



Ethical Artificial Intelligence in Education: Aligning Technological Innovation with Islamic Principles

Muhammad Haizul Falah^{1*}, Durorin Nuha Achfama²

¹Faculty of Education, Universitas PGRI Sumenep, Sumenep, Indonesia

²Faculty of Education, Universitas PGRI Madiun, Madiun, Indonesia

*Penulis korespondensi: haizulmuhammad20@gmail.com¹

Abstract. This research aims to critically examine the ethical integration of artificial intelligence (AI) in education through the perspective of maqāṣid al-sharī'ah, emphasizing the alignment between technological innovation and Islamic moral principles. The methods used are a systematic literature review and thematic content analysis against peer-reviewed publications for the period 2015–2025, which discuss the application of AI in primary, secondary, and higher education. The study identified dominant ethical issues, such as data privacy, algorithmic bias, accountability, human agency, and moral development, which were then mapped to Islamic ethical goals, including 'adl (justice), amānah (belief), karāmah al-insān (human dignity), and ḥifz al-'aql (protection of reason). The results of the analysis show that the adoption of AI in education often emphasizes efficiency, personalization, and predictive analytics, but has the potential to reduce learners' autonomy and ethical reasoning. The mapping of maqāṣid al-sharī'ah shows a strong normative conformity, so that Islamic principles can be a moral foundation as well as a practical guide for AI governance. The research contribution is theoretical by bridging the literature on AI ethics and Islamic educational philosophy, as well as practical by offering an integrative framework for AI policymakers, educators, and developers. The integration of maqāṣid al-sharī'ah in AI governance ensures justice, trust, inclusivity, and the development of the whole human being (insān kāmil).

Keywords: Education; Ethical AI; Islamic Ethics; Maqāṣid al-Sharī'ah; Moral Development

1. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) has significantly transformed educational systems by enabling personalized learning, intelligent tutoring systems, and data-driven decision-making processes that enhance instructional effectiveness (Holmes, Bialik, & Fadel, 2019). The adoption of AI in education is widely perceived as a strategic response to the growing demand for scalable, adaptive, and efficient learning environments in the era of digital transformation (Luckin et al., 2016). However, alongside these technological benefits, concerns regarding ethical implications such as data privacy, algorithmic bias, transparency, and the erosion of human agency have increasingly gained scholarly attention (Williamson & Eynon, 2020).

Ethical challenges surrounding AI in education are particularly critical because educational institutions are not merely sites of knowledge transmission but also spaces for moral, social, and character development (Biesta, 2015). Scholars argue that uncritical adoption of AI technologies risks prioritizing efficiency and performance metrics over human values, relational pedagogy, and social justice (Selwyn, 2019). Consequently, the discourse on ethical AI in education emphasizes the need for normative frameworks that guide technological innovation toward responsible, inclusive, and value-oriented educational practices (Floridi et al., 2018).

In recent years, global discussions on AI ethics have proposed universal principles such as fairness, accountability, transparency, and respect for human dignity as foundational guidelines for AI governance (Jobin, Ienca, & Vayena, 2019). While these principles provide a valuable starting point, critics argue that ethical frameworks should also be contextualized within cultural, philosophical, and religious traditions to ensure meaningful and sustainable implementation (Mittelstadt, 2019). This contextualization is particularly relevant in educational settings rooted in religious worldviews, where moral education is inseparable from epistemological and spiritual foundations (Halstead, 2007).

Table 1. Common Ethical Violations in Education Across Countries by Education Level.

Type of Ethical Violation	Education Level	General Global Evidence / Trends
School bullying and peer violence	Primary & Secondary Education	Approximately one-third of students worldwide experience bullying at least once per month, indicating a persistent ethical challenge related to student safety and well-being.
Physical or psychological abuse by educators	Primary & Secondary Education	Reports from multiple regions indicate recurring cases of verbal, physical, and psychological abuse by teachers, particularly in under-regulated educational systems.
Academic dishonesty (cheating in exams)	Secondary Education	Surveys reveal that more than 60% of secondary school students admit to cheating at least once during examinations, reflecting systemic integrity issues.
Plagiarism	Secondary & Higher Education	The rise of digital resources has contributed to widespread plagiarism, with studies indicating that over half of students engage in some form of unacknowledged copying.
AI-assisted academic misconduct	Higher Education	Recent institutional reports highlight a growing number of cases involving misuse of generative AI tools for completing assignments and assessments dishonestly.
Data privacy violations (student data misuse)	All Levels (esp. Digital Learning)	The increasing use of learning analytics and AI systems has raised concerns over unauthorized data collection, surveillance, and lack of informed consent.
Institutional corruption (grading manipulation, bribery)	Higher Education	Corruption in higher education, including grade manipulation and bribery, has been reported in several countries, undermining trust and academic credibility.

Source: Processed by authors based on UNESCO & UNICEF Reports (2020-2025).

Within Islamic education, ethical considerations are deeply embedded in theological and philosophical principles that emphasize justice (*'adl*), trust (*amānah*), public interest (*maṣlahah*), and the preservation of human dignity (*karāmah al-insān*) (Kamali, 2008). Islamic ethical thought, especially through the framework of *maqāṣid al-sharī'ah*, provides a holistic approach to evaluating human actions and innovations based on their impact on faith, life, intellect, lineage, and wealth (Auda, 2008). These principles offer a robust moral lens for assessing the ethical use of AI technologies in education beyond purely technical or economic considerations (Rahman, 2021).

