework for Malaysia

Given the constitutional and legal concerns associated with AI-assisted sentencing, Malaysia should develop a comprehensive regulatory framework capable of ensuring that technological innovation remains consistent with the principles of due process, judicial accountability, and the rule of law.

The first and most important safeguard is the adoption of a human-in-the-loop principle. Under this principle, AI systems should function solely as decision-support tools rather than autonomous decision-makers. Judges must retain exclusive authority to determine sentencing outcomes and should remain free to accept, modify, or reject algorithmic recommendations. The final sentence imposed should always be attributable to judicial reasoning rather than technological outputs (Khong & Ho, 2022).

Second, Malaysia should introduce mandatory explainability requirements for AI-assisted sentencing systems. Any algorithmic recommendation capable of influencing sentencing outcomes should be accompanied by a clear explanation of the factors considered and the reasoning underlying the recommendation. Explainability is necessary to ensure procedural fairness and enable parties to challenge recommendations where appropriate (Goodman & Flaxman, 2017; Malgieri & Comandé, 2017).

Third, independent auditing mechanisms should be established to monitor the operation of AI-assisted sentencing systems. Regular audits conducted by independent experts would assist in identifying algorithmic bias, inaccuracies, or unintended discriminatory effects. Such audits should assess not only technical performance but also compliance with constitutional principles and legal standards (Binns, 2018).

Fourth, clear accountability mechanisms should be implemented. Legislation should specify the responsibilities of judicial institutions, technology developers, and system administrators in relation to AI-assisted sentencing. Establishing clear lines of responsibility would ensure that affected individuals have access to remedies where algorithmic errors contribute to unjust outcomes (Citron, 2008).

Fifth, procedural safeguards and appeal mechanisms should be strengthened. Defendants should be entitled to access information concerning the role of AI in sentencing decisions and should have the opportunity to challenge algorithmic recommendations where relevant. Courts reviewing sentencing decisions on appeal should likewise be empowered to scrutinise the reliability and appropriateness of AI-generated recommendations.

Finally, Malaysia should consider enacting a dedicated regulatory framework governing the use of AI within the administration of justice. Existing laws such as the Penal Code (Act 574), Criminal Procedure Code (Act 593), and Subordinate Courts Act 1948 (Act 92) were not drafted with algorithmic decision-making in mind and therefore provide limited guidance on issues such as transparency, explainability, and algorithmic accountability. A specialised legislative framework would provide greater certainty and establish uniform standards governing the use of AI within judicial proceedings.

In summary, AI-assisted sentencing offers significant opportunities to enhance efficiency and consistency within the criminal justice system. However, these benefits can only be realised if appropriate safeguards are implemented to protect constitutional rights and preserve judicial accountability. A carefully designed regulatory framework

would enable Malaysia to harness the advantages of AI while ensuring that sentencing decisions remain fair, transparent, and consistent with the rule of law.

Conclusion

The integration of artificial intelligence into criminal sentencing represents a significant development in the modernisation of criminal justice systems. In Malaysia, the implementation of the Artificial Intelligence Sentencing System (AISS) demonstrates the judiciary's willingness to embrace technological innovation in pursuit of greater efficiency and consistency. Nevertheless, the use of AI-assisted sentencing raises complex legal and constitutional issues concerning due process, judicial discretion, transparency, accountability, and equality before the law.

This study has shown that while AI technologies may provide valuable assistance in identifying sentencing patterns and improving consistency, excessive reliance on algorithmic recommendations may undermine fundamental principles of natural justice and the rule of law. Concerns relating to explainability, algorithmic bias, and accountability remain particularly significant because sentencing decisions directly affect the liberty and rights of individuals. The study further demonstrates that existing Malaysian laws provide limited guidance on the regulation of AI-assisted sentencing, thereby creating uncertainty regarding the safeguards necessary to ensure fairness and constitutional compliance.

Comparative analysis reveals that jurisdictions such as the United States, the European Union, and the United Kingdom have recognised the importance of maintaining human oversight, transparency, and accountability when implementing AI technologies within judicial processes. These experiences provide valuable lessons for Malaysia as it seeks to develop an appropriate regulatory framework for AI-assisted sentencing.

Accordingly, this article argues that AI should function only as a supplementary judicial aid and should never replace human judgment in sentencing decisions. The preservation of judicial discretion, procedural fairness, and constitutional rights must remain central considerations in any future expansion of AI-assisted sentencing. By adopting safeguards such as explainability requirements, independent auditing mechanisms, accountability frameworks, and meaningful human oversight, Malaysia can ensure that technological innovation strengthens rather than undermines the administration of justice.

Ultimately, the successful integration of AI into criminal sentencing depends not merely on technological capability but on the ability of legal institutions to ensure that innovation remains consistent with constitutional principles, judicial accountability, and the rule of law.

