The Use of Artificial Intelligence in Death Penalty Decision-Making: An Analysis Based on Islamic Ethics and Maqāṣid al-Sharīʿah

Nursolehah Mat Razali
Universiti Malaya, Kuala Lumpur, Malaysia
Email: nursolehahmatrazali@gmail.com

Abstract

The increasing use of artificial intelligence (AI) in criminal justice and sentencing raises significant ethical concerns, particularly when algorithmic systems are applied to decisions involving the deprivation of human life. While AI may enhance efficiency, consistency, and access to legal information, its use in capital sentencing presents risks related to inaccurate data, algorithmic bias, opacity, automation bias, limited explainability, and the displacement of human responsibility. This study aims to examine the ethical implications of using AI in determining the death penalty from the perspectives of Islamic ethics and *maqāṣid al-sharīʿah*, and to develop a normative framework for responsible AI-assisted sentencing. The study adopts a doctrinal, normative, and conceptual research design based on qualitative analysis of primary Islamic sources, scholarly literature, legal and policy documents, and contemporary AI governance frameworks. The findings indicate that the use of AI in capital sentencing must be constrained by the principles of *ʿadl* (justice), *tabayyun* (verification), *amānah* (trust and responsibility), *masʾūliyyah* (accountability), *karāmah al-insān* (human dignity), *ḥifẓ al-nafs* (protection of life), and *dafʿ al-ḍarar* (prevention of harm). The study proposes the Islamic AI Death Penalty Decision Framework (IADPDF), comprising six layers: purpose, data integrity, algorithmic fairness, explainability and transparency, meaningful human oversight, and *maqāṣid* assessment. The study concludes that AI should function only as a decision-support instrument, while the final determination of capital punishment must remain with human judges. Accordingly, a presumption against AI-driven capital sentencing is proposed.

Keywords: artificial intelligence, death penalty, islamic ethics, maqāṣid al-sharīʿah, human oversight

Introduction

The development of artificial intelligence (AI) has brought significant changes across various sectors, including law enforcement, judicial administration, and sentencing systems. The ability of AI to process large volumes of data, identify patterns, classify information, and generate predictions has encouraged its use in criminal risk assessment, analysis of criminal records, and decision-support in sentencing. In the context of criminal justice, the use of AI is often associated with the potential to improve efficiency, consistency, and uniformity in decision-making processes. However, the use of such technology also raises concerns regarding accuracy, algorithmic bias, transparency, and accountability (*Berk,* 2021). The ethical concerns surrounding AI in the judicial system are not limited to the technological ability to produce accurate predictions,

but also extend to whether AI-generated decisions can be regarded as fair and accountable. Berk *et al.* (2021) demonstrate that the concept of *fairness* in criminal risk assessment has multiple interpretations and that different measures of fairness may conflict with one another as well as with predictive accuracy. This indicates that statistical accuracy does not necessarily result in substantive justice. A system may produce technically consistent outcomes while still potentially generating disproportionate effects on certain individuals or groups.

These concerns become more complex when AI-generated decisions influence sentencing processes. The data used by algorithmic systems may reflect existing social and institutional biases or imbalances. Consequently, AI has the potential to reproduce or reinforce existing patterns of injustice. At the same time, certain AI systems may be difficult for end users to understand because the processes through which outputs are generated can be highly complex or insufficiently transparent. This raises questions about the ability of relevant parties to understand, evaluate, and challenge decisions influenced by algorithms. Reliance on AI recommendations also creates the risk of *automation bias*, whereby humans may tend to accept automated outputs as more objective or accurate than their own judgment. In the judicial context, this risk is particularly significant because the mere presence of a human decision-maker does not necessarily guarantee meaningful human oversight. The more critical question is whether judges possess the knowledge, independence, and actual authority to evaluate, challenge, or reject AI-generated outputs. *Vogt* (2026) demonstrates that ethical concerns surrounding AI in sentencing encompass not only accuracy and fairness but also transparency, justification, responsibility, human dignity, and the legitimacy of decision-making.

The need for human oversight becomes even more important when AI is used in decisions involving serious and irreversible consequences. UNESCO (2021), through its *Recommendation on the Ethics of Artificial Intelligence*, emphasizes the principles of transparency, accountability, harm prevention, and human oversight. UNESCO further stresses that decisions involving irreversible consequences, particularly decisions concerning life and death, should not be delegated to AI systems. A similar approach can be observed in Regulation (EU) 2024/1689, or the *Artificial Intelligence Act*, which places emphasis on human oversight for high-risk AI systems.

Capital Punishment as a High-Risk Judicial Decision

The question of AI use becomes particularly critical when applied to capital punishment. Unlike other forms of punishment that may still allow opportunities for review or correction, the execution of a death sentence carries consequences that are almost entirely irreversible. An error in determining a death sentence is therefore not merely a technical error but may result in the irreversible loss of human life. Accordingly, the use of AI in this context requires a higher level of scrutiny, evidentiary verification, and human control than its use in lower-risk judicial tasks.

The distinction between using AI as a support tool and using it as a substantive decision-maker also requires careful consideration. AI may be used for tasks such as information retrieval, document management, analysis of previous cases, or identification of inconsistencies in sentencing. However, the use of AI to determine culpability, assess an individual's moral worth, predict future criminal risk, or determine whether the death penalty should be imposed involves

considerably more serious ethical concerns. Thus, the greater the risk and irreversibility of the consequences associated with a decision, the stronger the need for meaningful human oversight and accountability.

Islamic Ethics and *Maqāṣid al-Sharīʿah*

From an Islamic perspective, the use of AI in determining capital punishment cannot be evaluated solely on the basis of technical efficiency or accuracy. It must also be assessed according to Islamic principles concerning justice, the preservation of life, responsibility, harm, and human dignity. In this regard, *maqāṣid al-sharīʿah* provides an important normative framework for evaluating the objectives, benefits, and consequences of technological applications. The principle of ḥifẓ al-nafs, or the preservation of life, is particularly significant in the context of capital punishment. Although the criminal justice system aims to protect society and prevent harm, the preservation of life also requires that an individual's life should not be taken on the basis of insufficient evidence or an unaccountable decision-making process. The principle of ʿadl, or justice, similarly requires judicial decisions to be free from bias, discrimination, and irrelevant considerations. This principle has a direct connection with the issue of *algorithmic fairness* in AI systems.

