Philosophy & Technology (2025) 38:120
https://doi.org/10.1007/s13347-025-00922-4

RESEARCH ARTICLE

Islamic Ethics and AI: An Evaluation of Existing Approaches to AI using Trusteeship Ethics

Fatima Ali[1] · Karim Bouzoubaa[2,3] · Frank Gelli[4] · Boumediene Hamzi[5,6] · Suhair Khan[4]

Received: 21 January 2025 / Accepted: 29 June 2025

Abstract

Artificial Intelligence (AI) technologies are revolutionising key sectors such as healthcare, finance, and governance, while raising ethical challenges, including algorithmic bias, privacy violations, and environmental sustainability. Dominant Western ethical paradigms, such as Luciano Floridi's Information Ethics, emphasise procedural integrity and transparency but often lack spiritual and metaphysical grounding prevalent in global traditions. Most approaches to Islamic ethics for AI have employed Maqasid al-Shariah (objectives of Islamic law) and Qawaid Fiqhiyya (legal maxims), that apply legal principles to ethical questions but face limitations in addressing the complexities of emerging technologies especially that they, initially, were developed to address problems in Islamic law and not in Ethics.

This paper introduces the prospects of Taha Abdurrahman's I'timāni (trusteeship) framework as a unified ethical model for AI technologies. Rooted in the concept of divine trust (amana), the framework integrates three foundational covenants—Ontological, Epistemological, and Existential—offering a comprehensive vision of human responsibility toward God, knowledge, and creation. This approach provides ways to prioritise moral accountability with actionable governance strategies, as we demonstrate its practical applicability. We show that from the Islamic perspective, a philosophy of ethics approach is not only more appropriate, but more promising and comprehensive than ones based on jurisprudential rulings, as combating the ethical concerns of data-driven AI in a neoliberal environment requires an overhaul in worldview with cross-cultural respect and care for all.

Keywords Artificial Intelligence (AI) · Islamic Ethics · Trusteeship Ethics · Taha Abdurrahman · Divine Trust (Amana) · Ontological Covenant · Epistemological Covenant · Existential Covenant · Cross-cultural Collaboration · Ethical Governance · Algorithmic Bias

Extended author information available on the last page of the article

Published online: 09 August 2025

1 Introduction

Artificial Intelligence (AI) technologies are transforming modern society, driving innovation in healthcare, finance, and governance. From predictive analytics in medicine to automated decision-making in financial systems, AI promises unprecedented benefits. However, this progress is accompanied by profound ethical challenges, including algorithmic bias (racist and islamophobic biases, in particular— Abid et al., 2021; Hemmatian et al. 2022; Patel et al., 2023), systemic inequities, privacy violations, and environmental degradation as increased emissions are reported from data centres that power AI. The ethical challenges have become increasingly urgent with companies competing to deploy and release generative models in an expanding domain with no standard evaluation for safety. The increase in interest in AI ethics is well-documented in the recent 2025 AI Index Report from Stanford University, showing that there was a record high of 233 incidents and controversies related to AI in 2024, yet no consensus on benchmarks for responsible AI although there is agreement among companies and developers on benchmarks for general capabilities (Maslej et al., 2025). The report also showed that the number of papers on AI ethics published in 2024 was 1,278 (over half of which originated in the United States), an increase of 28.8% from the previous year.

Ethical discourse around AI has been dominated by Western paradigms, such as Luciano Floridi's Information Ethics (IE) (Floridi, 2011, 2023). IE extends moral considerations to all entities within the infosphere, emphasising transparency, accountability, and informational integrity. While valuable, these principles often lack cultural and metaphysical grounding, limiting their global applicability. As AI technologies increasingly shape societies with diverse cultural and spiritual orientations, the need for a globally inclusive and cross-cultural ethical framework is paramount. Ethical governance must transcend the limitations of secular, utilitarian approaches and incorporate metaphysical principles that resonate across traditions.

Both AI systems and publications on AI and ethics are heavily influenced by technologists and researchers in the United States (Chuang et al., 2022). This is a philosophically secular landscape, yet Sam Altman (OpenAI CEO) writes: "The challenge of who will lead on AI is not just about exporting technology, it's about exporting the values that the technology upholds." (Altman 2024). In the 2024 ACM Conference on Fairness, Accountability, and Transparency, the committee refused to make a stance on the AI used in the genocide in Palestine (Katibah, 2024), stating that "science should not mix with politics". It is clear that inclusive AI demands a systematic transformation that must be understood as an evolving process rather than a final state (Khan, 2024a, b). Buyl et al. (2025) further emphasize this perspective in their study highlighting how AI systems often mirror the ideologies of their developers.

The current global benchmark for ethics is based on UN Sustainable Development Goals. There is a problem in how to define 'AI for the Common Good' (Berendt, 2019) with varying philosophical constructs. Although techno philosophy

includes other schools of thought, including effective altruism, libertarianism and so on; all of the major technology AI companies are built within a traditional corporate governance structure with ultimate responsibility to the fiduciary duties to profit-seeking shareholders in a resource-constrained environment. The American business and legal landscape has allowed for most of the major corporations in AI to be rule-makers, to set their own evolving and dynamic ethical standards for what is moral.

Going beyond corporations and shareholder-led definitions of progress, the role of the collective in secular society should also be interrogated. Religious wisdom and indigenous knowledge provides an alternative to prevailing secular ideologies that are grounded in individualism and pragmatism, but instead strengthen our collective capabilities to live wisely and well. These paradigms have ontological roots and worldviews that are not built into current large language models.

What values are implied in the alignment problem? Is there an alternative ground epistemological and ontological framework that protects the 'human' with guidance on moral virtues that has universal appeal? In order to bridge communities worldwide, it is valuable to consider the future impact of artificial intelligence using the Islamic lens that unites across diverse cultures and ethnicities. Muslims are approximately a quarter of the world's population and roughly 10% more than those who are religiously unaffiliated[1]. With over four-fifths of the world's population affiliated with a religion, and as the Global South is increasingly impacted in severe ways, there is a growing need to not only include but lead with these voices in shaping the future directions of the AI technology landscape. Facilitating dialogue with communities traditionally alienated from the process will also increase the potential for such technologies to be adapted for their particular needs.

The governments of the Islamic world are also competing for rapid development in AI—the 'world's first AI fatwa service' was Virtual Ifta (IACAD, Dubai) and Sophia the first 'robotic citizen' by Saudi Arabia. Neom is the first smart city with AI as its 'beating heart' (Bell, 2022), yet is considered a 'surveillance city' (Farouk, 2022), among accusations of human rights violations and environmental damage. In these settings of heightened social inequality, it has become clear that there is some consideration for universal ethics, but hardly Islamic-inspired ethics (Azar & Haddad 2021; Chaudhary, 2020; Ghaly, 2023).

Philosophy and religion have historically played a central role in ethical frameworks. We can consider, in general, what the perspective of monotheistic religions offers relative to the secular position: a purpose, motivation and 'why' (or *telos*) to do a moral action; an objective morality, that cannot be achieved without God; feeling of responsibility and accountability; and virtue-based ideals in prophetic role models. Monotheistic religions provide both a purpose of life and a stable identity for humans, avoiding ontological crises such as issues relating to transhumanism. On postmodernism from a Christian perspective, Waters writes on the loss of purpose or telos:

[1] https://www.pewresearch.org/religion/feature/religious-composition-by-country-2010-2050/

> "The postmodern telos of no telos is in fact a telos of techne […] techne becomes telos with the recognition that the world's only underlying and universal feature is information. Since information has no inherent meaning it can be recast, conveyed and interpreted in virtually endless arrays." (Waters, 2006).

Heidegger's critique of modern technology further illuminates this shift. By introducing the concept of *enframing* (Gestell), he argues that technology reduces the world to a "standing-reserve": resources waiting to be exploited rather than a realm of intrinsic meaning. This enframing limits our engagement with the deeper ontological dimensions of being, pushing us toward a mode of existence that is detached and inauthentic. In contrast, the enduring ethical structures of monotheistic traditions nurture a rootedness in a transcendent source of meaning, thereby offering a buffer against the dehumanizing trends of technological reductionism. For a more extended Christian theological engagement with Heidegger's critique of technology, including the concept of enframing and its implications for religious thought in a technological age, see (Pattison, 2007).

In 2020, the Vatican sought "The Abrahamic Commitment to the Rome Call"[2], when representatives of Islam, Christianity, and Judaism, participated in a meeting at the Rome Summit and proposed five ethical principles that all AI designers must adhere to: "explainability", "comprehensiveness", "repeatability", "unbiased programs" and "humans always bear responsibility". Although the belief in one God is conceptually shared between the three religions, it seems to not have been taken into consideration on a philosophical level. The proposed principles already exist in the secular technological discourse, whereas there are spiritual aspects such as 'intentions' in Islamic tradition that have not been represented well. Islamic ethical traditions offer profound moral principles rooted in justice (*ʿadl*—عدل), intention (*niyyah*—نية), and trusteeship (*amāna*—أمانة), which prioritise societal well-being and spiritual accountability.

From the perspective of modern Islamic philosophy of ethics, we can consider several contemporary thinkers in the Islamic world. Taha AbdurRahman has made significant contributions to the understanding of ethics in the context of the Arabic-Islamic intellectual tradition. His approach to ethics (akhlaq—الأخلاق) revolves around grounding moral philosophy in both Islamic ontology and language philosophy, proposing an indigenous ethical framework that diverges from Western secular models. Rooted in the Qur'anic concept of divine trust (amana), the trusteeship (I'timāni—إئتماني) framework emphasises humanity's interconnected responsibilities toward God, knowledge, and creation. Its three foundational covenants — Ontological, Epistemological, and Existential — explicate ethical principles by integrating spiritual accountability, intellectual integrity, and practical stewardship.

This work explores the trusteeship (I'timāni) framework as foundational ethics for AI governance, analysing the Ontological, Epistemological, and Existential Covenants that can provide a deeper metaphysical grounding. We assess its augmentation

[2] romecall.org.

with the wider world, including practical applications and the role of AI ethicists. Existing ethical frameworks for AI, such as those based on Maqasid al-Shariah and Floridi's Information Ethics, are evaluated from the perspective of the trusteeship (I'timāni) framework.

2 Ethics and Jurisprudence

The word *ethics*, synonymous with moral philosophy, originates from the Greek word *ethos*, meaning character or disposition. Its Islamic counterpart is *akhlāq*, the plural of *khuluq*, which shares its root with *khalq* (creation), signifying the intrinsic link in Islamic metaphysics between inner character and outward form.

Despite this rich legacy, most Islamic approaches to AI ethics in contemporary literature adopt a *fiqhi* lens. *Fiqh*, the science of Islamic jurisprudence, concerns the derivation of laws not explicitly stated in the Qur'an or Sunnah. While jurisprudence is essential for regulating certain technological applications, law alone cannot encapsulate the moral and spiritual dimensions required to address the ethical complexities of emerging technologies like AI.