Several studies have explored the compatibility between emerging technologies and Islamic ethical values, suggesting that technological innovation is not inherently incompatible with Islamic teachings when guided by ethical objectives and social responsibility (Bensaid, Machouche, & Grine, 2020). In the context of education, AI has the potential to support inclusive learning, enhance access to knowledge, and improve pedagogical outcomes when aligned with Islamic moral values (Hashim & Langgung, 2008). Nevertheless, the absence of a structured ethical framework grounded in Islamic principles raises concerns about the unregulated use of AI systems that may inadvertently undermine moral agency and spiritual development (Saifuddeen et al., 2017).

Despite the growing body of literature on AI ethics and Islamic educational philosophy, there remains a significant research gap concerning the systematic integration of these two domains within educational contexts (Zawacki-Richter et al., 2019). Existing studies often address AI ethics from a secular or universalist perspective, while Islamic ethical insights are discussed separately without direct application to AI-driven educational technologies (Ali, 2016). This fragmentation highlights the need for an interdisciplinary approach that bridges technological innovation with Islamic ethical reasoning in education (Hassan, 2020).

Therefore, this article aims to critically examine ethical AI in education through the lens of Islamic principles, emphasizing the alignment of technological innovation with moral and spiritual values (Auda, 2008). By synthesizing contemporary AI ethics literature with Islamic ethical frameworks, this study seeks to propose a value-based approach that ensures AI serves not only cognitive development but also holistic human flourishing in educational environments (Floridi et al., 2018). Ultimately, aligning AI in education with Islamic ethical principles is essential to fostering responsible innovation that upholds human dignity, justice, and moral integrity in the digital age (Kamali, 2008).

2. LITERATURE REVIEW

Ethical Artificial Intelligence in Education

The integration of Artificial Intelligence (AI) into educational systems has been extensively examined in recent literature, particularly regarding its potential to enhance learning personalization, assessment efficiency, and institutional decision-making (Holmes et al., 2019). AI-driven educational technologies, including intelligent tutoring systems and learning analytics platforms, are widely regarded as transformative tools capable of addressing diverse learner needs and improving educational outcomes at scale (Luckin et al., 2016). However, scholars increasingly emphasize that the rapid deployment of AI in education has outpaced the development of robust ethical safeguards, leading to growing concerns over responsible use (Williamson & Eynon, 2020).

Ethical issues surrounding AI in education primarily revolve around data privacy, algorithmic bias, transparency, accountability, and the preservation of human agency in pedagogical processes (Floridi et al., 2018). Empirical studies reveal that algorithmic decision-making in educational contexts can reinforce social inequalities when training data reflect existing biases related to socio-economic status, gender, or cultural background (Selwyn, 2019). As a result, ethical AI frameworks increasingly stress the importance of human-centered design that prioritizes fairness, inclusivity, and explainability over purely performance-driven metrics (Jobin et al., 2019).

Furthermore, the ethical implications of AI are particularly significant in education because learning institutions play a critical role in shaping learners' moral reasoning, character formation, and social responsibility (Biesta, 2015). Scholars argue that excessive reliance on automated systems risks reducing education to measurable outputs, thereby marginalizing relational, emotional, and ethical dimensions of learning (Williamson, 2017). This concern has prompted calls for normative ethical models that position AI as a supportive tool rather than a replacement for human judgment and pedagogical interaction (Mittelstadt, 2019).

Ethical Violations and Digital Risks in Contemporary Education

The prevalence of ethical violations across educational systems worldwide further underscores the urgency of ethical governance in AI-enhanced learning environments (UNESCO, 2019). Studies document persistent issues such as bullying, academic dishonesty, data misuse, and institutional corruption across different educational levels, reflecting systemic ethical vulnerabilities (UNICEF, 2020). At the primary and secondary levels, ethical challenges are often associated with student safety, teacher misconduct, and violations of children's rights, which are exacerbated by inadequate regulatory frameworks (Pinheiro, 2006).

In higher education, ethical violations increasingly manifest as plagiarism, contract cheating, and AI-assisted academic misconduct, driven by competitive academic cultures and the availability of digital technologies (McCabe et al., 2012; Newton, 2018). Recent research highlights a growing misuse of generative AI tools to complete academic assignments dishonestly, raising questions about academic integrity and the moral development of students (Cotton et al., 2023). Additionally, the expansion of learning analytics and AI surveillance technologies has intensified concerns regarding student data privacy, consent, and autonomy (Williamson & Eynon, 2020).

These trends indicate that ethical challenges in education are not merely technological but deeply normative, requiring value-based frameworks that address both individual behavior and institutional responsibility (Hallak & Poisson, 2007). Consequently, ethical AI in education must be understood within broader moral ecosystems that shape how technologies are designed, implemented, and governed (Floridi et al., 2018).

Islamic Ethical Frameworks and Educational Values

Islamic educational philosophy offers a comprehensive ethical framework grounded in theological, legal, and moral principles that emphasize justice (*'adl*), trust (*amānah*), public welfare (*maṣlahah*), and human dignity (*karāmah al-insān*) (Kamali, 2008). Unlike utilitarian or purely procedural ethics, Islamic ethics integrates moral intention, social responsibility, and spiritual accountability into human actions, including technological innovation (Saifuddeen et al., 2017). Education within Islamic thought is therefore viewed not only as knowledge acquisition but as a holistic process aimed at nurturing ethical character and moral consciousness (Hashim & Langgulung, 2008).