Furthermore, the principle of tabayyun emphasizes the need to verify information before making a decision. This principle is particularly relevant when AI outputs are incorporated into judicial processes because such outputs depend on the quality of the underlying data, system design, and parameters employed. AI-generated outputs should therefore not be treated as inherently truthful or authoritative but should be examined and verified by human actors possessing appropriate authority. The principles of amānah and masʾūliyyah further indicate that responsibility for a decision cannot be transferred to technology, while karāmah al-insān emphasizes that human beings should not be reduced merely to risk scores or statistical profiles. The principle of dafʿ al-ḍarar, in turn, requires potential risks and harms arising from the use of AI to be minimized.

This approach is consistent with the growing body of scholarship on Islamic approaches to AI ethics. Raquib et al. (2022), for example, demonstrate that *maqāṣid al-sharīʿah* can provide a foundation for developing an AI ethics framework that emphasizes human values, welfare, and moral responsibility. Nevertheless, the specific application of these principles to the use of AI in capital punishment requires further systematic examination, particularly with regard to the limits of AI autonomy and human responsibility for decisions that have consequences for human life.

Research Gap

Although the literature on AI in criminal justice has expanded considerably, much of the existing research continues to focus on issues such as predictive accuracy, risk assessment, algorithmic bias, and system transparency. While these issues are important, they do not fully address the normative question of using AI in the most severe form of sentencing, namely capital punishment (Berk, 2021; Berk et al., 2021). At the same time, scholarship on AI ethics and *maqāṣid al-sharīʿah* has also developed, but the specific application of these principles to decisions affecting human life remains relatively limited (Raquib et al., 2022). This gap is particularly relevant in the Malaysian context, where the use of AI in sentencing has begun to receive institutional attention

within the judiciary. The Malaysian Judiciary has developed guidelines concerning the use of AI to assist judges in sentencing, with AI serving as a source of guidance and reference rather than as a replacement for judicial decision-making. This development demonstrates that the use of AI in sentencing is no longer merely a theoretical issue but requires closer examination of the boundaries of such technology when decisions involve substantially higher levels of risk.

Accordingly, there is a need to integrate perspectives from AI ethics, judicial ethics, and *maqāṣid al-sharīʿah* to establish a normative basis for determining the limits of AI use in capital punishment. This study adopts the premise that AI may function as a *wasīlah*, or supporting instrument, for human decision-makers, but that its use as an autonomous determinant of culpability, an individual's moral worth, or the imposition of capital punishment raises substantially greater ethical concerns. Based on this premise, the study proposes the principle of presumption against AI-driven capital sentencing, whereby AI should be presumed unsuitable for exercising direct decision-making authority over the imposition of capital punishment.

This study aims to examine the ethical implications of using artificial intelligence (AI) in determining capital punishment from the perspective of Islamic ethics and *maqāṣid al-sharīʿah*. It specifically seeks to evaluate whether and to what extent AI may be used as a judicial support tool in capital punishment cases while preserving the principles of justice (*ʿadl*), protection of life (*ḥifẓ al-nafs*), trust and responsibility (*amānah* and *masʾūliyyah*), verification (*tabayyun*), human dignity (*karāmah al-insān*), and prevention of harm (*dafʿ al-ḍarar*).

Methods

This study employs a qualitative research design incorporating doctrinal, normative, and conceptual approaches to examine the ethical implications of artificial intelligence (AI) in capital sentencing from the perspectives of Islamic ethics and maqāṣid al-sharīʿah. The research is based on documentary analysis of primary and secondary sources, including the Qur'an, classical works of Islamic jurisprudence, scholarly literature, relevant legal instruments, and contemporary AI governance frameworks. Particular emphasis is placed on the principles of *ʿadl* (justice), *tabayyun* (verification), *amānah* (trust), *masʾūliyyah* (accountability), *karāmah al-insān* (human dignity), *ḥifẓ al-nafs* (preservation of life), and *dafʿ al-ḍarar* (prevention of harm). The works of al-Ghazālī and al-Shāṭibī provide the theoretical foundation for examining the objectives of Islamic law and the protection of essential human interests. Contemporary governance instruments, particularly UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021), Regulation (EU) 2024/1689 (Artificial Intelligence Act), and the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment(2018), are examined to identify relevant standards of fairness, transparency, accountability, and human oversight. Peer-reviewed studies on algorithmic bias, criminal risk assessment, AI-assisted sentencing, and Islamic approaches to AI ethics are also reviewed to contextualise the analysis within contemporary scholarly debates.

The analysis is conducted in three stages. First, a doctrinal examination identifies the normative foundations and implications of the selected Islamic ethical principles for judicial decision-making. Second, a comparative normative analysis evaluates the relationship between

these principles and contemporary AI governance requirements, with particular attention to data reliability, algorithmic fairness, explainability, accountability, and meaningful human oversight. This stage also distinguishes between administrative or informational applications of AI and applications that may substantively influence assessments of culpability, criminal risk, or sentencing outcomes. Third, the findings are synthesised to develop the proposed Islamic AI Death Penalty Decision Framework (IADPDF), comprising six interrelated dimensions: purpose of use, data integrity, algorithmic fairness, explainability and transparency, meaningful human oversight, and maqāṣid assessment. The framework is intended to establish normative criteria for evaluating the ethical acceptability and limitations of AI-assisted decision-making in capital cases. As a doctrinal and conceptual study, the research does not involve empirical data collection, experimental testing, or the development of an AI system. Accordingly, the proposed framework is presented as a normative contribution rather than an empirically validated model. Its application and practical effectiveness remain matters for future empirical investigation.

Discussion

AI in Sentencing Is Not Merely a Question of Technological Efficiency

The use of artificial intelligence (AI) in judicial systems is commonly associated with improved efficiency, consistency, and the ability to process large volumes of data within a short period. In the context of sentencing, AI has the potential to assist judges in identifying patterns in previous cases, structuring case-related information, comparing sentencing decisions, and generating recommendations based on historical data. However, the central question is not merely whether AI is capable of producing technically consistent or accurate outcomes, but whether such outcomes can be regarded as fair, legitimate, and accountable from both moral and legal perspectives. *Vogt* (2026), for example, identifies various ethical concerns associated with the use of AI in sentencing, including issues of accuracy, fairness, transparency, justification, responsibility, dignity, and legitimacy.