In Islamic sciences, *fiqh* and *akhlāq* are treated as distinct yet interconnected disciplines (Mutahhari, 2002). The classical works of Islamic ethics—such as Ṭūsī's *Akhlaq-i Nasiri*, Ibn Miskawayh's *Tahdhīb al-Akhlāq*, and al-Ghazālī's *Iḥyāʾ ʿUlūm al-Dīn*—engage with Greek philosophy and rational ethics, not as imitations but as spiritual and moral refinements. *Fiqh* provides the minimum legal obligations; *akhlāq* inspires the higher aims of religion—spiritual proximity to God. For example, while retributive justice may be allowed, the Qur'an encourages a higher moral response: "If you retaliate, then retaliate with the like of that which you were afflicted. But if you are patient, it is better for those who are patient" (Qur'an, 16:126).

The Prophet Muhammad ﷺ was sent as a mercy to all creation (Qur'an, 21:107) and as one who came to perfect moral character. This *best of akhlāq* is offered as a model for all of humanity, while *fiqh* is contextually confined to Muslims. Mohammad Ghaly considers this conflation of *fiqh* with Islamic ethics in his introduction to *Islamic Ethics and the Genome Question* (Ghaly, 2019). While both are needed and related—akin to physics and chemistry in the natural sciences—their domains are different.

Islamic ethics is grounded in the *fiṭrah* (فطرة—natural disposition), which inclines the human soul toward truth, worship, and perfection. This ethical orientation, according to Islamic philosophers like Mutahhari and Mullā Ṣadrā (2003), is what distinguishes humans from animals—it is the capacity for moral discernment, not mere instinctual gratification. Like a seed, the soul contains inherent potential, which must be cultivated through ethical striving.[3]

[3] Mutahhari asserts "it is meaningless to talk of humanity without accepting the basic role of soul". The inner force to tend towards actualisation of one's potential is not physical, but external factors assist it. Hence the seed analogy—this is not like turning a plank into a chair as that is entirely dependent on external factors. If one does not actualise this humanity and perform their duty, their ontological state is thus considered 'dehumanised'.

Taha Abderrahmane is particularly adamant that *fiqh* and *akhlāq* are distinct yet ontologically linked. *Fiqh* addresses conformity to divine law in outward actions, whereas *akhlāq* addresses inward intentions and spiritual states. To reduce ethics to legal rulings is to cut off its vertical dimension—its connection to God and the unseen.

> ❝ The Sharīʿah is not reducible to legal prescriptions (*aḥkām*); it is a path toward ethical elevation (*tazkiyah*). ❞
> *(Paraphrased from Taha's writings in "Al-Amr bi-l-Maʿrūf" and "Rūḥ al-Ḥadātha")*

The *maqāṣid al-sharīʿah* (مقاصد الشريعة—objectives of Islamic law) were originally developed as a framework for use by jurists—experts in Islamic Law, who are in a position to understand the context and relevance of various factors in their decision-making—to expand legal reasoning in *fiqh*, not to function as a comprehensive ethical system. While useful, applying this framework to AI ethics presents significant limitations. Jurisprudence addresses the *what* and *how* of action; ethics addresses the *why*, focusing on cultivating right intentions, virtues, and moral character.

In Taha's philosophy, religion is nearly synonymous with ethics. It is no surprise, then, that he calls for a reformation of *fiqh* through ethical renewal. Wael Hallaq (2009) echoes this, arguing that the essence of Sharīʿah is ethics. However, Taha goes further, criticising how contemporary *maqāṣid* discourse often treats ethics as supplementary rather than foundational, as it places ethics as the 'enhancements' rather than the 'essentials' (Taha, 2022)—see Sect. 4.2.3.

He critiques the instrumentalisation of *maqāṣid* to serve state interests, utilitarianism, or political agendas. Many *maqāṣid* theorists reduce divine principles to human goals like welfare, development, or rights—resulting in a "*fiqh of outcomes*" rather than a *spirit of trust*. This misplacement prioritises expediency over divine command.

Taha writes that to treat *maqāṣid* as tools to negotiate modernity is to forget that they are rooted in amāna (divine trust), not human convenience. Ethics, for him, must be rooted in the *fiṭrah* and oriented toward the *ghayb* (unseen). When stripped of their metaphysical grounding, *maqāṣid* become secular rationalisations.

Taha's alternative is an ethics of entrustment (*iʿtimānī*) rooted in moral witnessing, humility, and divine alignment. True ethical action involves striving for *taʾalluh*—to mirror divine attributes such as mercy and justice. This vision necessitates reorganizing *maqāṣid*, making ethics central, not peripheral.

While efforts like Raquib et al. (2022) have attempted to integrate virtue ethics into *maqāṣid*, and Elmahjub (2023b) has proposed a *maqāṣid*-inspired AI ethics framework, both begin from a jurisprudential base. Raquib emphasises a non-Western Islamic worldview but ultimately remains within *fiqh*-oriented structures. These rule-based approaches constrain the moral imagination needed to address the evolving challenges of AI.

An instructive analogy is the delayed reaction to the harms of social media. It was not Big Tech but individuals, motivated by conscience and ethical concern, who

initiated resistance—eventually pressuring governments to act[4]. Technology ethics should largely be preventative not curative as by the time it reaches policy, the rapidly changing technology has already been released in public. ChatGPT was also released in controversial circumstances (Dixit, 2024).A valuable study from an Islamic perspective highlighted the importance of ethical virtues to tackle social media, yet again, suggested the framework Maqasid al-Shariah that does not actually encompass them (Zubair & Raquib, 2020). The design of AI should be informed on the whole by a philosophy that does not succumb to moral relativism or rule-based objectives.

Maqāṣid al-Sharīʿah faces critical challenges when applied to AI. It lacks the metaphysical depth and moral vision to offer comprehensive ethical guidance. As a legal theory, it can support jurisprudential rulings but cannot substitute for a philosophy. Taha Abderrahmane's *iʿtimānī* framework offers a more expansive alternative—grounded in the unseen, oriented toward trust, and animated by divine attributes.

What is needed is not merely a list of rules or principles, but an Islamic philosophical vision that resists moral relativism and reorients our relationship to technology around the metaphysical truth of human purpose and trusteeship. Only such a vision can guide the creation of technologies that uplift humanity, rather than degrade it.

3 Foundations of the I'timāni Framework

The trusteeship (I'timāni) framework, developed by Abdurrahman Taha (1994, 2000, 2001, 2005, 2006, 2020a, 2020b, 2022, 2023) (see also Hallaq (2019) and Hashas (2020) for a review of Taha's philosophy), offers a profound ethical vision rooted in the Qur'anic concept of amana (divine trust). It emphasises humanity's interconnected responsibilities toward God, knowledge, and creation, framing these obligations through three foundational covenants: the Ontological, Epistemological, and Existential[5]. Together, these covenants provide a comprehensive ethical foundation that integrates spiritual intentionality, intellectual integrity, and practical stewardship.

[4] Work by Smartphone Free Childhood (https://smartphonefreechildhood.co.uk/) has pushed for a complete ban in schools, a matter of focus now in UK Parliament: https://lordslibrary.parliament.uk/mobile-phones-in-schools-mandating-a-ban/

[5] Taha 'Abd al-Raḥmān never labels his pacts "ontological/epistemological/existential". Across several works he speaks of five (later six) mīthāq-s (covenants) using Qur'ānic language, the first three of which he uses when he proposes to reframe Maqasid al-Sharia through the lens of the Trusteeship Framework (2022 التأسيس الائتماني لعلم المقاصد). For readers trained in contemporary philosophy it is clearer to group these six pacts (subsequently numbered) under three broader headings: (a) Ontological – what grounds moral being – (1) Witnessing situates humanity before God; (4) Prophetic Distinction concretises that ground. (b) Epistemological – how truth is known and conveyed – (2) Entrustment regulates the acquisition and use of knowledge; (3) Mission secures its faithful transmission. (c) Existential – how moral agency unfolds in the world – (5) Stewardship obliges just governance of creation; (6) Mutual-Mercy extends that stewardship into compassionate solidarity. This triadic rubric keeps the Islamic substance of Taha's scheme while translating it into categories familiar to Western ethical discourse; it shows, in one phrase, how the covenants "integrate spiritual accountability, intellectual integrity, and practical stewardship."

3.1 The Central Concept of Divine Trust (Amana)

At the heart of the trusteeship (I'timāni) framework lies the Qur'anic concept of amana, which represents the sacred trust placed upon humanity to act as ethical stewards of creation. This trust is described in the Qur'an[6].

"We did indeed offer the Trust to the Heavens and the Earth and the Mountains; but they refused to undertake it, being afraid thereof: but man undertook it;- He was indeed unjust and foolish;" (Surah Al-Ahzab, 33:72).

This trust or privilege was granted to humans, signifying a special status over other creations that were unable to bear it. Yet when man did not recognise his own value or neglected his duties, he was 'unjust'. Amana has also been interpreted as perfection of worship by knowledge and deeds, or authority and free will that distinguishes humanity. By gaining knowledge and through self-purification, he is honoured above angels. This Divine trust is engagement and these privileges bring responsibilities for man, which is why it appeared burdensome. In fact, this highlights above all the special dignity of man who should not forget this honour. These responsibilities make man not only accountable to God but also to society, ensuring that ethical considerations transcend legal or material frameworks. Humans are trustees of creation, tasked with fostering social equity and justice, and safeguarding knowledge for the collective good. Their actions must align with the principles of justice (adl), compassion (rahma), and balance (mizan), reflecting sincere intentions (niyyah) aimed at upholding moral and spiritual integrity. Amana encapsulates accountability, intentionality and practical stewardship. It thus serves as the ethical nucleus of the trusteeship (I'timāni) framework, ensuring that all human endeavors are guided by moral and spiritual principles. Amana and I'timāni both share and are generated from the same Arabic root (أ م ن—A m n), as the Divine trust is central to the paradigm. Although this is a Quranic concept with common scholarly interpretations, Abdurrahman has used it as the basis on which he developed a robust theoretical ethical theory over the past four decades, considering man primarily as an ethical being and countering in his work postmodern thought and secularism that is incapable of taking care of humanity's needs as it considers only the biological and rational dimension of man. In this theory, the three covenants—Ontological, Epistemological, and Existential—form an integrated ethical framework that addresses the full spectrum of human responsibilities.