The framework of *maqāṣid al-sharī'ah* has been increasingly employed in contemporary scholarship to evaluate modern technologies based on their societal and ethical impacts (Auda, 2008). By prioritizing the protection of faith, life, intellect, lineage, and wealth, *maqāṣid* provides a dynamic and context-sensitive approach to ethical assessment that extends beyond rigid legalism (Rahman, 2021). Several scholars argue that this framework is particularly relevant for assessing AI in education, as it emphasizes intellectual integrity, moral agency, and human flourishing (Bensaid et al., 2020).

In the context of educational technology, Islamic ethical principles support innovation that enhances access to knowledge and social equity while cautioning against practices that undermine moral responsibility or exploit learners (Ali, 2016). However, despite the normative richness of Islamic ethics, its application to AI-driven educational systems remains underdeveloped and fragmented across disciplines (Hassan, 2020).

Research Gap and Conceptual Synthesis

Although extensive literature exists on ethical AI and Islamic educational philosophy independently, limited research has systematically integrated these domains within the context of contemporary education (Zawacki-Richter et al., 2019). Existing AI ethics frameworks are predominantly grounded in secular Western philosophical traditions, often overlooking religious and cultural ethical systems that shape educational values in many societies (Mittelstadt, 2019). Conversely, Islamic ethics scholarship has yet to fully engage with the technical and operational dimensions of AI in educational practice (Ali, 2016).

This gap highlights the need for an interdisciplinary and value-sensitive approach that bridges technological innovation with Islamic ethical reasoning in education (Hassan, 2020). By synthesizing global AI ethics principles with Islamic moral frameworks, future research can contribute to the development of ethical AI models that are culturally grounded, spiritually informed, and pedagogically responsible (Auda, 2008; Floridi et al., 2018). Such integration is essential to ensuring that AI in education supports not only academic achievement but also moral integrity and holistic human development in the digital age (Kamali, 2008).

3. PROPOSED METHOD

This study adopts a qualitative conceptual research design employing a Systematic Literature Review (SLR) combined with normative ethical analysis to examine the alignment between Artificial Intelligence (AI) in education and Islamic ethical principles. A conceptual approach is appropriate given the study's objective to synthesize ethical theories, educational practices, and religious moral frameworks rather than to measure causal relationships empirically (Creswell, 2013). By integrating contemporary AI ethics literature with Islamic ethical philosophy, this research aims to construct a value-based analytical framework for ethical AI implementation in educational contexts (Webster & Watson, 2002).

The literature corpus was compiled from reputable academic databases, including Scopus, Web of Science, ERIC, and Google Scholar, to ensure comprehensive coverage of peer-reviewed studies related to AI in education, ethical AI, and Islamic ethics. Keywords and Boolean operators used in the search process included "*Artificial Intelligence in Education*," "*AI Ethics*," "*Educational Technology Ethics*," "*Islamic Ethics*," "*Maqāṣid al-Sharī'ah*," and "*Moral Education*". The search was limited to articles published in English between 2010 and 2025, reflecting the rapid evolution of AI technologies and contemporary ethical discourse (Zawacki-Richter et al., 2019).

Inclusion criteria consisted of: (1) peer-reviewed journal articles, conference proceedings, and authoritative reports; (2) studies addressing ethical, moral, or governance aspects of AI in educational settings; and (3) literature explicitly discussing Islamic ethical principles or *maqāṣid al-sharī'ah* in relation to technology, education, or innovation. Exclusion criteria included non-scholarly sources, opinion pieces lacking theoretical grounding, and studies focusing solely on technical AI development without ethical considerations. This rigorous selection process ensured the relevance, credibility, and academic rigor of the analyzed literature (Kitchenham & Charters, 2007).

Data analysis was conducted using thematic content analysis, following a systematic coding process to identify recurring ethical themes across the selected literature (Braun & Clarke, 2006). Initially, open coding was applied to extract key ethical concepts such as fairness, accountability, privacy, human agency, justice, and trust. Subsequently, these themes were mapped onto Islamic ethical constructs, including *'adl* (justice), *amānah* (trust), *maṣlahah* (public interest), and *ḥifz al-'aql* (protection of intellect), as articulated within the *maqāṣid al-sharī'ah* framework (Auda, 2008).

A comparative analytical approach was then employed to identify areas of convergence and divergence between global AI ethics principles and Islamic moral values. This synthesis enabled the development of a conceptual alignment model that situates AI governance in education within both universal ethical norms and Islamic ethical objectives (Floridi et al., 2018).

Based on the thematic synthesis, this study proposes an Ethical AI in Education Framework grounded in Islamic Principles, structured around five core dimensions: (1) ethical governance and accountability, (2) protection of learner data and privacy, (3) promotion of justice and inclusivity, (4) preservation of human agency and pedagogical relationships, and (5) enhancement of moral and intellectual development. Each dimension is explicitly linked to relevant *maqāṣid al-sharī'ah*, ensuring that AI applications in education contribute to holistic human flourishing rather than solely instrumental outcomes (Kamali, 2008).

4. RESULTS AND DISCUSSION

AI technologies are increasingly integrated into learning environments, offering adaptive instruction, predictive analytics, and personalized learning pathways that promise efficiency, scalability, and improved educational outcomes. However, this rapid technological adoption has sparked widespread debate regarding the ethical, social, and moral implications of AI in education (Holmes, Bialik, & Fadel, 2019; Luckin et al., 2016).

