Tunjuk Ajar
Journal of Education and Culture

Vol. 1(2), December 2025, 130-146
DOI: 10.64929/ta.v1i2.38
ISSN 3110-8083

Integrating Islamic Ethical Principles into AI Education and Character Building in Digital Society

Yahya Shaykh Ahmad[1*], Yusuf Abdulrasaq[1], Ali Miqdad Ali[2], Farman Ali[3]

AFFILIATIONS

1. University of Ilorin, Nigeria
2. Kwara State University, Malete, Nigeria
3. Riphah International Univeristy, Islamabad, Pakistan

*Correspondence:
yahya.as@unilorin.edu.ng

Received 2025-12-12
Approved 2025-12-28
Published 2025-12-31

ABSTRACT

The rapid evolution of artificial intelligence (AI) presents unprecedented ethical challenges concerning fairness, transparency, and privacy in digital societies. While the global discourse on AI ethics has proliferated, it remains predominantly shaped by secular Western-centric frameworks, creating a significant gap in pluralistic and culturally grounded ethical perspectives. This conceptual study addresses this gap by systematically exploring how the rich ethical tradition of Islam can provide a robust normative framework for enhancing the moral capabilities of AI. Employing a qualitative hermeneutic methodology, the research analyzes primary Islamic sources (the Qur'an and Sunnah) alongside classical and contemporary scholarship to articulate three cardinal ethical principles: *'adl* (comprehensive justice), transparency rooted in *amanah* (sacred trust), and privacy as *ḥurma* (inviolable right). The findings demonstrate that these principles offer direct, actionable guidance to rectify prevalent AI ethical failures, including algorithmic bias, opaque decision-making, and exploitative data practices. Moreover, this study emphasizes the integration of these principles into AI education and character building, proposing curriculum models that cultivate ethical awareness, digital citizenship, and moral responsibility. By translating timeless Islamic ethics into contemporary technical and educational imperatives, this paper contributes a vital, culturally resonant perspective to the global AI ethics dialogue, advocating for interdisciplinary collaboration between AI developers, policymakers, and Islamic ethicists to foster equitable, transparent, and human-centered AI systems.

KEYWORDS

AI ethics, Islamic ethics, Justice (*'adl*), Transparency, Digital education

INTRODUCTION

The rapid and pervasive advancement of artificial intelligence (AI) has ushered in an era of profound transformation across global sectors, unlocking unprecedented opportunities for innovation and societal progress. Technologies ranging from machine learning algorithms to natural language processing models like GPT-5 have demonstrated remarkable capabilities in data analysis, automation, and complex problem-solving (Brown et al., 2020; Delipetrev et al., 2020). However, this technological surge has been paralleled by escalating ethical dilemmas concerning fairness, transparency, privacy, and the preservation of human dignity (UNESCO, 2023). Instances of algorithmic bias, such as discriminatory image search results or the propagation of anti-Muslim sentiment by language models, underscore the urgent need for robust ethical frameworks to guide AI's development and deployment (TRT World, n.d.; IBM, n.d.). The current discourse on AI ethics, while growing, often emerges from secular, Western-

centric philosophical traditions, leading to a significant gap in pluralistic and culturally-grounded ethical perspectives.

Literature Review

The contemporary discourse on artificial intelligence ethics has coalesced around several dominant, often secular and Western-centric, frameworks that emphasize human-centric values, accountability, and technical robustness. Foundational to this conversation is the work of institutions like the European Union's High-Level Expert Group on AI, which defines AI systems and establishes guidelines for "Trustworthy AI," prioritizing principles such as human agency, transparency, and fairness (Samoili et al., 2020). This institutional approach is mirrored by corporate ethical charters from leading technology firms like Microsoft, IBM, and Google, which publicly commit to developing AI that is fair, reliable, and private (Stryker & Kavlakoglu, 2024). Recent comprehensive analyses of global AI ethics guidelines reveal a predominance of Western perspectives, with limited engagement with religious or culturally-grounded ethical systems (Jobin et al., 2019; Hagendorff, 2020).

Philosophically, the field is underpinned by utilitarianism, deontology, and virtue ethics, applied to dilemmas of automation, bias, and machine agency. Scholars like Nick Bostrom and Stuart Russell have framed the long-term existential risks and alignment problems of advanced AI, while researchers such as Joy Buolamwini and Timnit Gebru have empirically exposed racial and gender biases in facial recognition and natural language processing algorithms, grounding ethical concerns in measurable harm. Systematic mappings of the algorithmic ethics debate further highlight persistent challenges related to bias, opacity, and accountability that remain inadequately addressed by current frameworks (Mittelstadt et al., 2016; Binns, 2018). These secular frameworks, however, often operate within a paradigm that separates ethical reasoning from religious or spiritually-grounded value systems, potentially overlooking rich reservoirs of moral wisdom that have guided human societies for millennia.

Concurrently, a robust and distinct body of scholarship explores the historical and epistemological relationship between Islam and the scientific enterprise. Academic work in this domain establishes that the Islamic tradition is not antagonistic to science but has historically been a profound catalyst for it. As elucidated by Nidhal Guessoum (2010), the Qur'anic revelation itself, beginning with the command "Iqra" (Recite), inaugurates an epistemic worldview that venerates the pursuit of knowledge (*ʿilm*) and demands empirical evidence, thereby fostering a culture of inquiry. Historical analyses, such as those by Yasmeen Mahnaz Faruqi (2006), meticulously document the monumental contributions of Muslim scholars during the Islamic Golden Age. Figures like Al-Khwarizmi, whose work on algebra and algorithmic calculation forms a bedrock of computer science, and Ibn Sina (Avicenna), whose Canon of Medicine structured medical thought for centuries, exemplify this synthesis. This scholarly tradition demonstrates that for Muslim intellectuals, scientific exploration was an act of piety, a means to understand Allah's creation (*āyāt*) and utilize its resources responsibly, as guided by the principle of stewardship (*istikhlāf*) mentioned in the Qur'an (Q45:13).