3.2 The Three Covenants as Ethical Pillars

The trusteeship (I'timāni) framework derives its ethical depth and structure from three foundational covenants. Each covenant reflects a distinct aspect of humanity's relational duties toward God, knowledge, and creation, ensuring human responsibilities are addressed in a comprehensive manner. This tri-partite structure is well-aligned with AI's multi-layered challenges: (a) the spiritual accountability

[6] Throughout the paper, all Quranic citations are presented not in their original Arabic text but as their corresponding translation of Abdullah Yusuf Ali.

dimension (Ontological), (b) the data- and knowledge-management dimension (Epistemological), and (c) the real-world impact dimension (Existential).

a. Ontological Covenant: Recognition of Divine Sovereignty

The Ontological Covenant is grounded in humanity's primordial acknowledgment of God's Lordship and establishes the basis for spiritual accountability. It grounds ethics in divine sovereignty, ensuring that *ultimate moral reference* is not purely human or secular. The Qur'an narrates this covenant in the following verse:

> "When thy Lord drew forth from the Children of Adam—from their loins—their descendants, and made them testify concerning themselves, (saying): "Am I not your Lord (who cherishes and sustains you)?"- They said: "Yea! We do testify!" (This), lest ye should say on the Day of Judgment: "Of this we were never mindful" (Surah Al-A'raf, 7:172).

By this testimony, the nature of man is established as an innate self-consciousness that is inclined towards recognising Divine Authority. This recognition based on the fitrah (or innate nature) of man means that seeking truth and the transcendent is part of everyone's natural disposition, and monotheism, that was confessed in the world of pre-existence, can be observed in the depth of each soul. By recognising God as the ultimate source of morality and guidance, ethical behaviour is not solely based on human reasoning but it is aligned to divine principles. This accountability fosters a sense of duty that transcends material or utilitarian concerns and is grounded in spiritual and moral obligations towards Divine will. Recognising God's sovereignty obliges individuals to act as ethical stewards (khalifa—خليفة) of creation, ensuring just governance over all aspects of life. Tied to a higher purpose enables one's actions to contribute positively to the well-being of society and the environment.

b. Epistemological Covenant: Pursuit of Ethical Knowledge

The Epistemological Covenant governs humanity's relationship with knowledge, emphasising its ethical acquisition and application. It emphasizes the ethical acquisition, dissemination, and application of knowledge that is, in the case of AI, crucial for data handling and algorithmic transparency. The Qur'an warns against distorting or misusing truth:

> "And cover not Truth with falsehood, nor conceal the Truth when ye know (what it is)." (Surah Al-Baqarah, 2:42)

Knowledge is regarded as a divine trust that must be utilised to promote justice (adl) and collective well-being. This guardianship of knowledge prohibits the use of knowledge as a tool for exploitation or deceit, ensuring that intellectual pursuits serve ethical purposes.The integration of revelation and reason is key. Human intellect (aql—عقل) must harmonize with divine revelation (wahy—وحي), ensuring that rational pursuits are informed by moral and spiritual values. This integration fosters a balanced approach to knowledge that respects both empirical evidence and spiritual wisdom thereby balancing the original and the analogic intellect. The dissemination and application of knowledge must prioritize truthfulness, equity, and accessibility. This commitment to truth and transparency ensures that information is shared

honestly and that ethical standards are maintained in the pursuit and distribution of knowledge and avoiding distortions for personal or collective gain.

The Epistemological Covenant establishes that intellectual endeavors are inherently moral acts, inseparable from their ethical consequences. By embedding ethical considerations into the acquisition and application of knowledge, this covenant ensures that intellectual pursuits contribute positively to society and uphold moral integrity.

c. Existential Covenant: Stewardship of Creation

The Existential Covenant defines humanity's responsibilities toward the natural world and society. It encompasses humanity's role as stewards of creation, including social equity and environmental care. Rooted in the Qur'anic principle of stewardship (khilafah—خلافة), this covenant emphasizes the balance and harmony of creation:

"He it is That has made you inheritors in the earth" (Surah Fatir, 35:39)

Humans must act as ethical caretakers of the environment, preserving its balance (mizan—ميزان) for future generations. This environmental stewardship involves sustainable practices that protect natural resources and mitigate environmental degradation. Stewardship extends to fostering justice (adl) and compassion (rahma—رحمة) within human societies, addressing inequalities, and ensuring that technological advancements benefit all members of society equitably. All aspects of creation deserve care and reverence, reflecting humanity's duty to act with responsibility and compassion. This respect fosters a harmonious relationship between humans and the natural world, ensuring that development does not come at the expense of ecological or social well-being.

The Existential Covenant ensures that human endeavors contribute to the flourishing of both society and the natural world. By balancing material progress with ethical stewardship, this covenant promotes sustainability, equity, and harmony, aligning human actions with the broader goals of creation.

3.3 A Unified Ethical Vision for AI

The covenants, as a basis for the ethical framework, address the means and motivations behind actions, ensuring that both processes and outcomes align with divine principles. These covenants address the limitations of existing models by emphasising intentions (niyyah), human dignity, and moral accountability. As a normative rather than descriptive ethical theory, Abdurrahman intended it to be actionable in the real world as he has critiqued modern theories that do not hold relevance in daily life. As a philosophy in the general sense of the word, it represents a world-view, and as an ethical framework, it guides human and technological actions.

The *Ontological Covenant* emphasises the innate moral disposition (fiṭrah) and dignity of humans, positioning them as the primary moral agents. This can be used to realise human-centric design. In AI ethics, this necessitates systems that prioritise human dignity and universal values. For instance, AI governance must ensure fairness, prevent algorithmic bias, and protect privacy, aligning with the divine mandate

of upholding justice. The *Epistemological Covenant* also shifts focus from outcomes to the means, embedding ethical processes in AI system design and governance. It emphasises intentions (niyyah) in actions. This ensures developers and users align their work with moral objectives. AI development and deployment should not only aim for beneficial outcomes but also ensure that the processes are morally sound. Lastly, the *Existential Covenant* frames humans as stewards accountable to God. In the context of AI, this entails ensuring transparency, fairness, and alignment with spiritual and ethical principles. Developers and policymakers must act as trustees, fostering technologies that promote societal harmony and ecological sustainability.

Thus, these covenants provide a foundation to expand the ethical scope beyond material and rational benefits to include spiritual well-being. The responsible management of AI will promote long-term well-being and environmental preservation, by spiritual accountability, intellectual integrity and practical stewardship. This approach challenges teleological frameworks like Maqasid and Floridi's Levels of Abstraction (LoA) by shifting the focus to the means and ethical motivations.

This unified vision ensures that ethical behavior reflects both spiritual intentionality and practical responsibility. By situating human actions within a divine cosmology, the trusteeship (I'timāni) framework provides a comprehensive model for addressing ethical challenges across spiritual, intellectual, and environmental domains.

As AI expands, it is branching out further into academic subdisciplines—'critical AI' focuses on highlighting the risks and dangers of AI, 'responsible AI' emphasises technical solutions, and 'decolonial' or 'indigenous AI' focus on representation and social justice. With the known issue of 'ethics-washing', AI ethics has been criticised for being useless as the underlying problems lie in the corporate environment itself and corporations escape real regulation (Munn, 2023). It is perhaps only an overarching philosophy that can tackle these complex issues. As Abdurrahman's theory sees ethic as the basis of religion, his philosophy has been a popular response to modernity in a technology-driven world—but does it hinder progress? In 'human-centred AI', AI is designed in a way that prioritises and supports human needs. The trusteeship paradigm, by emphasising the innate dignity of humans conferred by God, provides a framework to work in the technological sphere with responsibility, respect and care.

In today's 'attention economy', for Artificial Intelligence to work for humans and not the other way round, the ethical decision-making lies with the computer programmer (Blackwell, 2024). As Alan Blackwell alleges in Moral Codes, the speculative questions on AGI posed by the media circus shift this accountability and, much like the original Turing question[7], are philosophical in nature rather than science fiction fantasies to emulate. Are AI engineers trained to answer these moral quandaries? How far removed are the AI ethicists in driving the actual outputs? If the ultimate decision-making lies with the AI engineers, are they qualified to make philosophical and moral judgements that can impact humanity on such scale? The

[7] The Turing Test was initially presented because it was considered meaningless to ask if machines can think (Turing 1950).

trusteeship framework ensures that accountability is not shifted or erased, and can guide the design of responsible AI. It has the capacity, as a philosophical-based paradigm, to explore pertinent issues. Merely naming 'accountability' as one of the principles to aim for, is not sufficient, as its actuality requires a philosophical structure and worldview to encompass and encourage these values on an individual as well as collective level.

4 Review of Current Approaches

The trusteeship (I'timāni) framework, with its metaphysical grounding and emphasis on spiritual accountability, offers a critical lens for evaluating existing ethical paradigms. While existing ethical frameworks based on Maqasid al-Shariah and Floridi's Information Ethics (IE) provide valuable insights, they exhibit limitations in addressing the ethical complexities of emerging technologies.

4.1 Maqasid al-Shariah for AI

A common starting point in recent literature for discussing Islamic ethics for AI has involved the adoption of Maqasid al-Shariah (مقاصد الشريعة—Objectives of Shariah) alongside jurisprudential rules (legal maxims). Maqasid al-Shariah is defined as "a comprehensive issue that expresses the will of the legislator regarding the legislation of rulings, extrapolated from Shariah evidence to guide actions and decisions." Jurisprudential rules, by contrast, are practical, encompassing various detailed issues and applications of Shariah in daily life. While jurisprudential rules focus on individual accountability and actions, Maqasid rules provide broader guidance based on Shariah's overarching goals (Kamali, 2008; El-Mesawi, 2020; Zaim, 2023). In the context of AI, we discuss Maqasid rules due to their conceptual and theoretical relevance.

Maqasid al-Shariah, representing the understanding and purposes of Islamic rulings, centers on the preservation of five core objectives—faith (*din*—دين), life (*nafs*—نفس), intellect (*aql*—عقل), lineage (*nasl*—نسل), and wealth (*mal*—مال). These objectives can be broken down into essentials (daruriyyat—ضروريات), needs (hajiyyat—حاجيات), and enhancements (tahsiniyyat—تحسينيات), forming a hierarchy of priorities. For example, within the AI domain, Maqasid principles provide ethical guidelines to navigate novel challenges:

1. The Rule of Means and Ends: This rule states, *"The means take the same ruling as the objectives."* If the objective aligns with Shariah, the means to achieve it are permissible. For example, using intelligent systems like medical AI to save lives aligns with the preservation of life, an essential of Maqasid. Conversely, designing robots for idol worship directly contradicts the principle of Tawhid and is impermissible.
2. Flexibility of Means but Fixed Ends: This allows flexibility in methods to achieve objectives, provided they do not compromise fundamental values. For instance,

AI for crime prevention may permit the use of surveillance technologies if used responsibly, balancing public welfare with privacy concerns.

Suggestions of Islamic frameworks for evaluating AI ethics have mostly centered on Maqasid al-Shariah with various contemporary interpretations (Elmahjub, 2023a, 2023b; Mohadi & Tarshany, 2023; Raquib et al., 2022; Nawi et al., 2022, etc.). This is a teleological approach, prioritising the underlying aims of Shariah in evaluating modern challenges, including AI technologies. This methodology is valuable in centring human life above machines as altogether the objectives focus on the preservation of humanity.