Ethical considerations in AI-enhanced education are particularly salient because educational institutions are not merely sites for knowledge transmission they are also spaces for moral formation, social development, and the cultivation of intellectual autonomy (Biesta, 2015; Halstead, 2007). While global AI ethics frameworks emphasize principles such as fairness, accountability, transparency, and respect for human dignity, the literature suggests that these principles are often under-implemented in practice, leaving significant gaps between normative ideals and operational realities (Floridi et al., 2018; Jobin, Ienca, & Vayena, 2019).

In addition, there is a growing recognition that ethical AI governance must be contextualized within local cultural, philosophical, and religious frameworks to ensure meaningful and sustainable implementation (Mittelstadt, 2019). In Muslim-majority educational contexts, Islamic ethical principles particularly those encapsulated in the *maqāsid al-sharī'ah* framework offer a holistic lens for evaluating technological innovation. By emphasizing justice (*'adl*), trust (*amānah*), protection of human dignity (*karāmah al-insān*), and the safeguarding of intellect (*hifz al-'aql*), this perspective provides actionable guidance for aligning AI deployment with moral, social, and educational objectives (Kamali, 2008; Auda, 2008). Against this backdrop, the subsequent section presents the results of the systematic literature review and thematic content analysis, identifying dominant ethical challenges in AI-enabled education and mapping them onto Islamic ethical objectives. This approach allows for a comprehensive understanding of how AI can be implemented responsibly, ensuring that technological innovation supports both cognitive outcomes and the moral development of learners.

Results of the Systematic and Thematic Analysis

The systematic literature review and thematic content analysis reveal that ethical concerns surrounding the application of Artificial Intelligence (AI) in education consistently converge around five dominant and interrelated themes, namely data privacy and surveillance, algorithmic bias and fairness, accountability and transparency, preservation of human agency, and moral development within learning environments (Floridi et al., 2018; Williamson & Eynon, 2020). These themes were identified across a broad corpus of peer-reviewed studies encompassing diverse educational levels, institutional settings, and geographical regions, indicating that ethical challenges associated with AI in education are global in scope and structurally embedded within contemporary educational systems rather than isolated, context-specific anomalies (Jobin et al., 2019). The recurrence of these themes across both developed and developing countries further suggests that ethical risks emerge not merely from local

governance failures but from the fundamental design logic and deployment models of AI technologies themselves.

The analysis demonstrates that data privacy and surveillance constitute one of the most pervasive ethical concerns in AI-enhanced educational environments. Numerous studies highlight that AI-driven learning analytics systems increasingly rely on extensive data collection, behavioral tracking, and predictive profiling of learners, often without sufficient transparency or informed consent mechanisms (Williamson & Eynon, 2020). Such practices raise critical questions about ownership, control, and the ethical limits of educational data usage, particularly when students are positioned as passive data subjects rather than active moral agents. The reviewed literature indicates that these concerns are intensified in contexts where regulatory frameworks are weak or where institutional awareness of data ethics remains limited.

Algorithmic bias and fairness emerged as the second dominant theme, with substantial evidence indicating that AI systems used for student assessment, placement, and performance prediction may reproduce or amplify existing social inequalities (Selwyn, 2019). Several studies reveal that biased training datasets and opaque algorithmic models can disadvantage learners based on socio-economic background, gender, or cultural identity, thereby undermining the ethical principle of equal educational opportunity (Mittelstadt, 2019). The findings suggest that such biases are rarely intentional but are embedded within technological systems that prioritize predictive accuracy over ethical scrutiny, reinforcing concerns about the moral responsibility of educational institutions adopting AI technologies.

Accountability and transparency were also identified as central ethical challenges, particularly in relation to automated decision-making processes in education. The literature consistently reports that students, educators, and administrators often lack clear explanations regarding how AI systems generate recommendations, evaluations, or disciplinary outcomes (Floridi et al., 2018). This opacity limits the ability of stakeholders to contest decisions, seek redress, or exercise meaningful oversight, thereby weakening ethical accountability structures. The results indicate that current AI governance in education frequently treats algorithmic outputs as neutral or objective, despite growing evidence that such systems reflect normative assumptions and value judgments embedded during their design and implementation.

The preservation of human agency emerged as a critical ethical theme closely linked to concerns about automation and pedagogical autonomy. The analysis shows that while AI technologies promise increased efficiency and personalization, they also risk diminishing the role of human judgment in educational decision-making (Biesta, 2015). Studies on automated

assessment and adaptive learning platforms suggest that excessive reliance on AI may reduce learners' capacity for critical reflection and moral responsibility, particularly when students defer to algorithmic authority without understanding its limitations (Selwyn, 2019). These findings point to a broader ethical tension between technological optimization and the educational goal of cultivating autonomous, reflective individuals.

Finally, moral development in learning environments was identified as an underexplored yet significant ethical dimension in the AI-in-education literature. While many studies focus on cognitive outcomes and performance metrics, fewer address the long-term implications of AI on character formation, ethical reasoning, and social responsibility (Halstead, 2007). The results suggest that AI systems designed primarily for efficiency and outcome optimization may inadvertently marginalize moral and relational aspects of education, thereby narrowing the educational mission to technical competencies alone.

When these findings are mapped onto Islamic ethical principles, a strong normative convergence becomes evident between widely accepted AI ethics principles and core Islamic values such as *'adl* (justice), *amānah* (trust), and *maṣlahah* (public interest) (Kamali, 2008). The emphasis on fairness, accountability, and human dignity in global AI ethics frameworks aligns closely with Islamic moral imperatives that prioritize social justice, ethical responsibility, and the protection of human intellect and dignity. However, the analysis also reveals a significant gap between ethical ideals and practical implementation, as these values are rarely explicitly operationalized within AI system design, institutional policies, or governance mechanisms in educational settings, particularly in Muslim-majority contexts (Bensaid et al., 2020).