Within Islamic studies itself, a sophisticated literature exists on normative ethics (*akhlāq*) and jurisprudence (*fiqh*), which deal extensively with concepts directly relevant to modern technology ethics. The principle of *ʿadl* (justice and equity) is a central Qur'anic theme, extensively analysed by scholars like Shaykh Ahmad Yahya and Miqdad Ali (2023) as a

comprehensive social and moral imperative to place everything in its rightful place and ensure equitable treatment. The concept of *ḥurma* (inviolability) underpins the rigorous Islamic protections for privacy, prohibiting spying, trespassing, and the unauthorized disclosure of secrets, as discussed in classical and modern works on Islamic law and human rights (Mawdoodi, 1980; Memon Madani, 1998). Transparency and trust (*amānah*) in transactions and governance are also mandated, as seen in the Qur'anic injunctions on documenting contracts (Q2:282) and the Prophetic traditions condemning deceit.

Contemporary Islamic bioethics has successfully responded to modern medical challenges by setting a precedent for applying classical legal methodologies, such as the *maqāṣid al-sharīʿah*, to new technological contexts. However, this dynamic ethical discourse has only scratched the surface of the challenges of the digital era. Although a number of recent studies have emerged at the intersection of Islamic ethics and Artificial Intelligence, exploring virtue-based approaches (Raquib, 2022), pluralist ethical benchmarks (Elmahjub, 2023), and the ethics of guardianship (Ali, 2025), these contributions remain scattered and have not yet been integrated into a comprehensive framework. The systematic, in-depth, and scientific application of Islamic ethical frameworks to specific AI challenges such as algorithmic autonomy, transparency, data exploitation, and its social impacts remains very limited. This indicates a significant gap, given that the Islamic tradition not only offers historical inspiration, but is also a living, rational, and grounded value system that can provide unique insights into ensuring that technology upholds human dignity and the common good (*al-maṣlaḥah al-ʿāmmah*).

Research Gap

Despite the extensive literature in AI ethics and Islamic studies, a critical interdisciplinary gap remains. Existing AI ethics frameworks largely emerge from Western secular philosophical traditions, often lacking the rich value-based systems of other major worldviews. Instead, contemporary Islamic scholarship on science and technology often focuses on historical contributions or bioethical issues, with limited systematic application to the new ethical dilemmas posed by digital and autonomous systems such as AI. There is a dearth of research that actively translates comprehensive Islamic ethical teachings, such as the principles of justice, transparency, privacy, and governance, into structured and actionable frameworks specifically designed to address the operational and moral challenges of AI development, auditing, and governance. This gap represents a missed opportunity to enrich the global ethical dialogue with perspectives that have historically bridged spiritual morality with empirical scientific pursuits.

Aims and Contribution

This paper aims to address this gap by arguing that Islamic ethical principles offer a vital and coherent framework for enhancing the moral capabilities of artificial intelligence. It seeks to move beyond theoretical assertions and toward constructive applications. Its primary objective is to critically examine how Islamic principles such as justice (*'adl*), transparency, and privacy protection can inform and address common ethical deficiencies in AI systems. To achieve this, the study employs qualitative and conceptual research approaches, analyzing primary Islamic sources (the Qur'an and Sunnah) alongside contemporary ethical challenges

in AI. Its contribution is threefold: First, it provides a focused synthesis of Islamic ethics relevant to the digital age. Second, it proposes concrete applications of these principles to mitigate issues such as algorithmic bias and data exploitation. Finally, it advocates and outlines pathways for interdisciplinary collaboration between AI developers and Islamic ethicists. In doing so, this paper aims to contribute to a more inclusive, pluralistic, and effective global effort to ensure that artificial intelligence serves humanity with justice, clarity, and respect for fundamental rights.

METHODOLOGY

Research Approach

This study employs a qualitative, conceptual research design situated within the interpretive paradigm, utilizing a hermeneutic and applied ethics methodology. The primary objective is the construction of a normative ethical framework through the analytical synthesis of textual sources and principles (Creswell & Poth, 2018). The approach is fundamentally interpretive, as it involves retrieving, explicating, and applying ethical norms from Islamic tradition to the contemporary problem space of artificial intelligence ethics. This methodology is particularly appropriate for interdisciplinary inquiry that bridges religious ethics with technological discourse, as it enables deep engagement with values, meanings, and normative prescriptions (Denzin & Lincoln, 2018). By framing the research as an exercise in applied ethics (Beauchamp & Childress, 2019), the study moves from identifying concrete ethical failures in AI such as bias and privacy violation to proposing principled corrections derived systematically from Islamic teachings.

Data Sources and Data Collection Techniques

This research relies entirely on documentary data, collected through systematic and purposeful sampling of textual sources (Bowen, 2009). Primary Islamic sources constitute the normative foundation and include the Qur'an, consulted through thematic indexing and verified translations, and the prophetic tradition (Sunnah), drawn from major authenticated Hadith collections. Secondary sources provide context, interpretation, and contemporary dialogue; these include classical and modern Islamic exegesis and ethical studies (Guessoum, 2010), peer-reviewed literature on AI ethics, official policy documents from bodies such as the European Union (Samoili et al., 2020) and UNESCO (2023), and technical writings on AI development and capabilities. Data collection was guided by relevance and authority, ensuring the inclusion of key texts defining Islamic ethical discourse and the current state of AI ethics, thus creating a robust corpus for interdisciplinary analysis.