This issue becomes more complex because sentencing is not merely a mechanical process of calculation. Sentencing involves consideration of the facts of the case, the severity of the offence, the circumstances of the individual, mitigating and aggravating factors, the purposes of punishment, and the applicable principles of justice. Therefore, AI performance cannot be assessed solely on the basis of predictive accuracy or its ability to produce outcomes that closely resemble those of previous judicial decisions. *Ryberg* (2024, 2025a) argues that the question of whether algorithms can replace human beings in sentencing depends on normative considerations concerning what makes a punishment appropriate and just. In other words, algorithmic performance must be evaluated not only from a technical perspective but also in light of the moral and philosophical objectives of punishment.

In this context, an important distinction must be made between AI as an assistive tool and AI as a decision-maker. The use of AI to search previous cases, organise documents, or identify inconsistencies in sentencing decisions may still preserve the judge as the primary decision-maker. By contrast, the use of AI to generate risk scores or sentencing recommendations that effectively

determine the final outcome shifts the position of AI from a supporting instrument to a substantive actor in the judicial process. *Vogt* (2026) demonstrates that certain ethical concerns become more pronounced when AI is deployed autonomously rather than functioning merely as an advisory tool for judges. From an Islamic perspective, this distinction is consistent with the principle that decision-making authority involving human rights and life must be exercised through human trust, deliberation, and responsibility. The concept of *amānah* requires decision-making authority to remain connected to the party responsible for the consequences of the decision. The Qur'an emphasizes the obligation to fulfil trusts to those entitled to them and to judge between people with justice (*Qur'an, al-Nisāʾ 4:58*). Accordingly, technological efficiency cannot serve as the sole criterion for determining the appropriateness of AI in judicial processes.

In the context of capital sentencing, this consideration becomes even more critical because an erroneous decision may be irreversible. Therefore, the greater the potential harm and irreversibility of a decision, the greater the need for factual verification, human deliberation, and ethical justification. This approach leads to a fundamental principle of this study: AI may assist the judicial process, but it should not assume human responsibility for decisions that determine whether an individual lives or dies (*UNESCO*, 2021; *Vogt*, 2026).

Ḥifẓ al-Nafs and the Risk of Irreversible Error

The principle of *ḥifẓ al-nafs*, or the preservation of life, constitutes one of the core considerations within the framework of *maqāṣid al-sharīʿah*. The maqāṣid tradition articulated by al-Ghazālī and further developed by al-Shāṭibī places the preservation of fundamental human interests among the important objectives of the Sharīʿah (al-Ghazālī, n.d.; al-Shāṭibī, n.d.). In the context of the death penalty, this principle does not necessarily mean that every form of capital punishment must be categorically rejected. Rather, it requires an exceptionally high degree of caution when a judicial decision has the potential to result in the loss of human life. An important distinction between the death penalty and other forms of punishment lies in its irreversible nature. If a person is wrongfully imprisoned, release, compensation, or certain forms of redress may still be possible. By contrast, if a person is executed on the basis of an erroneous decision, no form of remedy can genuinely restore the original state of affairs. Therefore, risk assessment in this context is inadequate if it considers only the probability of error. The magnitude of the consequences arising from such an error must also be taken into account.

In AI systems, the risk of error may arise at multiple stages, including incomplete data, historical data containing bias, inappropriate variable selection, model errors, and the use of outputs beyond their original context. Berk (2021) and Berk et al. (2021) demonstrate that predictive and risk-assessment systems in criminal justice raise significant concerns regarding accuracy and fairness. In the context of sentencing, these problems become more complex when the historical data used to train a system reflect patterns of human decision-making that may themselves be affected by bias. *Ryberg* (2025b) further highlights what is known as the *input problem*, namely the question of what information should properly constitute the inputs to an algorithm when a sentence is being determined. This issue is significant because an algorithm cannot produce a decision that is more meaningful than the information upon which that decision

is based. If the information provided is incomplete, irrelevant, or contains problematic normative assumptions, an output that appears technically objective may nevertheless produce an unjust decision.

From a *maqāṣid* perspective, the principle of *dafʿ al-ḍarar*, or the prevention of harm, reinforces the need to adopt preventive measures before technology is employed in high-risk decisions. Where an error may result in an irreversible loss of life, the requirement of prevention should not be deferred until an error has already occurred. The principle of *sadd al-dharāʾiʿ* may likewise be applied conceptually to assess whether a particular technological use that is permissible in itself may potentially open the way to greater harm. Nevertheless, the principle of *ḥifẓ al-nafs* does not mean that all uses of AI in processes related to sentencing should be prohibited. Rather, there is a need to distinguish between administrative or informational uses of AI and uses that exert substantive influence on sentencing decisions. The use of AI to manage documents or retrieve previous case decisions entails a different level of risk from the use of AI to determine the severity of an offence or recommend the death penalty.

Accordingly, the findings of this study support a principle of proportionality between the level of risk and the level of human control: the more serious and irreversible the consequences of a decision, the stronger the requirements for human control, evidentiary verification, and ethical justification. This principle is consistent with contemporary AI governance approaches that require human oversight to be strengthened in accordance with the level of systemic risk (*UNESCO*, 2021; *European Union*, 2024).

Tabayyun as an Epistemic Principle in Assessing AI Outputs

One of the major challenges in using AI for sentencing is the tendency to regard algorithmic outputs as objective simply because they are generated through computational processes. In reality, AI outputs depend on the quality of the data, model design, selected variables, training methods, and assumptions incorporated into system development. Therefore, an AI output cannot automatically be accepted as fact or truth. In Islamic ethics, the principle of *tabayyun* provides an important basis for the verification of information. The Qur'an instructs believers to investigate information received so as to avoid making erroneous decisions on the basis of unverified information (Qur'an, al-Ḥujurāt 49:6). This principle has direct implications for the use of AI in judicial processes because an algorithmic output is, in essence, a form of information that must still be interpreted and verified before it is relied upon in decision-making.