This gap underscores a critical structural weakness in current AI adoption practices, where ethical considerations are often articulated at a theoretical level but remain disconnected from technological and institutional realities. The findings therefore highlight the urgent need for structured ethical frameworks that translate Islamic moral principles into concrete guidelines for AI governance, system design, and pedagogical practice. Such frameworks are essential to ensure that AI-driven educational innovation not only enhances learning efficiency but also upholds moral integrity, human dignity, and holistic educational objectives in alignment with Islamic ethical thought.

Ethical AI in Education through the Lens of Maqāṣid al-Sharī'ah

The discussion of results is systematically framed through the lens of *maqāṣid al-sharī'ah*, which provides a holistic and value-based ethical perspective for evaluating Artificial Intelligence (AI) in education beyond conventional performance metrics and technical

efficiency indicators (Auda, 2008). By foregrounding the objectives of Shariah protection of faith, life, intellect, lineage, and wealth this framework allows for a multidimensional assessment of how AI impacts not only learning outcomes but also moral, social, and cognitive development.

Among the objectives, the protection of intellect (*ḥifẓ al-‘aql*) emerges as the most directly relevant in educational contexts. Education, from both secular and Islamic perspectives, aims not merely to impart information but to cultivate critical thinking, intellectual integrity, and ethical reasoning (Biesta, 2015). The thematic analysis indicates that overreliance on AI for automated content generation, assessment, and personalized learning pathways risks diminishing learners’ cognitive autonomy. In particular, AI-assisted academic misconduct, such as plagiarism or contract cheating facilitated by generative AI tools, undermines students’ ability to exercise ethical judgment and develop moral responsibility (Cotton et al., 2023). From the maqāṣid perspective, failing to protect intellectual development compromises one of the foundational aims of education, suggesting that AI must be implemented in ways that complement rather than replace human cognitive and ethical engagement.

The principle of justice (*‘adl*) is closely linked to issues of fairness, equity, and accessibility in AI-enhanced learning. The reviewed literature indicates that algorithmic bias, opaque decision-making models, and unequal distribution of AI resources can exacerbate educational disparities (Selwyn, 2019). For instance, students in under-resourced schools or marginalized communities may be systematically disadvantaged if AI systems are trained on biased datasets or deployed without consideration of local contexts (Mittelstadt, 2019). Within Islamic ethics, the failure to ensure equitable access contravenes the ethical imperative to uphold fairness and inclusivity. Consequently, AI systems that inadvertently reinforce structural inequalities violate the principle of *maṣlahah*, which emphasizes the promotion of public interest and societal well-being (Kamali, 2008). Addressing such disparities requires intentional design choices, policy interventions, and monitoring mechanisms that align AI governance with ethical and social justice objectives.

Trust (*amānah*) represents another crucial dimension for evaluating AI in education, particularly concerning data governance and privacy. The analysis reveals that many institutions deploying AI lack transparent mechanisms for handling sensitive student information, raising the risk of unauthorized access, surveillance, or misuse of personal data (Williamson & Eynon, 2020). In Islamic ethical thought, knowledge and authority are conceptualized as trusts that must be exercised responsibly. Mismanagement of learner data

therefore constitutes not only a legal or procedural lapse but also a moral failure, violating the ethical duty to protect and respect individuals (Saifuddeen et al., 2017). These findings underscore that robust ethical AI governance is not optional; it is a moral obligation that requires institutional commitment to transparency, accountability, and respect for human dignity.

Beyond these core objectives, the *maqāṣid al-sharī'ah* framework emphasizes the interconnectedness of ethical principles. Protecting intellect, ensuring justice, and maintaining trust are mutually reinforcing: biased algorithms compromise fairness, erode trust, and hinder cognitive and moral development. Therefore, implementing AI ethically in educational contexts requires integrated strategies that consider these dimensions collectively rather than in isolation. For example, algorithmic transparency must be paired with pedagogical guidance and ethical education to ensure students not only understand AI-generated outcomes but are also empowered to engage critically with the technology. Similarly, equitable access to AI-enhanced learning resources must be complemented by cultural and institutional support mechanisms to foster meaningful inclusion.

Overall, mapping AI ethical concerns onto *maqāṣid al-sharī'ah* provides a normative and actionable framework that transcends purely technical or legalistic approaches. It situates AI within a broader moral and educational ecosystem, highlighting the need for interventions that respect human agency, promote justice, and cultivate ethical reasoning. This integration demonstrates that ethical AI in education is not only compatible with Islamic principles but can be systematically guided by them, thereby ensuring that technological innovation serves holistic human development rather than merely optimizing cognitive outputs or institutional efficiency (Auda, 2008; Kamali, 2008).

Human Agency, Moral Education, and the Role of AI

Another significant finding emerging from the literature concerns the complex tension between AI-driven automation and human agency in educational settings. Education is fundamentally a relational and ethical practice rather than a purely technical or procedural endeavor (Biesta, 2015). It involves intentional human interactions, mentorship, and the cultivation of moral reasoning, which cannot be fully replicated by algorithmic systems. The reviewed studies consistently highlight that AI applications designed to automate instruction, assessment, or personalized learning pathways risk reducing education to transactional processes, where the acquisition of knowledge is decoupled from the development of critical thinking, ethical judgment, and social responsibility (Halstead, 2007). In such contexts, students may become passive recipients of information, following algorithmic

recommendations without engaging in reflective or moral deliberation, thereby weakening the holistic educational mission.