Data Analysis Technique

Data analysis followed a structured two-stage qualitative process integrating thematic analysis and conceptual synthesis (Braun & Clarke, 2006). In the first stage, primary Islamic texts and scholarly commentaries were carefully read and inductively coded to identify, extract, and articulate the core ethical principles of justice (‘adl), transparency, and privacy. This involved tracing the semantic and juridical contours of each principle as established in classical Islamic thought. The second stage involved a deductive conceptual application,

where these extracted principles were systematically linked to documented ethical challenges in AI. Each challenge was examined through the lens of the corresponding Islamic principle. The analysis moved beyond comparison to synthesis, exploring how each principle reconfigures ethical issues and generates specific imperatives for the design, governance, and audit of AI. This method ensured that the proposed framework is grounded in traditional normative integrity and responsive to contemporary technical realities.

Research Ethics

As a conceptual and document-based study, this research adheres rigorously to academic integrity and ethical scholarly practice. All sources are accurately cited to honor intellectual contributions and avoid plagiarism, following established academic conventions. Special care is taken in the interpretation of religious texts, ensuring that representations are anchored in recognized scholarly traditions to prevent mischaracterization, a key ethical consideration in cross-cultural and interdisciplinary research (Tracy, 2010). The analysis maintains a stance of academic objectivity, striving to present Islamic ethics faithfully while engaging constructively and respectfully with secular ethical discourse. The research is conceived as a contribution to pluralistic dialogue rather than a polemical argument. Moreover, by centering its proposals on the enhancement of human dignity, fairness, and privacy, the study aligns its own ethical posture with the values it advocates, consciously considering the potential social impact of its recommendations toward more just and humane technological development.

FINDINGS AND DISCUSSION

A rigorous hermeneutic analysis of the foundational texts of Islam, specifically the Qur'an, the Prophetic traditions (Sunnah), and the works of classical Islamic scholars, reveals a coherent and deeply rooted ethical architecture. This architecture consists of three cardinal principles with profound and immediate implications for the governance of artificial intelligence. These principles are *'adl* (comprehensive justice), transparency rooted in *amanah* (sacred trust), and privacy as *ḥurma* (an inviolable sanctity). They are not merely isolated virtues but form an integrated, mutually reinforcing framework. This framework is itself grounded in Islam's overarching worldview of human stewardship (*istikhlāf*), the concept that humanity is a trustee of creation, and the paramount objective of pursuing universal societal benefit (*maṣlaḥah*). When applied to the domain of AI ethics, this framework serves two critical functions: first, as a diagnostic tool for identifying systemic ethical failures in existing technological systems, and second, as a source of constructive, normative guidance for designing, deploying, and regulating technology in a manner that fundamentally respects and upholds human dignity.

To systematically illustrate how these abstract principles translate into actionable governance measures, Figure 1 presents a conceptual framework that maps the relationship between Islamic ethical foundations and their practical implications for AI governance. This framework visually delineates the translational pathway through which *'adl*, *amanah*, and *ḥurma* are operationalized into tangible ethical outcomes, thereby bridging theological-ethical doctrine with applied technological governance.

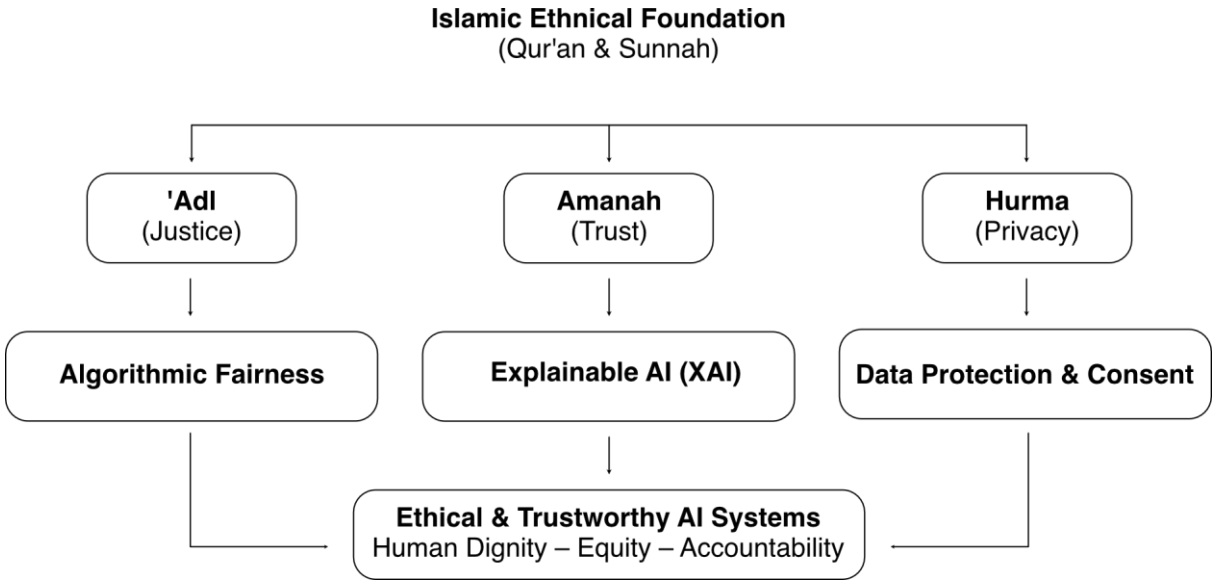

Figure 1. Islamic Ethical Framework for AI Governance

As demonstrated in Figure 1, the principle of *'adl* functions as a comprehensive imperative against algorithmic and structural injustice. Derived from verses such as *"Indeed, Allah commands justice and good conduct"* (Q16:90), this principle transcends retributive fairness to encompass proactive, restorative, and procedural equity in AI systems.

'Adl as a Comprehensive Imperative Against Algorithmic and Structural Injustice

The principle of *'adl*, derived from verses such as "Indeed, Allah commands justice and good conduct" (Q16:90) and "We sent Our Messengers with clear proofs and sent down with them the Scripture and the balance that the people may maintain justice" (Q57:25), transcends retributive fairness to encompass proactive, restorative, and procedural equity. Classical exegesis, as noted by scholars like Yahya and Ali (2023), defines *'adl* as placing everything in its rightful place, creating a normative standard that challenges both overt discrimination and embedded systemic bias. Applied to AI, this principle imposes ethical obligations across the technology's entire lifecycle, demanding a holistic approach to fairness that integrates distributive, restorative, and procedural dimensions.