Nuanced frameworks for integrating Maqasid into AI ethics have focused on various aspects. Elmahjub (Elmahjub, 2023a, 2023b) focuses on maslaha (مصلحة—public welfare) as it can be interpreted from both utility- and duty-based views (or consequentialism and deontologism respectively). When Al-Ghazali adopted al-Shatibi's five objectives in the eleventh century, he prioritised individual rights and values over welfare calculations. According to al-Ghazali, moral responsibility and human dignity should be protected even if collective welfare is not maximised. Elmahjub in his modern interpretation however proposes a nuanced hybrid interpretation of maslaha to be optimised. It is versatile and easily adaptable to any situation—being open to interpretation—as long as the Shariah objectives are not undermined. For instance, enhancing welfare through AI must not compromise essential Maqasid values like intellect or dignity.

Raquib and colleagues (Raquib et al., 2022) introduce an Islamic virtue ethics model with the overarching goal of Tazkiya (تزكية—self-purification), using Maqasid to shape ethical AI development as a holistic process integrating individual and societal dimensions. Again, this is a modern interpretation, as Islamic virtue theory is not traditionally part of Maqasid. She defines *maslaha* as benefit or human good, and considers benefits of both individual and society with priority of the latter. The holistic approach tackles structures of inequality thus: "Under this system, ethics is given a greater central stage in the individual's private life whereby an individual's ethical choices shape the society they live in; and in return, the society shapes the individual's ethical choices. This continuous feedback mechanism ensures that the private individual is intimately tied to the social whereby the individual becomes the social." (Raquib et al., 2022).

Another study considers the protection of the *'aql* (intellect) as the relevant principle in the Maqasid-al-Shariah for AI ethics (Abdul Rahman & Ibrahim, 2021). The authors use examples of alcohol and drugs that are prohibited as they affect the reasoning capacity of humans. The *'aql* is the contemplative faculty that connects human to God. Therefore, if using AI will degrade the normal function of human minds, then it should be avoided. Emphasising the 'aql also helps protect intellectual property, a serious conflict for generative AI that has repeatedly undermined these rights. On the practical implementation of Maqasid al-Shariah that is lacking in most other studies, this work considered how these principles can be used to impose laws and guide policy for better regulation. Using AI particularly for religious information can cause "shallow thinking (al-ummiyah al-fikriyah—الأمية الفكرية)"

(Rachmawati et al., 2024), so the focus on intellect is vital particularly for the Islamic world.

The variants of these frameworks are useful in offering principles of public welfare, preserving fundamental values and aligning with Divine will. They can help structure ethical AI development and deployment, but there are some shortcomings that have not been addressed well.

4.2 Limitations of Maqasid al-Shariah

Maqasid al-Shariah was initially proposed as a hermeneutical tool of reasoning in legal theory by al-Ghazali and al-Shatibi (Hallaq, 1997). However, it is not unifying across different schools of thought in Islam, especially as it is derived not only from primary sources (Quran and AHadith [traditions of Prophet Muhammad p.b.u.h.]) but secondary sources in Islamic epistemology. Scholars have concluded different versions of the fundamental Maqasid. Contemporary interpretations for wider applicability have evaluated the Maqasid in terms of community-level benefits rather than individual values that it originally prioritised. Furthermore, there is no unanimous agreement on the traditionally cited five core objectives. Kamali (2008), for instance, argues for an expanded list comprising eight categories.

As discussed in Sect. 2, the first critique is the Jurisprudence vs. Ethics debate. Our approach is universal and does not depend on Islamic Fiqh schools.

4.2.1 Law and Ethics

As established in Sect. 2, Islamic intellectual tradition maintains a fundamental distinction between *fiqh* (jurisprudence) and *akhlāq* (ethics), each operating within its own epistemological and ontological domain. *Fiqh* addresses the external dimension of action—what is permitted or prohibited—while *akhlāq* concerns the inner disposition, intention (*niyyah*), and spiritual refinement of the moral agent. Though interrelated, they are not interchangeable: *fiqh* is normative and prescriptive; *akhlāq* is aspirational and transformational.

This distinction becomes critical when considering the application of *maqāṣid al-sharīʿah*—a framework originally developed to broaden legal reasoning within *fiqh*—to the ethical challenges of artificial intelligence. Because *maqāṣid* was never intended as a comprehensive system of ethics, but rather as a juristic methodology to derive rulings based on overarching objectives of Islamic law, its scope is limited when it comes to addressing AI's complex moral and metaphysical implications.

Taha Abdurrahman critiques the overextension of *maqāṣid* into domains for which it was not designed. He argues that conflating legal reasoning with ethical reasoning reduces morality to rule-following and undermines the vertical dimension of ethical life—that is, the direct relationship between the human being and the Divine. As he asserts in *Rūḥ al-Ḥadātha* and *al-Amr bi-l-Maʿrūf*, true Islamic ethics is grounded not in outcomes or utility, but in *amāna* (trusteeship), *taʾalluh* (divine alignment), and spiritual intentionality.

Moreover, Abdurrahman's critique goes beyond methodology to question the ontological assumptions of *maqāṣid*-based frameworks. When ethics is treated as merely a *taḥsīnī* (complementary) aim rather than a *ḍarūrī* (essential) foundation, it

reflects a worldview in which moral values are subordinated to legal form. In contrast, his *iʾtimānī* paradigm elevates ethics as the essence of religion itself, calling for a reorientation in which moral philosophy becomes the organizing principle of legal and technological reasoning alike.

As shown in Sect. 2, the revival of *akhlāq* requires grounding ethics in *fiṭrah* (natural disposition) and *ghayb* (the unseen), emphasising the cultivation of virtues that cannot be captured by legalistic categories alone. While *maqāṣid*-based frameworks like those proposed by Raquib and Elmahjub attempt to bridge ethics and law—often by integrating virtue ethics into *maqāṣid*—they remain tethered to jurisprudential scaffolding that limits their philosophical scope.

Therefore, while *maqāṣid al-sharīʿah* can serve as a useful tool within a legal regulatory apparatus, it cannot fulfill the deeper role of ethical philosophy. To address the moral challenges posed by AI—challenges involving intention, responsibility, and human purpose—a rule-based framework is insufficient. What is needed is a comprehensive ethical vision, such as that provided by Abdurrahman's trusteeship (*iʾtimānī*) framework, which places the human being not merely as a legal subject, but as a moral trustee accountable before God, society, and creation.

4.2.2 Reduction of Divine Wisdom

Maqasid al-Shariah often employs rationalist methodologies that reduce the divine wisdom of Sharia to utilitarian or material objectives. Most of the aims that are centred are not related to the moral philosophy of religion but rather the secular concept of ethical objectives, missing the inner meanings of the worldview espoused by Islamic revelation. This approach, while pragmatic, risks prioritizing tangible benefits over transcendent spiritual goals, thereby marginalizing deeper spiritual accountability.

For instance, prioritizing economic growth (*mal*) through AI-driven financial systems might enable exploitative practices, such as biased credit scoring, which undermine human dignity (*karamah*—كرامة). Similarly, focusing solely on preserving life (*nafs*) in healthcare might lead to privacy violations, as data is commodified and patients' moral concerns are neglected. The Ontological Covenant in the I'timāni framework restores divine sovereignty to the center of ethical decision-making. This ensures that material outcomes are secondary to alignment with divine principles, fostering a sense of accountability that transcends secular utilitarian concerns.

4.2.3 Teleological Focus on Outcomes

The Maqāṣid framework's teleological orientation—which evaluates actions based on their intended outcomes—carries the risk of neglecting the ethical integrity of the means and the intentionality (niyyah—نية) underlying those actions. When disproportionate emphasis is placed on external outcomes, there is a danger of morally ambiguous means being justified for the sake of desirable ends. This issue becomes particularly acute when Maqāṣid reasoning is detached from the akhlaqī (ethical) dimension that traditionally integrates the form, process, and metaphysical character of the action.

For instance, an AI system designed to enhance educational performance (*ḥifẓ al-ʿaql*) might employ intrusive surveillance technologies, thereby violating students' privacy and infringing on principles of justice (*ʿadl*). While the ostensible outcome appears beneficial, the processes employed contradict core ethical values rooted in the Islamic conception of human dignity and trust.

In contrast, the Iʾtimānī Framework, drawing on Taha Abdurrahman's philosophy, places ethical intentionality at the heart of moral evaluation. Through the Epistemological Covenant (al-ʿAhd al-Maʿrifī), it insists that both means and ends must align with the higher values of trust (*amāna*), justice (*ʿadl*), and spiritual accountability. This approach firmly rejects any reasoning that would allow noble outcomes to justify ethically problematic means.

For clarity, we must emphasize that this paper does not claim that the Maqāṣid framework—whether classical or contemporary—explicitly endorses the maxim *"the end justifies the means."* Rather, our critique is directed at certain instrumentalist tendencies in modern applications of Maqāṣid reasoning, particularly when ethical intentionality and spiritual accountability are sidelined in favor of technocratic or utilitarian outcomes. This concern is precisely the focus of Taha Abdurrahman's critique of contemporary Maqāṣid discourses.

4.2.4 Static Categorization and Contextual Insensitivity

Contemporary applications of Maqāṣid al-Sharīʿah often treat the classical five objectives—preservation of religion, life, intellect, lineage, and property—as a fixed and exhaustive list, framing them as universal categories sufficient for all ethical challenges. However, this static approach struggles to adequately address emergent ethical issues posed by modern technologies, including algorithmic bias, environmental degradation, and the deep societal impacts of AI on human agency and identity.

This rigidity hampers the integration of broader concerns such as sustainability, social justice, and cultural diversity into ethical reasoning. For example, AI systems designed to enhance knowledge (*ḥifẓ al-ʿaql*) may rely on energy-intensive infrastructures, contributing to climate change and thereby contradicting Islamic principles of stewardship (khilāfah). Similarly, algorithmic biases that reinforce social inequities highlight a gap in how Maqāṣid frameworks have traditionally conceptualized justice and human dignity.

In response to these challenges, the Iʾtimānī Trusteeship Framework, inspired by Taha Abdurrahman's philosophy, introduces the concept of the Existential Covenant. This covenantal perspective offers a dynamic and context-sensitive ethical approach, rooted in the metaphysical principle of amāna (trust). Unlike static categorizations, it integrates environmental care, social equity, and cultural responsibility as essential dimensions of ethical deliberation, ensuring that Islamic ethics remain responsive to the complexities of technological modernity.