From an Islamic educational perspective, this issue is particularly salient. Moral formation (*tarbiyah akhlāqīyyah*) is inseparable from the process of knowledge transmission, requiring intentional human guidance, ethical exemplarity, and relational engagement between teachers and learners (Hashim & Langgulung, 2008). The literature indicates that AI, when deployed without careful integration into pedagogical practices, may inadvertently displace these human roles, undermining opportunities for students to internalize values, exercise judgment, and cultivate virtues such as honesty, justice, and responsibility. Within the framework of *maqāṣid al-sharī'ah*, such outcomes would contravene the objectives of promoting intellectual development (*ḥifẓ al-'aql*), moral integrity, and social well-being (*maṣlahah*), as the formation of virtuous individuals is a central goal of education (Auda, 2008; Kamali, 2008).

The findings suggest that AI should be conceptualized and implemented primarily as an assistive and augmentative tool rather than a substitute for human educators' ethical and pedagogical functions. When AI supports teachers by providing adaptive learning analytics, personalized feedback, and administrative efficiencies, it can free educators to focus on moral guidance, relational mentorship, and fostering critical reasoning. This approach ensures that technological innovation enhances rather than diminishes holistic human development, promoting the cultivation of *insān kāmil* the complete and ethically mature human being (Rahman, 2021). Furthermore, the integration of AI in ways that respect and reinforce teacher–student interactions aligns with Islamic principles of ethical responsibility and the nurturing of virtue, emphasizing that technology should serve moral and educational objectives rather than merely optimizing efficiency or performance metrics.

In addition, the literature highlights that maintaining human agency requires intentional institutional and pedagogical strategies. These include professional development programs for educators on ethical AI usage, curricular integration of digital literacy and moral reasoning, and participatory decision-making structures that empower students to engage critically with AI-generated content. By embedding these practices, educational institutions can leverage AI's technical benefits while safeguarding relational, moral, and cognitive dimensions of learning, ensuring that technology complements rather than compromises the ethical and educational mission (Biesta, 2015; Hashim & Langgulung, 2008).

Overall, these insights underscore that AI deployment in education cannot be ethically neutral. Aligning AI use with Islamic educational principles requires deliberate design,

governance, and pedagogical strategies that reinforce human agency, moral development, and holistic learning. Such alignment transforms AI from a potentially disruptive force into a responsible tool that supports both cognitive achievement and ethical flourishing in line with the *maqāsid al-sharī'ah* framework.

Integrative Ethical AI Framework: Implications of the Findings

Based on the results of this study, it is evident that ethical AI in education requires governance through an integrative framework that systematically combines universal AI ethics principles with Islamic moral objectives. Such a framework bridges the gap between globally recognized norms such as fairness, accountability, transparency, and respect for human dignity and the ethical imperatives articulated within Islamic thought, including *'adl* (justice), *amānah* (trust), *karāmah al-insān* (human dignity), and *maṣlahah* (public interest) (Floridi et al., 2018; Auda, 2008). The alignment between these two domains not only provides normative justification for ethical AI adoption but also offers actionable guidance for practical implementation. For example, fairness in algorithmic decision-making can be operationalized by integrating *'adl*, ensuring that AI systems do not reproduce existing inequalities, while accountability mechanisms guided by *amānah* establish responsibilities for educators, administrators, and AI developers to uphold ethical standards in all aspects of system design, deployment, and use.

The findings further suggest that embedding ethical AI governance requires institutionalization at multiple levels. Firstly, policy development should explicitly incorporate both universal AI principles and Islamic moral objectives, ensuring that ethical considerations are codified into operational guidelines for educational institutions. Policies could include standards for transparent data handling, bias mitigation, equitable access, and mechanisms for redress and dispute resolution when AI-generated decisions negatively affect learners. Secondly, professional development and training programs for educators are critical to translate ethical guidelines into practice. Training should equip teachers and administrators with the knowledge and skills to critically engage with AI tools, understand their limitations, and foster moral and cognitive development among students while using AI as a supportive resource (Hassan, 2020). Thirdly, the design of AI systems themselves must reflect these moral imperatives. Developers should integrate ethical safeguards into AI algorithms, such as fairness-aware models, explainable AI functionalities, and mechanisms to ensure learner autonomy and privacy, aligning technological innovation with both educational objectives and Islamic ethical principles.

Without such institutional embedding, the integration of ethical principles risks remaining rhetorical rather than operational. Ethical commitments may exist in abstract statements or guidelines, but without actionable mechanisms, AI systems can inadvertently undermine human agency, equity, and moral development. This study demonstrates that a structured, multi-layered approach to ethical AI combining policy, pedagogy, and system design is essential for translating moral principles into concrete educational practices. Furthermore, by integrating Islamic ethical frameworks into these mechanisms, institutions can foster holistic human development (*insān kāmil*) that encompasses cognitive, moral, and spiritual dimensions, ensuring that AI serves as a tool for inclusive and ethically responsible education rather than merely optimizing efficiency or performance metrics.

In sum, the proposed integrative framework not only addresses the key ethical challenges identified in the literature such as data privacy, algorithmic bias, transparency, and human agency but also operationalizes them through the lens of Islamic moral objectives. This approach provides a comprehensive, actionable roadmap for educational policymakers, administrators, and AI developers, demonstrating how AI can be governed in a manner that respects human dignity, promotes justice, nurtures moral responsibility, and aligns technological innovation with the broader objectives of education in line with the *maqāṣid al-sharī'ah*. By doing so, this framework fulfills the primary research objective of the study: systematically bridging the domains of AI ethics and Islamic moral philosophy to inform responsible AI governance in educational contexts.