At the foundational data stage, *'adl* necessitates rigorous auditing of training datasets for historical prejudices, particularly against underrepresented or marginalized groups. This obligation stems from the Qur'anic imperative against acting on conjecture without evidence (implied in Q53:28), which in a technological context translates to a mandate for evidence-based, representative, and balanced (*tawāzun*) data. The use of historically biased data to train AI systems without corrective measures perpetuates past injustices, violating the Islamic legal maxim *"lā ḍarar wa lā ḍirār"* (no harm shall be inflicted or reciprocated). Consequently, the principle of *'adl* requires proactive measures to identify and mitigate harmful biases in data, ensuring it does not cause new harm or reinforce existing societal inequities.

During the algorithm design and development phase, *'adl* moves beyond technical optimization for accuracy to demand the implementation of fairness constraints that achieve

equitable outcomes. An AI model might be statistically "accurate" yet socially discriminatory, a scenario fundamentally at odds with the Prophetic injunction to give rights to those entitled to them (*ta'hīl al-ḥaqq ilā ahlih*). This involves configuring algorithms to ensure *qisṭ* (equitable distribution) and prevent outcomes that systematically disadvantage any group. The Islamic conception of justice is active and substantive; it is not satisfied by procedural neutrality if the result is inequitable. Therefore, AI systems in domains like hiring, lending, or criminal justice must be designed with embedded fairness metrics that align with the restorative spirit of *'adl*, aiming to rectify rather than replicate structural imbalances.

In the deployment and post-deployment stages, *'adl* mandates continuous monitoring (*murāqabah*) for disparate impacts and the establishment of accessible, effective redress mechanisms. The Prophetic command to "Help your brother, whether he is an oppressor or is oppressed" is interpreted as a communal duty to correct wrongs whether the oppressor is a human or a systemic artifact like a biased algorithm. This translates to an operational requirement for ongoing impact assessments, transparent reporting on AI performance across demographic groups, and clear channels for grievance and repair (*iṣlāḥ*). This continuous ethical stewardship reflects the Islamic principle of *naṣīḥah* (sincere advice for the common good) and the collective responsibility (*farḍ kifāyah*) to ensure societal well-being. Thus, documented AI failures, such as racially skewed risk assessment algorithms or gender-biased hiring tools, are framed not as mere technical glitches but as fundamental violations of a divine command to establish comprehensive justice in human affairs.

This framework thus addresses documented AI failures, such as biased hiring tools or racially skewed risk assessment algorithms, by framing them not as technical glitches but as fundamental violations of a divine command to establish justice. This perspective aligns with but extends beyond secular fairness discourses in machine learning, which often focus narrowly on statistical parity rather than substantive, restorative justice (Binns, 2018; Floridi & Cowls, 2019).

Building on this understanding of 'adl, the analysis further reveals that the three principles *('adl, amanah*, and *ḥurma*) together form a coherent ethical architecture. They are not merely abstract virtues but actionable anchors that can directly inform AI governance. To illustrate their practical relevance and interconnectedness, Table 1 synthesizes each principle's conceptual foundation, its corresponding ethical imperative in the context of AI, and tangible applications in technical design and regulatory frameworks. This structured overview provides a foundational lens for the detailed discussion that follows.

Table 1: Islamic Ethical Principles and Their Applications in Artificial Intelligence Governance

Islamic Principle	Conceptual Meaning	Ethical Imperative in AI	Technical & Governance Applications
'*Adl* (Justice)	Holistic, procedural, and restorative justice; placing everything in its rightful place.	Prevent bias, ensure equitable distribution of benefits, provide redress mechanisms.	Dataset auditing, fairness constraints in algorithms, demographic impact monitoring, grievance channels.
Amanah (Trust & Trusteeship)	Transparency, accountability, and moral	Ensure explainability, maintain audit trails, uphold	Explainable AI (XAI) techniques, algorithmic

	responsibility as a sacred trust before the Divine.	accountability for AI outcomes.	impact assessments, transparent decision logs.
Ḥurma (*Inviolability*)	Privacy as an inherent, dignity-based right; protection of personal and spiritual sanctity.	Obtain informed consent, limit data collection to necessity, ensure data security and confidentiality.	Explicit consent mechanisms, purpose limitation, data anonymization, encryption, anti-surveillance policies.

Having outlined the conceptual and applied dimensions of these principles, the following sections delve deeper into each, examining their theological foundations, ethical implications, and specific relevance to contemporary AI challenges.

Transparency as a Function of *Amanah* (Trust) and Accountability

Transparency in the Islamic ethical system is fundamentally and inextricably linked to the concepts of *amānah* (sacred trust) and ultimate accountability before the Divine. The Qur'an's detailed prescriptions for documenting transactions with witnesses (Q2:282-283) and the command to "bring forth your proof" (Q2:111) establish a religious-legal norm of verifiability, clarity, and evidence-based interaction, all intended to prevent deceit (*ghish*) and oppression (*ẓulm*). This principle directly and powerfully challenges the ethical acceptability of opaque "black box" AI systems, especially in high-stakes domains like healthcare, justice, and finance. It frames explainability not as an optional technical feature or a means to user acceptance, but as a core moral obligation stemming from the developers' and deployers' *amānah* as stewards of powerful, decision-influencing technology.

The linkage of transparency to *amānah* (trust) and divine accountability adds a profound layer of moral responsibility for developers, suggesting that explainability is integral to their moral integrity, not merely a tool for user acceptance. This contrasts with debates in Western AI ethics about whether a "right to explanation" exists within current legal frameworks like the GDPR (Wachter et al., 2017; Kaminski, 2019), positioning transparency instead as a non-negotiable ethical imperative rooted in divine accountability.