While Maqāṣid al-Sharīʿah remains a valuable framework for practical decision-making within Islamic jurisprudence, its historical emphasis on legalistic reasoning and outcome-oriented assessments may fall short in confronting the ontological and

ethical disruptions of AI technologies. The Iʾtimānī Framework addresses this gap by unifying:

- Metaphysical grounding (tawḥīd and trust),
- Ethical intentionality (niyyah), and
- A dynamic understanding of stewardship (khilāfah and amāna).

This holistic approach ensures that technological developments are evaluated not only for their material utility, but also for their alignment with humanity's higher moral and spiritual purposes. By re-centering the human being as a morally responsible trustee (muʾtaman), the Iʾtimānī vision offers an ethical paradigm capable of confronting the unique challenges posed by AI and other emerging technologies.

Ultimately, it affirms that metaphysics, ontology, and ethics are inseparable. This integrated perspective is essential for any meaningful Islamic engagement with the ethical dilemmas of our technological age.

4.3 Ontological-Ethical Faultlines of AI: An Iʾtimānī Perspective

Mainstream AI ethics tends to frame its concerns around procedural issues: data privacy, algorithmic bias, environmental sustainability, and fairness. While these are valid and important, they operate largely within a technocratic and utilitarian paradigm, often detached from deeper metaphysical questions. From the vantage point of Islamic ethics in the iʾtimānī sense, however, such framing is insufficient. The real ethical crisis of AI lies not in its outputs alone, but in how it reconfigures human subjectivity, agency, and moral intentionality.

4.3.1 Displacement of Moral Intentionality (Niyyah)

At the heart of Islamic ethics is the principle of niyyah—the inner intention that orients an act towards its ethical and spiritual telos. Taha Abdurrahman emphasizes that ethical action cannot be reduced to external outcomes, as it is the intentionality that grants the act its moral worth (*Dīn al-Ḥayāʾ*, vol. 1, pp. 120–125). However, AI systems operate by decoupling actions from intentions, replacing human deliberation with algorithmic processes driven by statistical optimization.

In contexts like automated decision-making, the locus of moral intention is outsourced to systems that have no capacity for iʾtimān (trusteeship) or moral responsibility. This displacement represents an ethical rupture: it severs the link between human intentionality and the moral evaluation of action. From an iʾtimānī standpoint, this is not a mere technical issue, but an ontological corruption of human moral agency.

4.3.2 Erosion of Taklīf and Delegated Agency

Islamic ethics rests on the concept of taklīf—the moral responsibility conferred upon human beings by God. AI technologies, particularly in their autonomous or

semi-autonomous forms, blur the boundaries of agency. When humans rely on AI outputs to make decisions (in healthcare, finance, education), they engage in a subtle form of delegated agency that dilutes personal accountability.

As Taha warns in his critique of *ʿaql ālī* (instrumentalized reason), such delegation leads to a form of moral outsourcing, where humans abdicate their ethical responsibility to machinic processes designed without reference to divine trusteeship. The outcome is not just ethical laziness, but a progressive erosion of the human capacity for moral discernment (iḥsān).

4.3.3 Feedback Loops and Ethical Subjectivity

Another critical dimension is the reflexivity of AI systems: they do not merely process inputs, but actively shape human perceptions, desires, and behaviors through feedback loops (recommendation systems, predictive profiling). This recursive shaping of human subjectivity is an ontological intervention that subtly reconfigures what it means to act, to choose, and to be responsible.

From an iʾtimānī lens, this is a violation of the Epistemological Covenant (al-ʿAhd al-Maʿrifī), wherein knowledge is not neutral data, but a moral trust whose misapplication entails ethical breach. Taha would argue that AI systems, by modulating human ethical horizons through opaque algorithms, are agents of epistemic injustice—contributing to the moral deformation of the subject.

4.3.4 Beyond Procedural Ethics: Re-Centering the Muʾtaman

In sum, an iʾtimānī critique of AI ethics demands a shift from procedural governance to the re-centering of the moral subject (al-mukallaf, al-muʾtaman). Ethical evaluation must not stop at outputs (bias, fairness), but interrogate how AI reshapes the very capacities that define human moral agency: intentionality, responsibility, and spiritual accountability.

AI, in this light, is not merely a technological challenge but a spiritual-ontological crisis. Its ethical evaluation must therefore be grounded in a metaphysics of trusteeship, where every technological intervention is assessed by its impact on the integrity of human subjectivity and the fulfillment of divine trust (amāna).

4.4 Floridi's Information Ethics

In Information Ethics (IE) (Floridi, 2010, 2011, 2013, 2014a, 2014b, 2023), Luciano Floridi establishes an innovative digital ethics framework by considering the moral worth of all informational entities in the "infosphere," extending ethical consideration beyond human agents.[8] This approach redefines the boundaries of ethical discourse to include not only traditional biological entities but also digital artifacts,

[8] LuFlot Bot, a generative AI chatbot trained on Floridi's philosophy of information, answers questions on the ethics of digital technology: https://news.yale.edu/2024/04/16/student-developed-ai-chatbot-opens-yale-philosophers-works-all

artificial agents, and data structures, treating all these entities as having intrinsic moral value.

For AI systems to operate within ethically responsible bounds, with transparency, fairness, and accountability, Floridi advocates for what he calls an "info-ethical" approach, where information is a crucial element for defining moral agency and responsibility, not only for humans but also potentially for autonomous systems such as AI. Floridi utilizes the concept of Levels of Abstraction (LoA) as a key ontological tool in his framework, enabling the systematic categorization and analysis of these diverse informational entities at various levels of complexity and detail. This allows for a nuanced and scalable approach to ethics that can effectively address the ethical status and implications of entities ranging from simple data points to complex AI systems.

Furthermore, Floridi introduces the concept of "re-ontologizing the infosphere,"[9] arguing that digital technologies, particularly AI, are transforming the nature of our reality by creating new environments and forms of agency. This re-ontologization process fundamentally alters the intrinsic nature of systems and entities within the digital realm, thereby reshaping our understanding of agency, identity, and the environments we interact with. This ontological expansion is pivotal for accommodating the ethical complexities introduced by advanced technologies, ensuring that ethical discourse evolves in tandem with technological progress and reflects the profound changes these technologies bring to our conceptual and existential frameworks. He thus applies ethical reasoning in a systemic manner across different layers of AI applications, rather than restricting it to isolated cases. This encompasses individual interactions between AI systems and humans, broader societal impacts, and environmental considerations. Floridi promotes this as a step toward a more "deontological" approach to AI ethics, ensuring AI operates under principles of duty and moral responsibility. Although Easterns traditions share these values of responsibility in decision-making, in Floridi's theory they come at the expense of human dignity as non-living entities have equal rights and responsibilities. The re-ontologizing of the

[9] As an example, Floridi's ontology implies "artificial information, while distinct from human beings, is not necessarily of a lower ontological status"—this is what we get from https://luflot-bot.vercel.app/bot

"Luciano Floridi's work on the ontological status of artificial information suggests that it occupies a unique position within the broader framework of being. He argues that all entities, including artificial informational entities, share a common ontological status as informational entities. This perspective challenges traditional hierarchies that place human beings at the top of the ontological ladder.

Floridi posits that "all things—i.e., all informational entities—have some intrinsic value and are subject to some moral respect"[Book 4, Page 19]. This implies that artificial information, while distinct from human beings, is not necessarily of a lower ontological status. Instead, it is part of a continuum of being that includes both human and non-human entities. The principle of ontic equality suggests that all entities, regardless of their nature, deserve a minimal level of moral consideration[Book 4, Page 19][Book 4, Page 6].

However, this does not mean that artificial information is "higher" than human beings; rather, it indicates a shift in how we understand value and moral respect across different types of entities. Floridi's approach encourages a more inclusive ethical framework that recognizes the intrinsic worth of all informational entities, thus promoting a more egalitarian view of existence[Book 4, Page 19][Book 4, Page 6].

Source(s): [Book 4, Page 19]: The Ethics of Information—"IE argues that all things—i.e. all informational entities—have some intrinsic value and are subject to some moral respect." [Book 4, Page 6]: The Ethics of Information—"IE strives to provide a good unbiased platform from which to educate… the inhabitants of the infosphere.".

infosphere does not resonate well with Islamic ontology, where the ontological status of human beings is greater than that of the instruments they create.

Floridi's IE faces challenges from a metaphysical and cultural perspective, particularly when contrasted with frameworks like the I'timāni model, which embeds spiritual accountability into ethical reasoning by raising and not undermining the status of humans. Although Floridi's framework is comprehensive in its secular and systematic approach, it may not fully resonate with or incorporate the spiritual dimensions and metaphysical values that are central to other cultural and ethical systems. Its secular foundation treats moral worth as derived from systemic functionality rather than transcendental principles.

Floridi emphasises a universalist perspective grounded in five ethical principles—beneficence, nonmaleficence, autonomy, justice, and explicability. These principles are explicitly designed to align with a global vision for AI ethics that transcends jurisdictional boundaries and prioritize environmental sustainability, human dignity, and systemic accountability. This universalism has key strengths for the environmental sustainability goals as it highlights the dual responsibility AI poses to ecosystems and social systems alike. It also has procedural clarity—the emphasis on explicability and systemic accountability provides a robust mechanism for promoting trust in AI systems. Floridi's procedural neutrality risks reducing ethical frameworks to mechanical standards.

Floridi critiques ethical governance practices, such as "ethics bluewashing" and "ethics dumping," where organizations use superficial ethics as a facade. However, his framework may lack the cultural sensitivity needed to resonate universally. For example, IE's universalist principles may clash with regional and cultural norms that prioritize collective well-being over individual autonomy. This may therefore face challenges in implementation.

While IE sees informational ecosystems as ethically significant, the trusteeship (I'timāni) framework integrates divine trust (amana) as central, affirming that technology's alignment with justice and stewardship is inseparable from metaphysical accountability. The I'timāni model unites moral and spiritual dimensions with ethical governance.

4.5 4.5 Summary: Toward a Holistic Ethical Vision

While both Maqasid al-Shariah and Floridi's Information Ethics provide valuable insights, they exhibit significant limitations when addressing the ethical complexities of modern technologies. The focus of Maqasid al-Shariah on outcomes and static objectives limits its adaptability to novel challenges. The trusteeship (I'timāni) framework addresses these gaps through its emphasis on spiritual intentionality, dynamic stewardship, and holistic responsibility. The secular foundation and procedural neutrality in Floridi's Information Ethics fail to account for spiritual accountability and cultural diversity. The trusteeship (I'timāni) framework bridges these gaps by integrating divine trust (*amana*) and the three covenants into a unified ethical model.

By combining the strengths of these frameworks and addressing their limitations, the I'timāni perspective offers a holistic ethical vision. This vision aligns spiritual principles with practical governance, ensuring that emerging technologies, including AI, serve humanity's higher moral and spiritual goals. In particular, the trusteeship (I'timāni) framework emphasizes the metaphysical distinction between humans and AI, reaffirming humanity's unique role as stewards of creation while situating AI as a tool to fulfill ethical and spiritual objectives.