The concept of *tabayyun* is also closely connected to the Qur'anic principle of justice. Judges are required not to allow perceptions, assumptions, or particular inclinations to influence decisions in an unjust manner (Qur'an, al-Nisāʾ 4:135; al-Māʾidah 5:8). In the context of AI, this requirement means that judicial decisions cannot rely solely on algorithmic scores, classifications, or recommendations without examining the informational basis from which those outputs were generated. This issue is particularly important in risk-assessment systems. For example, a high-risk score does not necessarily indicate that an individual is morally more culpable or deserving of a harsher punishment. Such a score may merely reflect statistical patterns present in the training

data. The distinction between statistical correlation and moral and legal justification must therefore be maintained clearly (*Ryberg,* 2025b; *Vogt*, 2026).

In this regard, *tabayyun* can be translated into several technical and procedural requirements, including verification of data sources, validation of input accuracy, documentation of processes, examination of the variables used, and assessment of model limitations. AI outputs should also be open to challenge by relevant parties when such outputs are relied upon in judicial proceedings. The principles of transparency and auditability emphasised in contemporary AI governance have functional parallels with the requirement of *tabayyun*, although the two derive from different normative traditions (*Elmahjub*, 2023; UNESCO, 2021). Thus, *tabayyun* should not be understood merely as an obligation to verify whether data are factually correct. In the context of AI, it also encompasses questions concerning how an output was generated, what its limitations are, what assumptions it contains, and whether the output is appropriate for a particular purpose. This approach helps prevent *automation bias*, whereby humans place excessive reliance on system recommendations simply because they appear objective or technologically derived.

Amānah, Masʾūliyyah and the Need for Meaningful *Human-in-the-Loop* Oversight

The use of AI in judicial processes raises a fundamental question concerning who bears responsibility for decisions generated with AI assistance. If judges merely accept AI recommendations without conducting an independent assessment, there is a risk that substantive responsibility may be indirectly transferred to a system that lacks the moral or legal capacity to bear such responsibility. *Vogt* (2026) identifies accountability as one of the major challenges associated with AI in sentencing. From an Islamic perspective, the concepts of *amānah* and *masʾūliyyah* require decision-making authority to be exercised responsibly. Judges cannot relinquish their responsibility merely by stating that a decision was supported or recommended by an AI system. Responsibility remains with the human authority entrusted with the power to make that decision. This principle is consistent with the approach of *UNESCO* (2021), which emphasises that AI must not replace human responsibility and accountability, particularly in decisions concerning life and death.

Accordingly, *human-in-the-loop* should not be understood in a minimal sense as merely placing a human at the final stage of the process to press an "agree" button. Meaningful human oversight requires human decision-makers to possess the knowledge, authority, and genuine opportunity to assess, challenge, or reject AI outputs. This approach is also consistent with the requirements of *human oversight* under Regulation (EU) 2024/1689, which emphasises that high-risk AI systems should be designed so that human oversight can prevent or minimise risks to safety, health, and fundamental rights (*European Union*, 2024). In the context of the death penalty, these requirements must be strengthened further. Judges should not merely be aware that an AI system has generated a particular recommendation; they must also understand the basis of that recommendation, assess the reliability of the inputs, consider factors that cannot be adequately represented by the model, and retain full authority to reject the AI recommendation. The final decision must therefore arise from human judicial deliberation rather than from automatic adherence to an algorithmic score or recommendation.

On this basis, this study proposes the principle of AI-assisted but human-determined sentencing, whereby AI may provide analytical assistance, but substantive sentencing decisions must be determined by human beings who possess judicial authority. This principle is more consistent with the requirements of *amānah* and *masʾūliyyah* than an *AI-driven sentencing* model that grants substantive decision-making power to algorithms.

Karāmah al-Insān and the Danger of Reducing Individuals to Risk Scores

Another important issue concerns the impact of AI use on *karāmah al-insān*, or human dignity. The Qur'an affirms that human beings have been honoured and granted a position of distinction (*Qur'an, al-Isrāʾ* 17:70). This principle has significant implications when AI systems are used to assess individuals within criminal justice processes. The primary risk arises when an individual is reduced to a collection of data, categories, or risk scores. In algorithmic assessment systems, individuals may be evaluated according to particular variables, such as criminal history, social background, or other factors incorporated into the model. Although such variables may possess statistical predictive value, they do not necessarily represent the entirety of an individual's identity, intentions, circumstances, or moral worth. *Vogt* (2026) associates the use of AI in sentencing with concerns regarding human dignity and the risk of *dehumanization*. In certain circumstances, technology may transform the relationship between the judge and the accused from a process of human deliberation into one of classification based on numerical values and probabilities. This issue is particularly significant in sentencing because individuals appearing before a court are not merely data objects; they are human beings possessing rights and dignity whose circumstances must be considered as individuals.

The principle of *karāmah al-insān* does not mean that statistical data cannot be used. Rather, it prevents data from becoming a substitute for human judgment that takes into account the uniqueness of an individual's circumstances. Accordingly, a risk score must not be interpreted as a measure of a person's moral worth or used as a substitute for judicial assessment of culpability and punishment. This approach also demonstrates why AI must be used cautiously even where a system has demonstrated high statistical predictive accuracy. Predictive accuracy does not, in itself, make a decision just. A model may accurately predict a particular outcome while remaining inappropriate as a basis for sentencing if the variables it employs conflict with principles of justice, dignity, or individual rights (*Berk et al.,* 2021; *Ryberg*, 2025a). Therefore, *karāmah al-insān* functions as a normative limit on the use of technology. AI may assist human beings in understanding information about an individual, but it must not transform a person into merely a "risk profile." In the context of the death penalty, this limitation becomes even more important because reducing an individual to a score may contribute to a decision whose consequences cannot be reversed.