The theological weight of *amānah* imbues the development and governance of AI with a profound sense of moral responsibility. In Islam, knowledge and the power it confers are trusts from God for which humanity will be held accountable. The Hadith regarding the neglect of a camel, where the Prophet Muhammad (peace be upon him) demanded an account from its owner (Ṣaḥīḥ Muslim 4405), provides a potent analogy for the accountability of AI operators. Just as the camel's owner was responsible for its well-being, those who create and deploy AI systems are accountable for their societal impacts, intended or otherwise. This perspective mandates that institutions maintain auditable decision logs, conduct and publish transparent Algorithmic Impact Assessments, and establish clear lines of responsibility. It strengthens secular calls for transparency by rooting them in a framework of sacred responsibility, where obscurity and non-auditability constitute a potential breach of divine trust and a failure in one's duty to the community (*ḥuqūq al-'ibād*).

This principle has direct implications for technical development and educational practice. It provides a robust ethical mandate for prioritizing Explainable AI (XAI) techniques, ensuring that AI decision-making processes can be understood and interrogated by human stakeholders. Furthermore, it underscores the need to integrate these values into the education of future computer scientists and engineers. Curricula in technology and AI should

incorporate ethical modules grounded in diverse traditions, including Islamic perspectives on *amānah* and accountability. This educational integration aims to cultivate a generation of technologists who view transparency not as a regulatory burden, but as an integral component of professional integrity and moral responsibility in the digital age, thereby fulfilling the trust placed in them as knowledge-holders and innovators.

Ḥurma (privacy) as an Inviolable Right and Foundation for Digital Dignity

The analysis establishes that privacy is a protected and inviolable right (*ḥurma*) in Islamic law, intrinsically linked to the preservation of human dignity (*ʻird*), personal autonomy, and spiritual sanctuary. Strong Qur'anic prohibitions against entering homes without permission (Q24:27-29) and the unequivocal condemnation of spying (*tajassus*), alongside Prophetic declarations on the inviolability of a Muslim's person, life, and honour, create a robust ethical and legal boundary. This boundary extends conceptually and normatively to personal data in the digital realm, which is viewed as an extension of the self (*dhāt*) and one's private domain. Therefore, the Islamic framework offers a stringent, dignity-based critique of the pervasive data surveillance and extractivist practices that fuel much of contemporary AI development.

From an Islamic perspective, personal data is not a freely exploitable economic resource but a trust (*amānah*) that carries obligations. This viewpoint leads to several operational imperatives for ethical AI governance. First, the collection and processing of personal data require explicit, informed, and uncoerced consent (*riḍā*), reflecting the principle of mutual agreement in all transactions. Second, data practices must be guided by *ḍarūrah* (necessity) and purpose limitation, meaning data should only be collected for a specific, legitimate need and not retained or used beyond that declared purpose. Third, entities that collect data bear the *amānah* of safeguarding it with confidentiality and security, preventing unauthorized access or disclosure. These principles directly challenge prevalent practices such as non-consensual web scraping, opaque data profiling, deceptive consent mechanisms, and biometric mass surveillance.

Furthermore, this conception of privacy as *ḥurma* provides a potent counter-narrative to discourses that frame privacy as a commodity to be traded for convenience or security. It asserts that violations of digital privacy are not merely contractual or legal issues but moral transgressions that harm the human spirit and societal trust. This is particularly salient in addressing AI-facilitated harms, such as the propagation of anti-Muslim hatred through biased language models (TRT World, n.d.), which constitutes a dual violation of both *ʻadl* (justice) and *ḥurma* (the inviolable dignity of the community). For educational and policy contexts, this principle underscores the importance of digital literacy programs that teach individuals about their digital rights from a holistic, value-based perspective, empowering them to navigate online spaces with an awareness of their inherent dignity and the ethical responsibilities of technology stakeholders.

This perspective challenges practices like non-consensual web scraping, biometric surveillance, and opaque data profiling. It strengthens global data protection norms but from a foundation that views privacy violations as moral transgressions harming the human spirit, thereby providing a potent, dignity-based argument against data-extractivist economic models. This human rights-based approach to data protection resonates with emerging frameworks for designing AI with human rights considerations at its core (Aizenberg & van den Hoven, 2020).

Integrating Islamic AI Ethics into Education and Character Building

Beyond their relevance for technological governance and policy, the Islamic ethical principles articulated in this study carry profound implications for education, particularly in the context of character building and ethical formation in digital societies. Education has historically functioned as the primary medium through which moral values, cultural norms, and social responsibilities are transmitted across generations. In an era where artificial intelligence increasingly shapes cognition, decision-making, and social interaction, the integration of Islamic AI ethics into educational frameworks becomes a strategic and ethical necessity. The principles of *ʻadl* (justice), *amanah* (trust and accountability), and *ḥurma* (inviolability of dignity and privacy) provide a culturally grounded foundation for cultivating ethical awareness and moral responsibility among learners navigating AI-mediated environments.

Figure 2 visualizes the educational translation of Islamic AI ethics proposed in this study. The framework positions ʻ*adl*, *amanah*, and *ḥurma* as foundational ethical values that inform curriculum design and pedagogical strategies, ultimately shaping character formation and ethical consciousness in digital society. By situating education as a mediating domain between ethical principles and societal outcomes, the model reinforces the role of education as a transformative space for cultivating responsible engagement with artificial intelligence.