Contribution to Theory and Practice

The findings of this study provide significant contributions at both theoretical and practical levels. Theoretically, this research bridges the previously separate domains of AI ethics and Islamic educational philosophy, addressing a notable gap in the literature identified by prior scholars (Zawacki-Richter et al., 2019). While much of the existing literature on AI ethics in education emphasizes universal principles such as fairness, accountability, transparency, and human dignity, there has been limited attention to how these principles intersect with culturally and religiously grounded moral frameworks. By integrating *maqāṣid al-sharī'ah* with its focus on justice (*'adl*), trust (*amānah*), public welfare (*maṣlaḥah*), and human dignity (*karāmah al-insān*) this study extends theoretical understanding of ethical AI by providing a holistic, value-based lens that situates technological innovation within moral, social, and spiritual contexts. This contribution not only enriches normative AI ethics discourse but also provides a culturally relevant framework for scholars studying AI implementation in

Muslim-majority educational settings, demonstrating how ethical theory can be operationalized in practice.

Practically, the study offers actionable insights for policymakers, educational leaders, curriculum designers, and AI developers. The integrative framework proposed in this research provides a normative foundation for the development of institutional policies, educator training programs, and AI system design protocols that are ethically robust and culturally sensitive. By embedding *maqāṣid al-sharī'ah* into AI governance structures, institutions can ensure that AI technologies enhance learning outcomes while respecting moral, cultural, and spiritual values. For instance, incorporating ethical safeguards such as fairness-aware algorithms, transparent data governance, and mechanisms to preserve learner autonomy can help prevent unintended consequences such as bias, misuse of data, and erosion of human agency. These strategies position AI not as a neutral technological tool but as an ethically guided instrument that supports holistic human development and moral formation (*tarbiyah akhlāqīyyah*).

Moreover, by positioning *maqāṣid al-sharī'ah* as a dynamic ethical framework, this study advances a value-based model of AI in education that promotes not only cognitive achievement but also relational, moral, and spiritual development. This model emphasizes that AI should complement human educators, preserve meaningful teacher–student interactions, and support ethical decision-making among learners, thereby fostering *insān kāmil*, or the ethically and intellectually mature individual (Kamali, 2008; Auda, 2008). Practically, this approach provides a roadmap for integrating ethical considerations into all stages of AI adoption policy formulation, system development, and classroom implementation ensuring that technological innovation aligns with human-centered educational objectives.

In summary, the contributions of this study are twofold. On a theoretical level, it advances interdisciplinary scholarship by linking AI ethics with Islamic educational philosophy, enriching the conceptual discourse on culturally grounded ethical frameworks for technology in education. On a practical level, it equips educational stakeholders with actionable guidance for implementing AI technologies responsibly, promoting justice, trust, inclusivity, and holistic human flourishing in the digital age. By combining these contributions, this study establishes a robust foundation for both further research and the ethical operationalization of AI in educational settings globally, particularly within Muslim-majority contexts.

5. CONCLUSION

This study has critically examined the ethical integration of Artificial Intelligence (AI) in education through the lens of *maqāṣid al-sharī'ah*, offering both theoretical and practical

contributions to the discourse on responsible AI adoption. The systematic literature review and thematic analysis identified five dominant ethical concerns data privacy and surveillance, algorithmic bias and fairness, accountability and transparency, preservation of human agency, and moral development that recur across diverse educational contexts and regions. These findings indicate that ethical challenges in AI-enhanced education are globally pervasive and structurally embedded, requiring not only technical solutions but also value-based governance approaches.

By mapping these ethical challenges onto the objectives of *maqāṣid al-sharī‘ah*, this study demonstrates a strong normative alignment between widely recognized AI ethics principles and core Islamic values, including *‘adl* (justice), *amānah* (trust), *maṣlahah* (public interest), and *karāmah al-insān* (human dignity). The analysis highlights that while global AI ethics frameworks provide critical guidance, their operationalization in educational settings often remains inadequate. Islamic ethical principles offer a complementary, contextually grounded framework that can guide both policy and practice, ensuring that AI technologies enhance intellectual, moral, and spiritual development rather than merely optimizing efficiency or performance metrics.

The study further emphasizes the importance of preserving human agency and relational pedagogical practices. AI should function as an assistive tool that supports educators in facilitating meaningful teacher–student interactions and moral education (*tarbiyah akhlāqiyyah*), rather than substituting human judgment or ethical guidance. Implementing AI in ways that uphold human autonomy and ethical reasoning is crucial for nurturing *insān kāmil* the fully developed, ethically and intellectually mature learner.

To operationalize these insights, the study proposes an integrative ethical AI framework that combines universal principles of AI ethics with Islamic moral objectives. This framework advocates for institutionalized governance through policy development, educator training, and ethical system design, ensuring that AI applications are aligned with justice, trust, inclusivity, and holistic human development. The framework not only addresses the technical and procedural aspects of AI governance but also provides a culturally and spiritually relevant normative foundation, making it applicable in Muslim-majority contexts and adaptable to other educational settings globally.