Explainability as a Condition of Procedural Justice

Transparency and *explainability* are among the important requirements for the use of AI in judicial systems. If a system generates a sentencing recommendation but the judge, accused, or defence counsel cannot understand the basis of that recommendation, this raises concerns regarding the ability to challenge the decision and assess its justification. *Ryberg* (2025a) and *Vogt* (2026)

identify transparency and justification as among the major challenges associated with AI in sentencing. In judicial proceedings, a decision must not only be correct; it must also be capable of being justified. Judges must be able to explain why a particular decision has been made on the basis of the relevant facts, law, and considerations. If part of the reasoning underlying a decision originates from an AI system that cannot be adequately explained, there is a risk that the justificatory structure of the decision will be weakened.

In this context, *explainability* is not merely a technical question concerning how an algorithm operates. It is also a question of procedural justice. Individuals affected by a decision should have a reasonable opportunity to understand the basis of that decision and to challenge it. UNESCO (2021) similarly emphasises the importance of transparency, auditability, and explainability in ensuring that AI-assisted decisions can be understood and subjected to accountability.bThe principle of *tabayyun* provides a corresponding basis within Islamic ethics. If an AI output is used as part of the decision-making process, the basis of that output must be capable of examination. At the same time, the principle of *masʾūliyyah* requires decision-makers to possess sufficient understanding of the reasons supporting the decision for which they are responsible.

Nevertheless, *explainability* does not necessarily mean that every judge must understand every technical aspect of an AI model. More importantly, there must be sufficient explanation to understand the principal inputs, the factors influencing the output, the limitations of the system, and the reasons for a particular recommendation. In high-risk cases, the system should also maintain an audit trail that enables the use of AI throughout the decision-making process to be traced and reviewed. Accordingly, in death penalty cases, an AI system that cannot provide an adequate level of transparency and auditability should not serve as a substantive basis for the decision. This requirement does not arise because technology is categorically unsuitable for judicial use, but because the level of transparency required must be proportionate to the level of risk and the consequences of the decision.

Presumption Against AI-Driven Capital Sentencing

Based on a synthesis of the principles of *ḥifẓ al-nafs*, *ʿadl*, *tabayyun*, *amānah*, *masʾūliyyah*, *karāmah al-insān*, and *dafʿ al-ḍarar*, this study proposes a normative principle termed the presumption against AI-driven capital sentencing. This principle means that there should be a presumption against allowing AI to autonomously make or determine capital sentencing decisions. This principle does not amount to a prohibition of all uses of AI in cases involving the death penalty. Rather, it distinguishes between the use of AI for supportive functions and its use for substantive decision-making. AI may still be used for legal research, document management, comparison of judicial decisions, detection of inconsistencies, or certain analytical functions, subject to appropriate safeguards. What is rejected is the allocation of decision-making authority to an algorithm to determine whether an individual should receive the death penalty.

This presumption is supported by three principal considerations. First, the death penalty carries irreversible consequences. Second, AI systems depend on inputs and system designs that may contain errors or bias. Third, sentencing is not merely a predictive decision but involves normative judgments concerning culpability, justice, responsibility, and human dignity (*Ryberg*,

2025a, 2025b; Vogt, 2026). The development of debates concerning the automation of decisions involving life and death further reinforces these concerns. *AbuMusab* et al. (2026) highlight the dangers that arise when decisions leading to the execution of a death sentence become connected to autonomous mechanisms. Although this issue involves a more extreme level of automation than AI-assisted sentencing, the argument remains relevant in demonstrating that the further human beings are removed from the decision-making chain, the greater the concerns regarding accountability and human control.

This approach is also consistent with the UNESCO (2021) principle that decisions concerning life and death should not be entrusted to AI. Regulation (EU) 2024/1689 likewise emphasises the need for *human oversight* in high-risk AI systems and recognises risks to fundamental rights, safety, and individual interests (*European Union*, 2024). Accordingly, the principle may be formulated as follows: AI may assist judges in understanding information, but AI must not become the entity that determines the value of a person's life or replace human judicial deliberation in capital sentencing decisions. Where there is uncertainty concerning the accuracy, appropriateness, or explainability of an AI output, the decision should not automatically move towards a more severe punishment.

Comparison between Islamic Ethics and Contemporary AI Governance

A comparison between Islamic ethics and contemporary AI governance frameworks reveals significant areas of convergence, particularly concerning justice, the prevention of harm, human dignity, transparency, and accountability. However, such similarities do not mean that the two approaches share identical epistemological and normative foundations. Islamic ethics is rooted in revelation, Sharīʿah principles, and the Islamic intellectual tradition, whereas contemporary AI governance frameworks generally develop from principles of human rights, technology ethics, the rule of law, and the protection of public interests. Within Islamic ethics, the principle of *ʿadl* requires justice in treatment and decision-making; *tabayyun* requires the verification of information; *amānah* and *masʾūliyyah* emphasise responsibility; *karāmah al-insān* safeguards human dignity; while *ḥifẓ al-nafs* and *dafʿ al-ḍarar* provide a basis for preventing harm to human life and well-being (*al-Ghazālī, n.d.; al-Shāṭibī, n*.d.). In contemporary AI governance frameworks, principles such as fairness, transparency, explainability, accountability, human oversight, and do no harm serve broadly corresponding practical objectives (*UNESCO,* 2021; *European Union,* 2024).

Studies of AI ethics from an Islamic perspective also indicate that the Islamic ethical tradition has the potential to contribute to the development of broader ethical benchmarks for AI. *Elmahjub* (2023) emphasises the importance of developing AI ethical benchmarks that take Islamic sources and values into account, while *Raquib et al*. (2022) demonstrate that Islamic value- and virtue-based ethics can be applied to assess issues relating to the development and use of AI. Nevertheless, there are differences in the foundations of justification. The principle of *ḥifẓ al-nafs*, for example, is not merely a principle of technical safety. It concerns the status and value of human life within the framework of the objectives of the Sharīʿah. Similarly, *amānah* is not merely an administrative accountability mechanism, but a moral responsibility associated with human

obligations towards God and society. Therefore, the application of Islamic principles to AI governance should not be undertaken through mechanical matching of terms, but through substantive analysis of the purpose and implications of each principle.