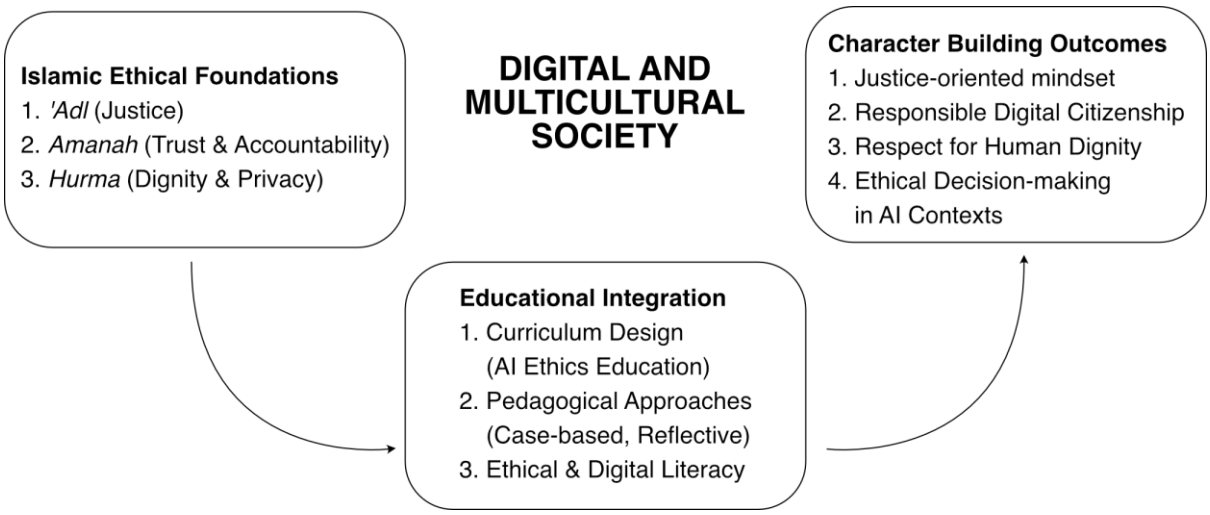

Figure 2. Conceptual Framework for Integrating Islamic AI Ethics into Education and Character Building

The principle of ʻ*adl* offers a powerful pedagogical anchor for addressing issues of fairness, equity, and social responsibility in AI education. As highlighted in both Islamic ethics and contemporary AI ethics literature (Binns, 2018; Floridi & Cowls, 2019), algorithmic bias and unequal technological outcomes are not merely technical flaws but reflections of deeper moral and social failures. Integrating ʻ*adl* into educational curricula encourages students, particularly those studying computer science, data science, and information systems, to critically examine how datasets, design choices, and institutional contexts can reproduce injustice. This approach aligns with the Islamic educational tradition that frames knowledge (ʻ*ilm*) as inseparable from ethical responsibility (Guessoum, 2010). Pedagogically, this can be operationalized through case-based learning, critical reflection on real-world AI failures, and

interdisciplinary dialogue that connects technical competencies with moral reasoning. In doing so, education functions not only to produce skilled technologists but also morally conscious individuals committed to equitable outcomes and restorative justice in digital systems.

Amanah, understood as sacred trust and accountability, further reinforces the ethical role of education in shaping professional integrity and civic responsibility. In Islamic thought, knowledge is a trust bestowed upon humanity, and its misuse constitutes a moral breach with social consequences. When applied to AI education, *amanah* reframes transparency, explainability, and accountability as ethical obligations rather than optional technical features. This perspective resonates with contemporary debates on Explainable AI and algorithmic accountability (Wachter et al., 2017; Kaminski, 2019), while grounding them in a value system that emphasizes moral stewardship. Educational programs that incorporate *amanah* can foster a culture of responsibility among future AI developers, policymakers, and educators by emphasizing documentation, ethical auditing, and reflective practice. Embedding these values into curricula, particularly in higher education and professional training, helps cultivate graduates who perceive transparency not as regulatory compliance but as an expression of ethical character and trustworthiness.

The principle of *ḥurma*, which establishes privacy and dignity as inviolable rights, has equally significant implications for education and character formation in digital contexts. As AI-driven surveillance, data extraction, and profiling practices become normalized, learners must be equipped with ethical frameworks that challenge the commodification of personal data. Islamic ethics conceptualizes privacy as an extension of human dignity, aligning with human rights–based approaches to data protection (Aizenberg & van den Hoven, 2020), yet offering a deeper moral grounding rooted in spiritual and social sanctity. Educational integration of *ḥurma* encourages critical digital literacy, enabling students to understand consent, data ownership, and the ethical limits of technological intervention. This is particularly relevant in educational institutions themselves, which increasingly rely on learning analytics, biometric systems, and AI-based monitoring tools. Teaching *ḥurma* as a core ethical principle fosters respect for personal boundaries and reinforces the moral dimensions of digital citizenship.

Collectively, the integration of *‘adl*, *amanah*, and *ḥurma* into education supports a holistic model of character building that aligns cognitive development with ethical formation. This approach resonates with Islamic educational philosophy, which has historically emphasized the cultivation of *adab*, moral conduct and intellectual humility, alongside technical knowledge. In contemporary settings, this model can inform curriculum design across disciplines, including religious education, social sciences, and STEM fields. It also supports interdisciplinary pedagogies that bridge technology, ethics, and cultural studies, reinforcing the journal’s focus on education within multicultural and digital societies. Moreover, embedding Islamic AI ethics into education contributes to the development of value-sensitive curricula that reflect local cultural contexts while engaging global technological challenges.

In the broader educational landscape, this integration also has implications for teacher education and institutional policy. Educators play a critical role in mediating ethical discourse and modeling responsible engagement with technology. Training programs that familiarize teachers with Islamic ethical perspectives on AI can enhance their capacity to guide students through complex moral dilemmas associated with automation, data use, and digital power

structures. At the policy level, educational institutions, particularly in Muslim-majority contexts, can draw on these principles to develop ethical guidelines for AI adoption that are culturally resonant and socially legitimate.

Ultimately, positioning Islamic AI ethics within education transforms ethical reflection from a peripheral concern into a core educational objective. It reinforces the role of education as a site for moral cultivation and cultural continuity in the face of rapid technological change. By integrating these principles into curricula, pedagogy, and institutional practice, education can serve as a transformative force that ensures artificial intelligence evolves in harmony with human dignity, social justice, and ethical responsibility.