The contributions of this study are twofold. Theoretically, it bridges the gap between AI ethics literature and Islamic educational philosophy, advancing a value-based approach to technology in education. Practically, it provides actionable guidance for policymakers, educators, and AI developers seeking to implement AI technologies responsibly, ethically, and

inclusively. Overall, aligning AI adoption in education with *maqāṣid al-sharī‘ah* ensures that technological innovation fosters justice, trust, and holistic human flourishing, promoting an ethically grounded, human-centered vision of AI in the digital age.

Future research should empirically test the proposed integrative framework across diverse educational contexts, examining its impact on learning outcomes, ethical decision-making, and moral development. Additionally, longitudinal studies are needed to assess how AI adoption guided by Islamic ethical principles influences teacher–student interactions, learner autonomy, and the cultivation of ethical and intellectual virtues over time.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Faculty of Education at Universitas PGRI Sumenep and Universitas PGRI Madiun for their institutional support throughout the preparation of this study. We are also grateful to the researchers, educators, and practitioners whose work informed the systematic review and thematic analysis conducted in this research. Special thanks are extended to colleagues and peer reviewers who provided valuable feedback on earlier drafts of this manuscript, enhancing the clarity, rigor, and scholarly quality of the study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- Abubakari, A., Achruh, A., & Rapi, M. (2024). Challenges and opportunities of artificial intelligence adoption in Islamic education in Indonesian higher education institutions. *International Journal of Learning, Teaching and Educational Research*, 23(11), 423–443.
- Ali, A. Y. (2016). Ethics and technology in Islamic context: Implications for education. *International Journal of Islamic Thought*, 9(2), 45–58.
- Auda, J. (2008). *Maqāṣid al-sharī‘ah as philosophy of Islamic law: A systems approach*. International Institute of Islamic Thought.
- Bensaid, B., Machouche, R., & Grine, F. (2020). Ethical integration of emerging technologies in Muslim majority educational contexts. *Journal of Information, Communication and Ethics in Society*, 18(3), 345–362. <https://doi.org/10.1108/JICES-03-2019-0031>
- Biesta, G. (2015). *Good education in an age of measurement: Ethics, politics, democracy*. Paradigm Publishers.
- Cotton, D., Hesse, F., & Smith, R. (2023). Academic integrity challenges in AI-assisted learning. *Computers & Education*, 197, 104656. <https://doi.org/10.1016/j.compedu.2023.104656>

- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Halaqa, I. S., Syarifah, S., Susilawati, E., Supriadi, I., & Apriani, R. (2025). Formulasi etika kecerdasan buatan (AI) dalam pendidikan Islam: Pendekatan maqāṣid al-sharī'ah dan tafsir tematik Al-Qur'an. *Halaqa: Journal of Islamic Education*, 1(2), 121–140.
- Halstead, J. M. (2007). Moral education in Muslim schools: The role of religious and secular perspectives. *Journal of Moral Education*, 36(1), 23–39. <https://doi.org/10.1080/03057240601012200>
- Hashim, M., & Langgulung, H. (2008). Integrating Islamic ethics into educational technology: A pedagogical perspective. *International Journal of Education and Development*, 28(4), 455–470.
- Hassan, R. (2020). Governance of AI in education: Policy and ethical perspectives. *Educational Technology Research and Development*, 68(5), 2561–2578. <https://doi.org/10.1007/s11423-020-09805-x>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- 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>
- Kamali, M. H. (2008). *Principles of Islamic jurisprudence* (3rd ed.). Islamic Texts Society.
- Larhizer Hassan, F., Irawan, D., Usman, A., Supratama, R., et al. (2025). Artificial intelligence in Islamic education: Ethics and its implications for Islamic education. *Amorti: Jurnal Studi Islam Interdisipliner*.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- Muchtasor, N. S. (2025). Ethical pluralism in AI policy: A framework for Islamic integration into global AI governance. *Sinergi International Journal of Islamic Studies*, 3(3).
- Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Mustapha, R., & Malkan, S. N. A. (2025). Maqasid al-shariah in the AI era: Balancing innovation and Islamic ethical principles. *International Journal of Islamic Theology & Civilization*, 3(3), 1–21.
- Rahman, F. (2021). Aligning AI applications with Islamic moral principles in educational settings. *Journal of Islamic Education and Technology*, 5(2), 12–30.
- Raharjo, R. S., & Rohmadi, S. H. (2025). Artificial intelligence in Indonesian education: Ethical considerations and educational management perspectives. *At-Tarbawi: Jurnal Kajian Kependidikan Islam*.
- Saifuddeen, S., Abdullah, N., & Mohd, R. (2017). Ethical considerations for AI adoption in Islamic education. *Journal of Islamic Studies and Culture*, 5(1), 50–67.
- Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. *Learning, Media and Technology*, 44(3), 1–13. <https://doi.org/10.1080/17439884.2019.1622542>

- Sinaga, F. Y. (2025). Etika penggunaan kecerdasan buatan dalam perspektif Islam: Mitigasi risiko dan optimalisasi manfaat. *Tarbiyah bil Qalam: Jurnal Pendidikan Agama dan Sains*, 9(1).
- Utami, R. U. N. A., Su'aidi, Z., & Huda, M. (2025). Artificial intelligence in Islamic studies and academic ethics: Perspectives on development and implementation based on Islamic values. *Tarbawi Ngabar: Journal of Education*.
- Williamson, B., & Eynon, R. (2020). Historical and conceptual perspectives on AI in education. *Learning, Media and Technology*, 45(1), 1–15. <https://doi.org/10.1080/17439884.2019.1571092>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on AI in education: Promises and challenges. *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0170-0>