This comparison produces a form of complementarity without equivalence. Contemporary AI governance frameworks can provide institutional mechanisms such as auditing, documentation, human oversight, and impact assessment. Islamic ethics, meanwhile, can provide a normative foundation for explaining why human beings, dignity, life, justice, and the prevention of harm should be treated as priorities. The combination of these approaches can produce a stronger basis for assessing the use of AI in the context of capital sentencing. Within this framework, AI should not be evaluated solely by asking, "*Does the system work?"* but also by asking, *"Is the use of the system just?",* "*Has the information been verified*?", "*Who bears responsibility?*", "*Is human dignity preserved?*", and "*What are the implications of its use for human life?*" These questions provide the foundation for the conceptual framework proposed in this study.

Islamic AI Death Penalty Decision Framework (IADPDF) as a Conceptual Contribution

Based on a synthesis of doctrinal analysis, Islamic ethics, and AI governance, this study proposes the Islamic AI Death Penalty Decision Framework (IADPDF) as a normative framework for assessing the use of AI in processes related to the death penalty. IADPDF is not an algorithm for determining whether a person should be sentenced to death. Rather, it constitutes an ethical and governance screening framework for defining the boundaries of AI use before and during the deployment of such technology in judicial processes.

The framework is constructed around six interrelated layers: purpose of use, data integrity, algorithmic fairness, explainability and transparency, meaningful human oversight, and maqāṣid assessment.

Table 3 Islamic AI Death Penalty Decision Framework (IADPDF)

No	Framework Layer	Key Question	Islamic Ethical Principle	Governance Requirement
1	**Purpose**	Why is AI being used, and what objective is it intended to achieve?	*Maṣlaḥah*, *ʿadl*	Justification of the purpose of use and assessment of necessity
2	**Data Integrity**	Are the data accurate, valid, relevant, and sufficient?	*Tabayyun*, *amānah*	Data verification, source documentation, and auditing
3	**Algorithmic Fairness**	Does the system produce bias or discrimination?	*ʿAdl*	Bias auditing, fairness testing, and review of variables
4	**Explainability & Transparency**	Can the output be understood, traced, and challenged?	*Tabayyun*, *masʾūliyyah*	Explainability, traceability, and audit trail
5	**Meaningful Human Oversight**	Who makes the final decision, and can humans reject the AI output?	*Amānah*, *masʾūliyyah*	Human review, override authority, and the right to reject AI recommendations
6	**Maqāṣid Assessment**	What are the effects of AI use on human life, dignity, and interests?	*Ḥifẓ al-nafs*, *karāmah al-insān*, *dafʿ al-ḍarar*	Maqāṣid impact assessment and final human decision

The first layer, Purpose, functions as an initial screening mechanism. Before AI is deployed, judicial authorities must determine the actual purpose of its use. AI should not be

introduced merely because it is available or perceived as more technologically advanced. Its use must have a clear and reasonable purpose and provide a meaningful benefit to the judicial process. The principle of *maṣlaḥah* requires the benefits of technological use to be assessed against the risks it may create (*al-Ghazālī, n.d.; al-Shāṭibī, n.d.).* The second layer is Data Integrity. At this stage, attention is given to the accuracy, validity, relevance, and sufficiency of the data. The principle of *tabayyun* requires information to be verified before it is used as a basis for decision-making, while *amānah* requires that data not be manipulated or used beyond their legitimate purpose. This requirement is particularly significant because problems at the input stage may affect the entire AI output (*Ryber*g, 2025b).

The third layer, Algorithmic Fairness, assesses whether the system produces biased outcomes or discriminates against particular groups. The principle of *ʿadl* provides the normative basis for ensuring that the use of technology does not produce injustice. Fairness audits and bias assessments are therefore necessary because a system with high statistical accuracy does not necessarily produce fair outcomes for all individuals (*Berk et al*., 2021). The fourth layer is Explainability & Transparency. AI outputs must be sufficiently understandable to enable judges and relevant parties to identify the basis of the recommendations provided. Audit records, documentation, and traceability mechanisms should be maintained so that the process of AI use can be examined if the resulting decision is challenged. This requirement is consistent with the emphasis on transparency and auditability in international AI ethics frameworks (*UNESCO*, 2021).

The fifth layer, Meaningful Human Oversight, is the most critical component in death penalty decisions. Human beings must not merely be present in the process; they must possess substantive authority to assess and reject AI outputs. The principles of *amānah* and *masʾūliyyah* preclude the transfer of moral responsibility to an algorithm. Accordingly, judges must remain the final decision-makers and must not be automatically bound by AI recommendations (*European Union,* 2024; *UNESCO*, 2021).The final layer is Maqāṣid Assessment, which functions as a comprehensive assessment of the effects of AI use on human beings. At this stage, the question is no longer merely whether the system is technically accurate or fair, but whether its use is consistent with the preservation of life, human dignity, and human interests. In death penalty cases, *ḥifẓ al-nafs* and *daf*ʿ *al-ḍarar* provide strong grounds for establishing the highest threshold of control.

The six layers should not be regarded as independent procedures. Rather, they constitute an interconnected chain of safeguards. Failure at any one layer may undermine the overall suitability of AI use. For example, an algorithm that is highly accurate in technical terms may nevertheless be unsuitable as a basis for a death penalty decision if the data used are invalid, the process cannot be adequately explained, or the judge lacks the authority to reject the output. For this reason, IADPDF does not propose that AI must "*make better decisions than humans*" before it can be permitted to determine the death penalty. Instead, the framework establishes that the burden of justification rests on those seeking to use AI in high-risk decisions. The greater the risk to human life and individual rights, the higher the level of evidence, transparency, oversight, and justification that must be satisfied. The principal conceptual contribution of IADPDF lies in integrating Islamic ethical principles with contemporary AI governance mechanisms within the

specific context of the death penalty. The framework connects *ʿadl* with algorithmic fairness, *tabayyun* with data verification and explainability, *amānah* and *masʾūliyyah* with human oversight, and *ḥifẓ al-nafs*, *karāmah al-insān*, and *dafʿ al-ḍarar* with assessments of the impact on human life and dignity. These relationships do not mean that each concept carries an identical meaning, but rather demonstrate how the two traditions can complement one another in establishing boundaries for AI use (*Elmahjub*, 2023; *Raquib et al*., 2022; *UNESCO*, 2021).