Discussion

The findings demonstrate that Islamic ethics converges with and critically enriches the global AI ethics discourse. While principles like fairness, transparency, and privacy are recognized in secular frameworks (e.g., EU's HLEG guidelines, Samoili et al., 2020), and are part of unified ethical frameworks proposed for AI in society (Floridi & Cowls, 2019), the Islamic tradition provides them with a coherent metaphysical foundation that may enhance their normative force and cultural resonance. For instance, the Islamic concept of *'adl* reframes algorithmic fairness from a technical optimization problem into a religious and moral imperative for restorative justice, potentially motivating more proactive efforts to rectify embedded biases than compliance-driven approaches might inspire.

The linkage of transparency to *amanah* (trust) and divine accountability adds a profound layer of moral responsibility for developers, suggesting that explainability is integral to their moral integrity, not merely a tool for user acceptance. Similarly, the concept of *ḥurma* fortifies privacy arguments with a dignity-based absolutism that can powerfully counter narratives framing privacy as a trade-off for convenience or security. This is particularly relevant in addressing AI-facilitated harms like the propagation of anti-Muslim hatred documented in language models (TRT World, n.d.), which violates both *'adl* and *ḥurma*.

Furthermore, the integrated nature of the Islamic framework connects technology ethics to higher objectives (*maqāṣid al-sharī'ah*), such as preserving intellect, lineage, and property. This encourages a holistic evaluation of AI's impact on community cohesion and moral cognition, a dimension often overlooked in principles-based checklists. Recent studies have begun to explore the application of *maqāṣid al-sharī'ah* specifically to AI governance (Zarqa, 2020), while multi-religious perspectives on AI highlight the importance of engaging diverse religious traditions in technology ethics (Ahmed, 2025). The framework thus offers a vital, culturally-grounded resource for policymakers in Muslim-majority nations to develop resonant AI governance strategies and for the global community to achieve a more pluralistic and robust ethical consensus.

Research Implications

This research carries significant multi-stakeholder implications. For AI developers and corporations, it underscores the necessity of inclusive ethical review processes that incorporate diverse cultural and religious perspectives to identify context-specific harms. This aligns with calls for more critical evaluation of existing AI ethics guidelines and their

implementation (Hagendorff, 2020). For policymakers, especially within the OIC member states, it provides a theological and ethical basis for crafting national AI strategies and regulations that align with local values, potentially fostering greater public trust and responsible innovation. Studies on integrating Islamic ethics with modern technology highlight both challenges and opportunities in such cross-disciplinary work (Rahman & Al Farisi, 2021). For the global AI ethics community, it demonstrates the necessity of intellectual pluralism, showing how engagement with religious ethical systems can reinforce universal values through diverse justifications and uncover blind spots in dominant paradigms. For academia, it catalyzes the nascent field of "Islamic Digital Ethics," calling for interdisciplinary work to translate these principles into practical tools, audit frameworks, and cross-disciplinary curricula that bridge computer science and humanities.

Research Limitations

This conceptual study has inherent limitations that delineate its scope and point to future research directions. First, as a normative textual analysis, it proposes an ethical framework but does not provide empirical validation of its implementation efficacy in real-world development environments. Collaborative action research with tech teams is needed. Second, the interpretation, while grounded in mainstream classical scholarship, represents one methodological approach within a pluralistic tradition; other Islamic schools of thought may yield different emphases. Third, the framework primarily addresses design and governance ethics. The formidable structural challenges posed by the global political economy of AI require separate, critical socio-political analysis. Finally, the practical application of this framework relies on ongoing interdisciplinary collaboration between ethicists, Islamic scholars, and computer scientists. These challenges are not trivial and must be strategically addressed for the principles to move from theory to practice.

CONCLUSION

This study has systematically demonstrated that the rich ethical tradition of Islam offers a profound and coherent framework for addressing the pressing moral challenges posed by artificial intelligence. By examining the principles of *'adl* (comprehensive justice), transparency grounded in *amanah* (trusteeship), and privacy as *ḥurma* (inviolability), the research has translated foundational Islamic concepts into actionable normative guidance for AI development and governance.

The findings affirm that these principles are not antithetical to technological progress but provide essential safeguards to ensure such progress aligns with the preservation of human dignity, equitable benefit, and societal welfare. The proposed framework directly addresses critical shortcomings in contemporary AI systems, from algorithmic bias and opaque decision-making to exploitative data practices, by offering a value-based justification rooted in a worldview shared by over a billion people. Ultimately, this paper argues that integrating such culturally-grounded ethical perspectives is not merely an academic exercise but a practical imperative for developing AI that is truly trustworthy, equitable, and respectful of the diverse fabric of global society. Future work must focus on the interdisciplinary collaboration necessary to operationalize these principles into technical standards, policy

instruments, and educational curricula, thereby ensuring that the pursuit of artificial intelligence remains firmly anchored in the service of humanity.

AI Usage Declaration

During the preparation of this work, the authors used Dimension AI and DeepSeek AI for literature exploration, conceptual structuring, and preliminary drafting assistance. These tools were employed to enhance research efficiency and support analytical framing. After using these services, the authors reviewed and edited the content critically and take full responsibility for the final manuscript's content, integrity, and originality.

Author Contribution Statement

MZM: Conceptualization, methodology, original draft preparation, project administration, and final manuscript review. SHA: Literature review, data curation, formal analysis, and manuscript revision. SAMB: Theoretical framework development, validation, visualization, and editing. All authors have read and approved the final version of the manuscript.

REFERENCES

Ahmed, S. (2025). Exploring multi-religious perspective of artificial intelligence. *Theology and Science*, 23(1), 104–128. https://doi.org/10.1080/14746700.2024.2436783

Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, 7(2). https://doi.org/10.1177/2053951720949566

Al-Bayhaqī, A. B. A. (2003). *Shuʿab al-Īmān* (Vol. 2). (A. A. A. Ḥāmid, Ed.). Maktabat al-Rushd.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular biology of the cell* (4th ed.). Garland Science.