5 Integration with the Wider World

In this section, we demonstrate how the trusteeship (I'timani) framework as a holistic framework, derived from Quranic principles that speak to humanity, has practical applicability and the potential to be augmented with the wider world.

5.1 Applications

This section illustrates how we are going to conceptualize and plan to interrogate the applications of the trusteeship (I'timāni) framework to the ethical challenges posed by AI. By employing its three foundational covenants—Ontological, Epistemological, and Existential—, as shown in Fig. 1, the framework provides actionable guidance for aligning AI development and deployment with justice, sustainability, and spiritual accountability.

For example, Fig. 2 illustrates the unfolding of the Trusteeship framework from the three covenants to practical guidelines in areas like Healthcare, Finance and Surveillance. We leave other applications like Warfare and Education to future work.

5.2 Conjunction with Regulations: Addressing Ethical Threats Using Sadd al-Dharai (Blocking Harmful Means)

Legal prohibitions or restrictions, whether secular or based on religion, play a crucial role in demarcating the limits within which one can develop and use a new technology. How can the I'timini framework be used in conjunction with regulations? The work of both Elmahjub (2023a, 2023b) and Raquib (2022) employed this as a way to halt and block the means of unethical ends, if any of the Maqasid al Shariah were undermined. Here, we show practically how a philosophical approach like the trusteeship paradigm is not lacking in its ability to be integrated with laws and regulation.

While the trusteeship (I'timāni) framework provides a positive framework for ethical AI governance—focusing on what should be done to align AI with justice (*adl*), stewardship (*khilafah*), and spiritual accountability—the principle of *sadd al-dharai* (سد الذرائع) augments it by defining *what must not be done*. Rooted in Islamic jurisprudence, *sadd al-dharai* emphasizes the proactive prevention of harm by prohibiting practices that could lead to unethical outcomes. This dual approach ensures

a balanced and comprehensive strategy for addressing both the opportunities and risks of AI technologies.

For example, manipulative algorithms pose a rising ethical threat by undermining individual autonomy, spreading misinformation and weakening social cohesion. Applying the three covenants of the trusteeship (I'timāni) framework, human responsibilities can be structured in the ontological, epistemological and existential dimensions. *Sadd al-Dharai* integrates seamlessly by ensuring that these covenants are not breached through harmful practices. The practical measures as a result of this augmentation include prohibiting exploitative practices that target vulnerable populations and create divisions through misinformation. The algorithms should be transparent in design, promote truthful content and societal well-being. This aligns with the principles of justice (adl) and spiritual accountability. This and other examples, shown in Table 1, apply *sadd al-dharai* to key ethical challenges in AI, highlighting actions that must be avoided to safeguard human dignity, equity, and sustainability. One of the main challenges of AI technologies is the loss of accountability and threat of shifting blame within a changing legal landscape. The trusteeship (I'timāni) framework recenters the human, ensuring there is no risk of exploitative powers shifting responsibility of consequences to digital entities. This is essential for the future safeguarding of this rapidly expanding technology.

5.3 Recommendations for Building Ethical AI Communities

The rapid proliferation of Artificial Intelligence (AI) technologies requires governance models and educational initiatives that reflect inclusivity and cross-cultural collaboration. The trusteeship (I'timāni) framework, with its metaphysical grounding and universal principles, provides a comprehensive foundation for achieving justice, sustainability, and spiritual accountability in AI development and governance.

AI technologies transcend national borders, influencing societies with diverse cultural and spiritual values. However, current governance models often prioritize Western secular paradigms, neglecting alternative ethical perspectives. A cross-cultural approach grounded in the trusteeship (I'timāni) framework ensures inclusivity and global relevance. Interdisciplinary international forums that bring together Islamic scholars, ethicists, technologists, and policymakers can foster dialogue and shared understanding of ethical AI challenges and solutions. The framework can also be integrated with ethical systems of other Abrahamic religions or ideologies, to form ethical alliances that can advocate for universal principles such as justice (adl), sustainability (istidama—إستدامة), and accountability (hisab—حساب). Regional research hubs can be established to explore and develop AI applications tailored to local contexts while adhering to global ethical standards. For example, an alliance addressing AI's environmental impact might draw on Islamic principles of stewardship (khilafah) and the common good (maslahah). The emphasis of the existential covenant on sustainability can encourage nations to adopt green AI practices while respecting cultural sensitivities.

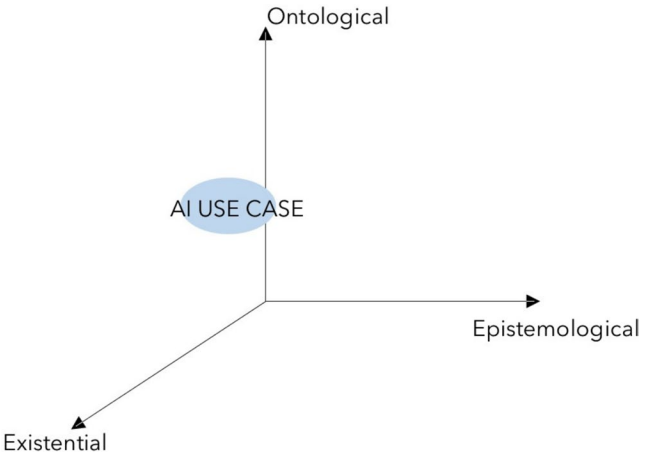

Fig. 1 Ontological, Epistemological and Existential Covenants as dimensions of the framework for a specific AI use case

For ethical AI education, ethical models such as the trusteeship (I'timāni) framework that have global appeal and practical case studies demonstrating how they can address real-world AI challenges, can be integrated into curricula such as interdisciplinary university courses. AI systems can be assessed to evaluate whether the algorithms uphold dignity (karamah), equity (adl), and sustainability (mizan). Public awareness campaigns should be launched to educate the general public about AI's ethical dimensions, such as data privacy, algorithmic bias, and environmental sustainability. Culturally resonant narratives can be used to connect ethical principles to local contexts and enhance community engagement. Most importantly, in order to incorporate the trusteeship (I'timāni) framework into AI design principles, developers should be trained to balance innovation with ethical responsibility. Professional workshops and certifications can teach ethical coding practices aligned with these principles.

There is also a need for universally accepted global standards that address technical, moral, and cultural dimensions. The trusteeship (I'timāni) framework's integration of divine trust (amana) and the three covenants provides a robust foundation for creating inclusive and actionable guidelines. An international code of ethics incorporating the principles of the trusteeship (I'timāni) framework may ensure spiritual and cultural inclusivity in AI governance. These standards can be advocated for in global forums. AI governance standards can be aligned with the Existential Covenant, mandating environmental sustainability by encouraging green computing practices and prioritising underserved communities in AI development. Independent oversight bodies can be established to monitor compliance with ethical standards, along with transparent reporting systems for addressing ethical violations in AI deployment. This will help maintain accountability mechanisms, particularly

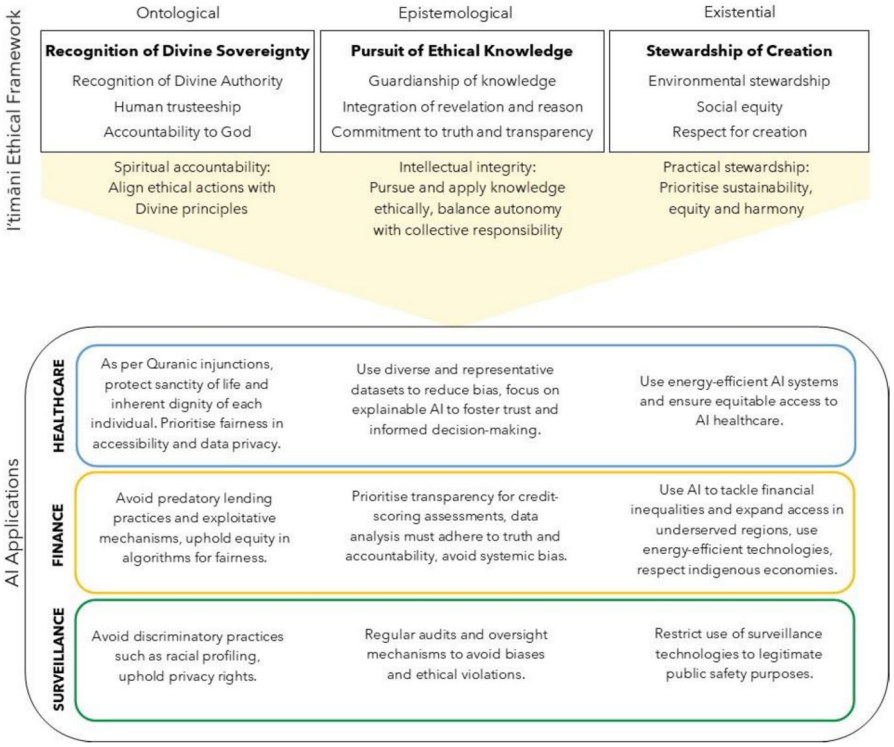

Fig. 2 Practical application of the trusteeship (I'timāni) framework for AI. The comprehensive trusteeship (I'timāni) framework rests on three covenants that designate responsibilities towards God, knowledge and creation. These unfold top to bottom. The implications of these trusts are ethical intentionalities that yield imperatives, practical guidelines and implementation strategies for artificial intelligence applications in various use cases

for large corporations. Forming such cross-cultural collaborations fosters trust and cooperation among nations.

Grassroots involvement is essential for ensuring that AI technologies reflect local values and address community needs. The trusteeship (I'timāni) framework's emphasis on collective responsibility and stewardship aligns naturally with community-driven initiatives. Community projects that address local challenges, such as improving healthcare access or mitigating harm to local farmers, should be supported. Ethical principles that reflect local values should be integrated into AI design and deployment. Advocating for empowering marginalised populations would assess whether AI technologies are addressing their needs and concerns while upholding their dignity (karamah). They can also become involved in participatory design processes to include underrepresented voices in AI development. Developing open-source AI tools based on the trusteeship (I'timāni) framework would make ethical technologies accessible to diverse communities.

The trusteeship (I'timāni) framework offers a comprehensive model for cross-cultural AI governance and ethical education by ensuring that AI technologies align with principles of justice, sustainability, and accountability. By integrating these recommendations into governance and education systems, the trusteeship (I'timāni) framework paves the way for an ethical AI future that respects human dignity on an individual level and advances collective well-being across communities.

5.4 The role of AI ethicists: a dialogue with Floridi's perspective

The ethical challenges surrounding AI development demand a profound examination of the moral responsibilities and capacities required of AI ethicists. Luciano Floridi and collaborators provide a framework for understanding the role of AI ethicists in "What is an AI Ethicist? An Empirical Study of Expertise, Skills, and Profiles to Build a Competency Framework" (Floridi, 2019), emphasising their multidimensional expertise in moral reasoning, facilitation, research, and education. Integrating these insights with the trusteeship (I'timāni) framework reveals complementary strengths and mutual contributions to shaping AI governance.