Accordingly, the final model proposed by this study is AI-assisted but human-determined capital sentencing. AI may assist in controlled support functions, but the decision to impose the death penalty must remain a human judicial decision. This approach also reinforces the principle of presumption against AI-driven capital sentencing, whereby AI should not be granted the authority to autonomously determine whether a human being should lose their life. This principle constitutes a normative conclusion derived from the combined requirements of *ʿadl*, *tabayyun*, *amānah*, *masʾūliyyah*, *karāmah al-insān*, *ḥifẓ al-nafs*, and *dafʿ al-ḍarar*. Overall, IADPDF emphasises that the central question is not whether AI can be used within judicial systems, but where the boundaries of AI use should be drawn when a decision affects the most fundamental and irreversible rights. In the context of the death penalty, these boundaries should be set at the strictest level. Technology may strengthen the processes of information retrieval and assessment, but it must not replace the moral judgment, legal responsibility, and human accountability that constitute the foundation of judicial decision-making.

Conclusion

This study demonstrates that the use of artificial intelligence (AI) in capital sentencing cannot be evaluated solely on the basis of its technological capacity to improve efficiency, consistency, or decision-making accuracy. In the context of punishment involving the loss of human life, the more fundamental question is whether the use of AI is consistent with the principles of justice, responsibility, human dignity, and the preservation of life. The analysis indicates that although AI has potential as a supportive tool within judicial systems, its use as an autonomous determinant of capital sentencing raises ethical and normative risks that are difficult to justify. The findings reveal several major risks that require particular attention, including inaccurate data, algorithmic bias, opacity in decision-making processes, *automation bias*, the use of risk scores, and the potential transfer of responsibility from human decision-makers to AI systems. These risks become more serious when AI outputs are used to make decisions that have direct consequences for an individual's liberty, dignity, and life. Therefore, statistical accuracy or algorithmic efficiency cannot, in itself, be regarded as sufficient evidence that an AI system is appropriate for use in capital sentencing.

From the perspective of Islamic ethics and *maqāṣid al-sharīʿah*, this study finds that the principles of *ʿadl*, *tabayyun*, *amānah*, *masʾūliyyah*, *karāmah al-insān*, *ḥifẓ al-nafs*, and *dafʿ al-ḍarar* provide a strong normative basis for limiting the use of AI in high-risk judicial decisions. The principle of *tabayyun* requires AI data and outputs to be verified before they are used as a basis for decision-making, while *ʿadl* requires that systems not produce discrimination or injustice. The principles of *amānah* and *masʾūliyyah* further establish that responsibility for a decision

cannot be transferred to an algorithm. At the same time, *karāmah al-insān* prevents human beings from being reduced merely to scores or risk profiles, while *ḥifẓ al-nafs* and *dafʿ al-ḍarar* require the highest level of precaution where a decision may result in the loss of human life. The study also finds that AI use should be differentiated according to the extent of its influence on judicial decision-making. The use of AI for administrative functions, legal research, document management, or analytical support entails a different level of risk from its use to predict criminal risk, recommend sentences, or autonomously determine punishment. Therefore, not all uses of AI in judicial processes should be rejected. Rather, the level of human control should be increased in proportion to the level of risk and the irreversibility of the decision.

Based on these findings, this study proposes the principle of presumption against AI-driven capital sentencing, namely the normative presumption that AI should not be granted the authority to autonomously determine the death penalty. This principle does not reject the use of AI altogether; rather, it establishes that substantive decisions concerning capital punishment must remain subject to human deliberation and responsibility. In this regard, the model most consistent with Islamic ethics is AI-assisted but human-determined sentencing, whereby AI functions as a support tool for judges while the final decision remains with human beings who possess judicial authority and responsibility. As a conceptual contribution, this study develops the Islamic AI Death Penalty Decision Framework (IADPDF), incorporating six layers of control: *purpose*, *data integrity*, *algorithmic fairness*, *explainability and transparency*, *meaningful human oversight*, and *maqāṣid assessment*. The framework is not an algorithm for determining who should be sentenced to death. Rather, it constitutes an ethical and governance framework for determining whether, how, and to what extent AI may be used in processes related to the death penalty. IADPDF therefore positions Islamic principles as a normative filter for the use of technology while incorporating contemporary AI governance mechanisms such as auditing, transparency, auditability, and human oversight.

In conclusion, this study emphasises that the more irreversible a decision is, the greater the level of human control, evidentiary verification, and ethical justification required. In the context of the death penalty, this principle reaches its highest threshold because an erroneous decision may result in an irreversible loss of human life. Accordingly, AI should be positioned as a tool that assists human beings in understanding and managing information, rather than as a substitute for judicial judgment and human moral responsibility. This approach is not only more consistent with the principles of Islamic ethics and *maqāṣid al-sharīʿah*, but also aligns with contemporary developments in AI governance that emphasise accountability, transparency, and human oversight for high-risk decisions. Overall, the findings of this study demonstrate that technological advancement must not override the principles of justice and the value of human life; rather, technology should remain subject to principles that ensure judicial decisions remain just, entrusted, accountable, and respectful of human dignity.

References

AbuMusab, S., Sirbu, R., & Floridi, L. (2026). Execution by lethal autonomous mechanism: How anti-capital punishment arguments might lead to automated executions—and why such automation must be prevented. *The Journal of Ethics, 30*, 71–96. doi:10.1007/s10892-025-09541-9

Berk, R. A. (2021). Artificial intelligence, predictive policing, and risk assessment for law enforcement. *Annual Review of Criminology, 4*, 209–237. doi:10.1146/annurev-criminol-051520-012342

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

Castro-Toledo, F. J., Miró-Llinares, F., & Aguerri, J. C. (2023). Data-driven criminal justice in the age of algorithms: Epistemic challenges and practical implications. *Criminal Law Forum, 34*, 295–316. doi:10.1007/s10609-023-09454-y

El.Fiqy, A. M. (2023). Deployment of artificial intelligence algorithms in the criminal justice system: Opportunities and challenges. *Journal of Law and Emerging Technologies, 3*(2). doi:10.54873/jolets.v3i2.158

Elmahjub, E. (2023). Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI. *Philosophy & Technology, 36*, Article 73. doi:10.1007/s13347-023-00668-x

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)*. European Union.

Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. *Journal of Criminal Justice, 81*, Article 101902. doi:10.1016/j.jcrimjus.2022.101902

Fernando, Z. J., & Anditya, A. W. (2024). AI on the bench: The future of judicial systems in the age of artificial intelligence. *Jurnal Hukum dan Peradilan, 13*(3), 523–550. doi:10.25216/jhp.13.3.2024.523-550

Giovanola, B., & Tiribelli, S. (2022). Beyond bias and discrimination: Redefining the AI ethics debate. *AI & Society, 37*, 1–15.

Greenstein, S. (2022). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law, 30*, 291–323. doi:10.1007/s10506-021-09294-4

Hakim, M. K. I. M., & Shamsuddin, M. M. J. (2024). Maqasid Al-Shariah in the age of AI: A critical examination of ChatGPT usage among International Islamic University Malaysia students. *Online Journal of Research in Islamic Studies, 11*(2), 1–18. doi:10.22452/ris.vol11no2.1

Kigerl, A., Hamilton, Z., Kowalski, M., & Mei, X. (2022). The great methods bake-off: Comparing performance of machine learning algorithms. *Journal of Criminal Justice, 81*, Article 101946. doi:10.1016/j.jcrimjus.2022.101946

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), 135–148. doi:10.33093/ajlp.2022.9

Kannike, U. M. M., & Fahm, A. O. (2025). Exploring the ethical governance of artificial intelligence from an Islamic ethical perspective. *Jurnal Fiqh, 22*(1), 134–161. doi:10.22452/fiqh.vol22no1.5

Medvedeva, M., Wieling, M., & Vols, M. (2023). Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law, 31*, 195–212. doi:10.1007/s10506-021-09306-3

Miron, M., Tolan, S., Gómez, E., & Castillo, C. (2021). Evaluating causes of algorithmic bias in juvenile criminal recidivism. *Artificial Intelligence and Law, 29*, 111–147. doi:10.1007/s10506-020-09268-y

Mustapha, A., Senik@Kamaruddin, Z., & Che Haron, S. (2025). An overview of artificial intelligence (AI) issues from the perspective of Islamic jurisprudence. *Online Journal of Research in Islamic Studies, 12*(1). doi:10.22452/ris.vol12no1.6

Raquib, A., Channa, B., Zubair, T., & Qadir, J. (2022). Islamic virtue-based ethics for artificial intelligence. *Discover Artificial Intelligence, 2*, Article 11. doi:10.1007/s44163-022-00028-2

Ryberg, J. (2023). Sentencing disparity and artificial intelligence. *The Journal of Value Inquiry, 57*, 447–462.

Ryberg, J. (2024a). Criminal justice and artificial intelligence: How should we assess the performance of sentencing algorithms? *Philosophy & Technology, 37*, Article 9. doi:10.1007/s13347-024-00694-3

Ryberg, J. (2024b). Punishment and artificial intelligence. In J. Ryberg (Ed.), *The Oxford handbook of the philosophy of punishment* (pp. 690–704). Oxford University Press. doi:10.1093/oxfordhb/9780197750506.013.43

Ryberg, J. (2025a). Artificial intelligence at sentencing: When do algorithms perform well enough to replace humans? *AI and Ethics, 5*, 1009–1018. doi:10.1007/s43681-024-00442-5

Ryberg, J. (2025b). Criminal sentencing and artificial intelligence: What is the input problem? *Criminal Law and Philosophy, 19*, 203–220. doi:10.1007/s11572-024-09739-2

Ryberg, J. (2025c). Artificial intelligence and criminal justice: How to use algorithmic sentencing support in real life (and ethically non-ideal) penal systems? *AI and Ethics, 5*, 3255–3263. doi:10.1007/s43681-024-00655-8

Ryberg, J., & Petersen, T. S. (2022). Sentencing and the conflict between algorithmic accuracy and transparency. In J. Ryberg & J. V. Roberts (Eds.), *Sentencing and artificial intelligence* (pp. 57–73). Oxford University Press.

Ryberg, J., & Roberts, J. V. (Eds.). (2022). *Sentencing and artificial intelligence*. Oxford University Press.

Sachoulidou, A. (2023). Going beyond the "common suspects": To be presumed innocent in the era of algorithms, big data and artificial intelligence. *Artificial Intelligence and Law*. doi:10.1007/s10506-023-09347-w

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

Tan, P. Y. (2023). The need for artificial intelligence in solving unsolved criminal cases and sentencing in Malaysia. *Asian Journal of Law and Policy, 3*(3), 89–103. doi:10.33093/ajlp.2023.7

Taylor, I. (2023). Justice by algorithm: The limits of AI in criminal sentencing. *Criminal Justice Ethics, 42*(3), 193–213. doi:10.1080/0731129X.2023.2275967

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization.

Vogt, D. C. (2026). AI in criminal sentencing: Mapping the ethical terrain. *AI and Ethics, 6*, Article 225. doi:10.1007/s43681-026-01062-x

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

Al-Ghazālī, A. H. (1993). *Al-Mustaṣfā min ʿilm al-uṣūl*. Dār al-Kutub al-ʿIlmiyyah.

Al-Shāṭibī, A. I. (1997). *Al-Muwāfaqāt fī uṣūl al-sharīʿah*. Dār Ibn ʿAffān.

Al-Quran al-Karim. (n.d.). *Al-Quran al-Karim*.

Al-Nawawī, Y. S. (1991). *Al-Majmūʿ sharḥ al-Muhadhdhab*. Dār al-Fikr.

Ibn ʿĀshūr, M. T. (2006). *Treatise on maqāṣid al-sharīʿah* (M. El-Tahir El-Mesawi, Trans.). International Institute of Islamic Thought.

International Covenant on Civil and Political Rights. (1966). *International Covenant on Civil and Political Rights*. United Nations.

United Nations Human Rights Committee. (2018). *General comment No. 36 (2018) on article 6 of the International Covenant on Civil and Political Rights, on the right to life*. United Nations.

European Commission for the Efficiency of Justice. (2018). *European ethical charter on the use of artificial intelligence in judicial systems and their environment*. Council of Europe.

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD.

World Commission on the Ethics of Scientific Knowledge and Technology. (2019). *Preliminary study on the ethics of artificial intelligence*. UNESCO.