Ali, F. (2025). Islamic ethics and AI: An evaluation of existing approaches to AI using trusteeship ethics. *Philosophy and Technology*, 38(3), Article 922. https://doi.org/10.1007/s13347-025-00922-4

Arena, C. (2022, June 14). *7 Disadvantages of Artificial Intelligence Everyone should know about*. Liberties.eu. https://www.liberties.eu/en/stories/disadvantages-of-artificial-intelligence/44289

AvicenaTech Corp. (2024). *The history of artificial intelligence*. https://avicena.tech/resources/the-history-of-artificial-intelligence/

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (FAT)* , 1–10. https://doi.org/10.1145/3287560.3287580

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *arXiv*. https://arxiv.org/abs/2005.14165

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Delipetrev, B., Tsinaraki, C., & Kostić, U. (2020). *AI watch: Historical evolution of artificial intelligence: Analysis of the three main paradigm shifts in AI*. Publications Office of the European Union. https://doi.org/10.2760/70660

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2018). *The SAGE handbook of qualitative research* (5th ed.). SAGE Publications.

Dickmanns, E. (2025, August 6). *Pioneering work in autonomous driving*. Wikipedia. https://en.wikipedia.org/wiki/Ernst_Dickmanns

Duggal, N. (2025, January 16). *Advantages and disadvantages of artificial intelligence*. Simplilearn. https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article

Elmahjub, E. (2023). Artificial intelligence (AI) in Islamic ethics: Towards pluralist ethical benchmarking for AI. *Philosophy and Technology*, 36(4), Article 6. https://doi.org/10.1007/s13347-023-00668-x

Faruqi, Y. M. (2006). Contributions of Islamic scholars to the scientific enterprise. *International Education Journal, 7*(4), 391–399.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Ghaly, M. (2024). What makes work “good” in the age of artificial intelligence (AI)? Islamic perspectives on AI-mediated work ethics. *Journal of Ethics*, 28(3), 429–453. https://doi.org/10.1007/s10892-023-09456-3

Guessoum, N. (2010). Science, religion, and the quest for knowledge and truth: An Islamic perspective. *Cultural Studies of Science Education, 5*(1), 55–69. https://doi.org/10.1007/s11422-009-9207-4

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Ibn Mājah. (n.d.). *Sunan Ibn Mājah* (Vol. 1). (M. F. A. al-Bāqī, Ed.). Dār al-Fikr.

Imperial War Museums. (2020, March 5). *How Alan Turing cracked the Enigma code*. https://www.iwm.org.uk/history/how-alan-turing-cracked-the-enigma-code

Investopedia Team. (2025, August 9). *What is artificial intelligence (AI)?* Investopedia. https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kaminski, M. E. (2019). The right to explanation, explained. *Berkeley Technology Law Journal*, 34(1), 189–218.

Kearey, P., Klepeis, K. A., & Vine, F. J. (2009). *Global tectonics* (3rd ed.). Wiley-Blackwell.

Mawdoodi, A. A. (1980). *Human rights in Islam* (2nd ed.). Islamic Foundation.

Memon Madani, M. I. (1998). *The Islamic commandments of Hijab*. Kazi Publications.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

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

Muslim ibn al-Hajjaj. (1976). *Ṣaḥīḥ Muslim* (A. H. Siddiqui, Trans.). Kazi Publications.

Oloyede, I. O. (2024). *Artificial intelligence and the future of the humanities*. Unilorin Press.

Rahman, F., & Al Farisi, M. Z. (2021). Integrating Islamic ethics into modern technology: Challenges and opportunities. *Journal of Religion and Technology*, 9(2), 112–130. https://doi.org/10.1080/24751839.2021.1895210

Raquib, A. (2022). Islamic virtue-based ethics for artificial intelligence. *Discover Artificial Intelligence*, 2(1), Article 28. https://doi.org/10.1007/s44163-022-00028-2

Samoili, S., Lopez Cobo, M., Gomez Gutierrez, E., De Prato, G., Martinez-Plumed, F., & Delipetrev, B. (2020). *AI watch: Defining artificial intelligence* (EUR 30117 EN). Publications Office of the European Union. https://doi.org/10.2760/382730

Smith, C. (2006). *The history of artificial intelligence*. University of Washington.

SRI International. (2025). *Shakey the robot*. Retrieved August 9, 2025, from https://www.sri.com/hoi/shakey-the-robot/

Stryker, C., & Kavlakoglu, E. (2024, August 9). *What is artificial intelligence (AI)?* IBM Think. https://www.ibm.com/think/topics/artificial-intelligence

Sulaymān ibn al-Ash'ath al-Sijistānī, A. D. (2007). *Sunan Abī Dāwūd* (N. al-Khattab, Trans.). Darussalam.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837–851. https://doi.org/10.1177/1077800410383121

TRT World. (n.d.). *Rigged AI: How algorithms fuel anti-Muslim hatred in Europe and beyond*. Retrieved August 9, 2025, from https://www.trtworld.com/magazine/rigged-ai-how-algorithms-fuel-anti-muslim-hatred-in-europe-and-beyond-59241

UNESCO. (2023). *Artificial intelligence: Examples of ethical dilemmas*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45. https://doi.org/10.1145/365153.365168

Yahya, S. A., & Ali, M. A. (2023). The Prophet Muhammad's 'adālah: A catholicon for injustice in Nigeria. Al-Basirah, 13 (2), 61–78.

Zarqa, M. A. (2020). Maqasid al-Shari'ah in the ethical governance of artificial intelligence. *Journal of AI and Ethics*, 1(1), 45–58. https://doi.org/10.1007/s43681-020-00005-4