Floridi outlines AI ethicists as possessing three overlapping types of moral expertise: academic, performative, and practical. This aligns with the trusteeship (I'timāni) framework 's covenants of Ontological, Epistemological, and Existential responsibility, suggesting a shared emphasis on grounding ethics in both theoretical rigor and practical application.

1. Academic Moral Expertise: AI ethicists are expected to delve deeply into normative ethical theories, akin to the trusteeship (I'timāni) framework 's metaphysical grounding in the Ontological Covenant. For example, while the trusteeship (I'timāni) framework draws on divine sovereignty to inform ethical stewardship, Floridi's perspective integrates broader philosophical paradigms, including soft ethics and principles-based reasoning, to align technology with human values.
2. Performative Moral Expertise: This aspect resonates with the trusteeship (I'timāni) framework 's emphasis on spiritual intentionality (niyyah), as ethicists must demonstrate ethical integrity in decision-making and inspire trust through consistency and empathy.
3. Practical Moral Expertise: Both Floridi's model and the trusteeship (I'timāni) framework highlight the need for ethicists to facilitate ethical decision-making in real-world contexts. The facilitator role requires balancing competing values, akin to managing trade-offs in the trusteeship (I'timāni) framework, while ensuring that ethical principles are actionable and context-sensitive.

Floridi's depiction of AI ethicists as researchers and educators complements the trusteeship (I'timāni) framework 's call for intellectual stewardship (Epistemological Covenant). As researchers, ethicists contribute to understanding and addressing novel ethical dilemmas, particularly in unregulated or emerging technological domains. Similarly, as educators, they play a vital role in fostering ethical awareness

Table 1 Deriving regulatory solutions to prohibit manipulative algorithms, mitigate algorithmic bias, safeguard privacy and promote sustainability

Ethical threat	Application of the I'timāni covenants	Practical measures	Case study
AI-powered algorithms in advertising, recommendation systems, and social media platforms often exploit user vulnerabilities to maximize engagement or profit. Such manipulation undermines individual autonomy, fosters misinformation, and weakens social cohesion.	Ontological: Upholds human dignity (karamah) by ensuring that algorithms respect individual autonomy and avoid exploitation	Governments and institutions should enforce regulations prohibiting manipulative algorithmic practices, particularly in targeting vulnerable populations	A social media platform redesigns its recommendation algorithm to prioritize content that fosters community engagement and well-being, rather than polarizing content that maximizes screen time. This aligns with the principles of justice (adl) and spiritual accountability
	Epistemological: Demands transparency and accountability, preventing algorithms from misleading users or distorting truth	Platforms must be transparent in design and provide users with clear explanations of how content recommendations are generated, enabling informed decision-making	
	Existential: Encourages the use of AI technologies to promote societal harmony rather than creating division	Developers should design algorithms to promote truthful content and constructive engagement rather than exploiting behavioural vulnerabilities	
Algorithmic bias in AI systems perpetuates discrimination, disproportionately harming marginalized groups in domains such as hiring, lending and policing. These biases often stem from unrepresentative datasets or flawed design processes.	Ontological: Demands equity and fairness in AI decision making, protecting the dignity of all individuals	Establish rigorous auditing processes to identify and mitigate biases in AI algorithms	A hiring platform redesigns its algorithms to exclude historically biased data points, such as gender or race, and implements regular fairness audits. This reflects the Qur'anic emphasis on justice: "Indeed, Allah commands you to render trusts to whom they are due and to judge with justice." (Surah An-Nisa, 4:58)
	Epistemological: Emphasises the ethical use of knowledge to eliminate bias and promote justice (adl)	Train AI systems using inclusive datasets that reflect diverse populations to ensure equitable outcomes	
	Existential: Calls for inclusivity and social equity, ensuring that AI systems benefit all members of society	Develop independent review boards to oversee the fairness and inclusivity of AI applications	

Table 1 (continued)

Ethical threat	Application of the I'timāni covenants	Practical measures	Case study
AI systems collect, store, and analyse vast amounts of personal data, often without user consent. This raises privacy concerns and risks enabling authoritarian surveillance or data misuse.	Ontological: Recognizes privacy as a fundamental right tied to human dignity	Enforce strict data protection regulations to safeguard user information	A healthcare AI system implements end-to-end encryption for patient data and ensures data access is limited to authorized personnel. This aligns with the Qur'anic directive: "Do not spy, nor speak ill of each other behind their backs." (Surah Al-Hujurat, 49:12)
	Epistemological: Advocates for transparency in data usage, ensuring users are informed and protected	Empower users by providing individuals with control over their data, including the ability to consent to and delete collected information	
	Existential: Stresses the collective good, balancing security needs with individual freedoms	Limit the deployment of AI surveillance to contexts that ensure public safety while protecting individual rights	
Energy-intensive AI systems, such as large-scale machine learning models, contribute to environmental degradation. The carbon footprint of data centres poses significant sustainability challenges	Ontological: Recognizes the moral responsibility to preserve the environment as part of humanity's role as stewards (khilafah)	Design AI systems using energy-efficient algorithms and hardware	A cloud service provider transitions to fully renewable energy sources for its data centres, reducing its carbon footprint while maintaining efficiency. This reflects the Qur'anic principle of balance (mizan): "Do not commit abuse on the Earth, spreading corruption." (Surah Al-A'raf, 7:56)
	Epistemological: Advocates for acknowledgment and reduction of environmental impacts through responsible innovation	Promote the use of renewable energy sources in data centres	
	Existential: Mandates sustainable practices to protect natural resources for future generations	Implement policies to monitor and mitigate the environmental impact of AI development	

and literacy among stakeholders, echoing the trusteeship (I'timāni) framework 's emphasis on ethical knowledge dissemination.

The trusteeship (I'timāni) framework's integration of spiritual accountability offers a unique lens to address the perceived neutrality of Floridi's secular ethical paradigms. By incorporating metaphysical principles, such as divine trust (amana), the trusteeship (I'timāni) framework adds depth to discussions about the ultimate ends of technology and its alignment with transcendent values.

Conversely, Floridi's structured competency framework provides actionable insights for operationalizing the trusteeship (I'timāni) framework. For instance, the codification of AI ethicists' skills into categories of ethical assessment, interpersonal engagement, and systemic facilitation enables more consistent and measurable application of ethical principles.

Combining Floridi's competency model with the trusteeship (I'timāni) framework's metaphysical grounding therefore highlights the potential for a unified ethical approach. This integration can:

- Ensure that AI technologies are developed with both procedural integrity and spiritual intentionality.
- Equip ethicists with the tools to navigate diverse cultural and moral landscapes, fostering cross-cultural collaboration and inclusivity.
- Enhance the ethical maturity of organizations by embedding both universal principles and context-sensitive values in their practices.

Further research and dialogue are needed to refine the intersection of these paradigms. This includes developing shared educational initiatives, competency frameworks, and governance models that integrate Floridi's practical guidelines with the spiritual depth of the trusteeship (I'timāni) framework. Such efforts can ensure that AI ethics evolves as a multifaceted profession that not only addresses technological challenges but also aligns human creativity with moral and spiritual imperatives.

6 Discussion

This paper found that the main gaps in current approaches to Islamic ethics and Artificial Intelligence were (a) a jurisprudential law approach that was restricted in usefulness and did not wholly encompass Islamic ethical values and (b) an ambiguity in how the current proposed frameworks can be applied to complex ethical scenarios specific to the challenges of AI. We addressed these issues by considering an approach based on philosophy of ethics rather than law. Existing studies are a poor representation of Islamic Ethics, in which duty and intention matters more than the result. Using the work of a contemporary Muslim philosopher, we explored how the I'timani or trusteeship framework can supersede both differences across religious denominations as well as the range of opinions in response to the challenges of modernity. Finally, we showcased how

the comprehensive framework could be applied to ethical threats and integrated across governance and technology. A philosophical approach is not only capable of providing inputs for policy and regulation but also more conducive than a rule-based vision in the long run, as the technology rapidly evolves.

There are epistemic risks involved in producing more but understanding less causing an 'illusion of understanding' (Messeri & Crockett, 2024). In AI applications, there is a danger of solutionism (Morozov, 2013) that Berendt describes is a result of a confusion between solvable engineering problems and ill-defined social problems (Berendt, 2019). AI ethics should encompass guidelines for all relevant stakeholders on choosing whether to participate in a project or not, making sure the execution is safe and beneficial, and ensuring that the used technology is contributing to global good and is not harmful.

By a dialogue between frameworks, we showed that a maqasid-based approach does not adequately address the ethical complexities introduced by modern AI technologies, particularly the processes and intentions behind their development. There was little in the literature on how this approach can practically inform on the technicalities of the design of AI. Floridi's framework, while comprehensive, also lacks the spiritual and intentional dimensions that the Islamic frameworks provide.

The trusteeship (I'timāni) framework, rooted in Quranic principles, offers a profound ethical paradigm for navigating the complexities of Artificial Intelligence (AI). Its integration of the Ontological, Epistemological, and Existential Covenants provides a holistic foundation that bridges the metaphysical, moral, and practical dimensions of human action. By emphasising spiritual accountability, ethical intentionality, and stewardship, the framework addresses key ethical gaps in both Islamic and Western paradigms. The trusteeship (I'timāni) framework bridges critical gaps in existing ethical models by:

1. Providing a metaphysical grounding that connects ethical principles to divine accountability, surpassing the procedural neutrality of secular approaches.
2. Emphasising the moral quality of processes (niyyah) alongside outcomes, countering the teleological bias of frameworks like Maqasid al-Shariah.
3. Offering a holistic vision of stewardship (khilafah) that integrates justice (adl), sustainability (istidama), and equity (ihsan) into technological governance.

The trusteeship (I'timāni) framework promotes inclusivity by integrating diverse spiritual and cultural perspectives. It complements Western paradigms like Floridi's Information Ethics with Islamic metaphysical principles, fostering a global ethical model that resonates across traditions.

By applying the three covenants to real-world scenarios in healthcare, finance, and surveillance, the framework demonstrates its capacity to guide ethical AI development. It ensures technologies serve humanity's higher moral and spiritual goals while addressing pressing challenges such as privacy violations, algorithmic bias, and environmental harm.

However, implementing the trusteeship (I'timāni) framework may face challenges in a pluralistic world, especially cultural resistance from secular organisations that prefer procedural neutrality. Interdisciplinary teams that include ethicists, technologists and theologians can mitigate these and encourage engagement in advocacy and dialogue.