References

Casey, B., Farhangi, A., & Vogl, R. (2019). Rethinking explainable machines: The GDPR's right to explanation debate and the rise of algorithmic audits in

enterprise. *Berkeley Technology Law Journal*, *34*(1), 143–188. https://btlj.org/data/articles2019/34_1/04_Casey_Web.pdf

Citron, D. K. (2008). Technological due process. *Washington University Law Review*, *85*(6),1249–1313. https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview

Criminal Procedure Code (Act 593). (2017). *Laws of Malaysia: Criminal Procedure Code (Act 593)*. https://www.forestry.gov.my/images/stories/pdf/Akta/Kanun%20Tatacara%20Jenayah%20CPC-%20BI.pdf

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Federal Constitution. (2016). *Federal Constitution*. https://www.rcrc-resilience-southeastasia.org/wp-content/uploads/2017/12/2010_federal_constitution__2016_version_.pdf

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Khong, D. W. K., & Ho, C. C. (2022). Artificial Intelligence in Malaysian Courts: PP v Denis P Modili. *Asian Journal of Law and Policy*, *2*(2), 127–136. https://doi.org/10.33093/ajlp.2022.9

Lim, C., & Gong, R. (2020, August 18). *Artificial intelligence in the courts: AI sentencing in Sabah and Sarawak*. Khazanah Research Institute. https://old.krinstitute.org/Views-%40-Artificial_Intelligence_in_the_Courts-%3B_AI_sentencing_in_Sabah_and_Sarawak.aspx

Malgieri, G., & Comandé, G. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(4), 243–265. https://doi.org/10.1093/idpl/ipx005

Mayson, S. G. (2019). Bias in, bias out. *Yale Law Journal*, *128*, 2218–2300. https://yalelawjournal.org/pdf/Mayson_p5g2tz2m.pdf

Meyer, M., Horowitz, A., Marshall, E., & Lum, K. (2022). Flipping the script on criminal justice risk assessment: An actuarial model for assessing the risk the federal sentencing system poses to defendants. *arXiv*. https://arxiv.org/abs/2205.13505

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Oswald, M. (2018). Algorithm-assisted decision-making in the public sector. *Philosophical Transactions of the Royal Society A*, *376*(2128). https://doi.org/10.1098/rsta.2017.0359

Penal Code (Act 574). (2023). *Laws of Malaysia: Penal Code (Act 574)*. https://www.sprm.gov.my/admin/files/sprm/assets/pdf/penguatkuasaan/Akta_574_%2831.5.2023%29_BI.pdf

Sanghvi, H. (2022). Application of artificial intelligence in the sentencing process. In *Proceedings of the International Conference on Law and Digital Governance*. Atlantis Press. https://doi.org/10.2991/assehr.k.220105.012

Subordinate Courts Act 1948 (Act 92) (Malaysia). (2025). https://tcclaw.com.my/wp-content/uploads/2025/08/Subordinate-Courts-Act-1948.pdf

Toh, J., Lee, A., & Partners. (2023). Artificial intelligence and judicial decision-making: Emerging legal challenges in common law jurisdictions. *Legal Practice Review*, *15*(2), 45–68. https://ssm-italia.eu/wp-content/uploads/2025/02/JuLIA_handbook-Justice_final.pdf

Verma, S., & Rubin, J. (2018). Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness* (pp. 1–7). https://doi.org/10.1145/3194770.3194776

Zainol, Z., Abdul Rahman, N. A., & Mohamed, M. F. (2023). Artificial intelligence-powered criminal sentencing in Malaysia. *Environment-Behaviour Proceedings Journal*, *8*(24), 89–96. https://ebpj.e-iph.co.uk/index.php/EBProceedings/article/view/3813/2118

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum*, *20*(4), 567–583. https://doi.org/10.1007/s12027-020-00602-0

Zilka, M., Fogliato, R., Hron, J., Butcher, B., Ashurst, C., & Weller, A. (2023). The progression of disparities within the criminal justice system: Differential enforcement and risk assessment instruments. *arXiv*. https://arxiv.org/abs/2305.07575

Ziosi, M., & Pruss, D. (2024). Evidence of what, for whom? The socially contested role of algorithmic bias in a predictive policing tool. *arXiv*. https://arxiv.org/abs/2405.07715

Acknowledgment

The authors would like to express their sincere appreciation to the Faculty of Law, Universiti Teknologi MARA (UiTM) and all individuals who contributed directly or indirectly to the completion of this research.

Competing Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Generative AI Statement

During the preparation of this manuscript, the authors used ChatGPT (OpenAI GPT-5.5) as a language assistance tool to improve vocabulary and clarity. All generated content was subsequently reviewed, verified, edited, and approved by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the final manuscript.