One of the conclusions of a conference on Islamic Ethics and AI in Lahore, Pakistan was: "Ethical use of AI cannot be realized just through principles. An entire system of complementary subsystems (moral, ethical, educational, economic, legal) promoting human-beneficial AI is needed" (Raquib, 2022). Requiring such an overhaul to tackle the power imbalance imposed by a techno-monopoly that believes in 'the end justifies the means' would need an approach that rejects this, not affirms it. This requires a top-down approach such as that shown in this paper. This also requires a framework that can extend beyond merely stating what is allowed or not within any legal system—it requires a philosophy that will encompass such a system yet extend beyond it to battle the technicalities in designing the technology. With regards to technology and utilising one's surroundings, Tabātabā'i considers moderation as a key virtue based on the Quranic verse on Muslims as the 'middle nation' (Arzroomchilar & Olamaiekopaie, 2022).

The interpretation of the concepts in the I'timāni framework have been brief for the purposes of this paper and interpreted for explicit application within the context of AI. There is a rich literature in Islamic sciences and philosophy that others have drawn upon or have not been explored yet. For AI-mediated work, Ghaly suggests an ihsan-based approach (moral excellence), such as that proposed in his past work in Islamic bioethics (Ghaly, 2016, Ghaly, 2023). Another contemporary philosopher of ethics, Syed Muhammad Naquib al-Attas (Al-Attas, 1995), focuses on the concept of adab (proper conduct) in his process-oriented ethical scheme. In Al-Attas's view, adab is more than just good behavior; it represents the proper placement of knowledge and the just relationship between all beings. If applied to AI, adab would require that AI systems operate within their rightful place in the ethical hierarchy, supporting human well-being without undermining autonomy or dignity. It would also be useful to explore Islamic peripatetic or transcendental philosophical approaches to technology driven by technical efficiency. An alternative vision by Wael Hallaq that is similar, in spirit, to that of Taha Abdur-Rahman's albeit using different tools is worth exploring as well (Hallaq, 2019).

This work showed how a philosophical approach can inform design of AI in tangible ways, as from the perspective of Islamic traditions, how we build AI is just as important as the 'information' we get from it. Whether that information contributes to humanity's knowledge domain, or it increases redundancy and distracts us from our higher aims, is a global challenge that requires active engagement moving forwards.

This paper is a contribution to the emergent discourse on *non-Western AI ethics*, addressing epistemic injustice (Fricker, 2007) and illustrating how a philosophically robust Islamic tradition can intersect with the complexities of modern technology. The trusteeship (I'timāni) framework offers a comprehensive vision of ethical stewardship, calling humanity to align its technological advancements with spiritual

and moral imperatives. By emphasising divine trust (amana), holistic responsibility, and collective well-being, it ensures that AI development serves humanity's higher ideals.

Departing from mainstream procedural frameworks, which focus primarily on issues of privacy, bias, and sustainability, we have argued that AI raises a deeper ontological and ethical crisis—one that concerns the very nature of human subjectivity, moral agency, and intentionality in an age increasingly dominated by algorithmic abstraction.

At the heart of this analysis, from the perspective of Taha Abdurrahman's iʾtimānī (trusteeship-based) ethics, is the recognition that ontology, ethics, and trusteeship cannot be separated. Ethical evaluations of AI must move beyond utilitarian or rights-based models that isolate outcomes from intentions, or means from ends. Instead, every technological act is embedded within a moral and metaphysical trust (*amāna*), requiring alignment between the purpose of action (maqṣad), the integrity of the means, and the moral intentionality (niyyah) of the actor.

Taha Abdurrahman's critique of *ʿaql ālī* (instrumentalized, automated reason) has provided the conceptual basis for this position, offering a path to resist the reduction of ethics to procedural governance or technocratic optimization. AI, as a paradigmatic instance of modern technē, demands such a comprehensive ethical response—one that re-centers the human being as a morally accountable trustee (muʾtaman), not merely as a data point in algorithmic systems (Hamzi, 2025). Together, the current paper and (Hamzi, 2025) present a coherent vision of how Islamic ethics—rooted in trusteeship—offers an alternative paradigm for evaluating AI and related technologies. Islamic well-being (falāḥ) is not merely a sum of welfare indicators; it is the realization of human potential in relation to God (tawḥīd), the unseen (ghayb), and the afterlife (ākhirah), through dignified agency (karāma), accountable freedom (ḥurriyya), and the fulfillment of moral trust (amāna). In short, the ethical, ontological, and epistemological dimensions of technology are inseparable. The iʾtimānī framework restores this unity, ensuring that technological development remains accountable not just to material outcomes, but to the higher moral and spiritual ends of human flourishing (falāḥ) in both worlds.

Acknowledgements The authors thank Dr Amina Inloes, Dr. Shoaib Malik, Dr. Sohail Hanif for useful comments and discussions.

Authors' contributions NA.

Funding The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data Availability NA.

Declarations

Ethics approval and consent to participate NA.

Consent for publication NA.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

Abdul Rahman, S.N.M. & Ibrahim, A.B. (2021). Maqāṣid al-Sharī'ah: Protection of 'Aql in the Era of Industrial Revolution 4.0. In *The 14th ISDEV International Islamic Development Management Conference*. Universiti Sains Malaysia Pulau Pinang.

Abubakar, A., Maheen, F., Zou, J. (2021). Persistent anti-muslim bias in large language models. https://arxiv.org/abs/2101.05783

Al-Attas, S. M. N. (1995). Prolegomena to the Metaphysics of Islam: An Exposition of the Fundamental Elements of the Worldview of Islam. *International Institute of Islamic Thought and Civilization (ISTAC).*

Altman, S. (2024). Who Will Control the Future of AI? *The Washington Post*. https://www.washingtonpost.com/opinions/2024/07/25/sam-altman-ai-democracy-authoritarianism-future/.

Arzroomchilar, E., & Olamaiekopaie, M. (2022). Where Technology Meets Islam: Towards an Islamic Perspective on Technology. *Journal of Islamic Thought and Civilization, 12*(2), 14–27.

Azar, E. & Haddad, A.N. (2021). *Artificial intelligence in the Gulf: Challenges and opportunities*. Palgrave Macmillan, Springer.

Babak, H., Varshney, L.R. (2022). Debiased large language models still associate muslims with uniquely violent acts. https://arxiv.org/abs/2208.04417

Bell, J. (2022). AI Will Be the 'Beating Heart' of Saudi's NEOM, THE LINE: Summit. Al Arabiya English. https://english.alarabiya.net/News/gulf/2022/09/16/AI-will-be-the-beating-heart-of

Berendt, B. (2019). AI for the Common Good?! Pitfalls, Challenges, and Ethics Pen-Testing. *Paladyn, Journal of Behavioral Robotics, 10*(1), 44–65.

Blackwell, A.F. (2024). *Moral codes: Designing alternatives to AI*. The MIT Press.

Buyl, M., Rogiers, A., Noels, S., Dominguez-Catena, I., Heiter, E., Romero, R., Johary, I., Mara, A.- C., Lijffijt, J., De Bie, T. (2025). Large *Language models reflect the ideology of their creators*. https://arxiv.org/abs/2410.18417v1

Chaudhary, M. Y. (2020). Initial Considerations for Islamic Digital Ethics. *Philosophy & Technology, 33*, 639–657.

Chuang, C.-W., Chang, A., Chen, M., Selvamani, M. J. P., & Shia, B.-C. (2022). A Worldwide Bibliometric Analysis of Publications on Artificial Intelligence and Ethics in the Past Seven Decades. *Sustainability, 14*(18), 11125.

Dixit, P. (2024). OpenAI's board allegedly learned about ChatGPT launch on Twitter" Engadget. https://www.engadget.com/openais-board-allegedly-learned-about-chatgpt-launch-on-twitter-235643014.html

El-Mesawi, E. T. (2020). "مقاصد الشريعة: المعنى والنطاق والآثار" (Maqasid Al-Shari'ah: Meaning, Scope, and Ramifications). *Al-Shajarah (ISTAC Journal of Islamic Thought and Civilization), 25*(2), 263–295.

Elmahjub, E. (2023a). Artificial Intelligence (AI) in Islamic Ethics: Towards Pluralist Ethical Benchmarking for AI. *Philosophy & Technology, 36*(2), 345–362. https://doi.org/10.1007/s13347-023-00598-7

Elmahjub, E. (2023b). Navigating the intersections of AI and Islamic Ethics: A response to Dr Raquib's commentary. *Philosophy & Technology, 36*, 85.

Farouk, M. A. (2022). Saudi 'Surveillance City. *Would You Sell Your Data to The Line?*" Reuters. https://www.reuters.com/article/markets/commodities/feature-saudi-surveillance-city-would-you-sell-your-data-to-the-line-idUSL8N2ZL0CM.

Floridi, L. (2010). *Information: A Very Short Introduction*. Oxford University Press.

Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press.
Floridi, L. (2013). *The Ethics of Information.* Oxford University Press.
Floridi, L. (2014a). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.
Floridi, L. (2014b). Computerizing Minds and Anthropomorphizing Machines. In P. Grim (Ed.), *The Philosophy of Artificial Intelligence* (pp. 261–275). Springer. https://doi.org/10.1007/978-3-319-10274-8_16
Floridi, L. (2019). What is an AI Ethicist? *Minds and Machines, 29*(3), 273–281. https://doi.org/10.1007/s11023-019-09512-4
Floridi, L. (2023). *The Ethics of Artificial Intelligence:* Principles, Challenges, and Opportunities. Oxford University Press.
Fricker, Miranda (2007) Epistemic Injustice: Power and The Ethics of Knowing (Oxford University Press)
Ghaly, M. (2016). *Islamic perspectives on the principles of biomedical ethics*. World Scientific & Imperial College Press.
Ghaly, M. (Ed.). (2019). Islamic Ethics and the Genome Question (Vol. 1). Brill. http://www.jstor.org/stable/https://doi.org/10.1163/j.ctvbqs7qf
Ghaly, M. (2023). What Makes Work 'Good' in the Age of Artificial Intelligence (AI)? Islamic Perspectives on AI-Mediated Work Ethics. *Journal of Ethics, 28*, 429–453.
Hallaq, W. B. (1997). *A History of Islamic Legal Theories: An Introduction to Sunni Usul al-Fiqh.* Cambridge University Press.
Hallaq, W. B. (2019). *Reforming Modernity: Ethics and the New Human in the Philosophy of Abdurrahman Taha.* Columbia University Press.
Hallaq, W. B. (2009). *Sharia: Theory, Practice*. Cambridge University Press.
Hamzi, Boumediene, Redefining Reality: An Islamic Metaphysical Critique of AI's Data-Centric Worldview, https://www.academia.edu/129072897/Redefining_Reality_An_Islamic_Metaphysical_Critique_of_AI_s_Data_Centric_Worldview, 2025.
Hashas, M., al-Khatib, M. (Eds.) (2020). Islamic ethics and the trusteeship paradigm: Taha Abderrahmane's Philosophy. In *comparative perspectives.*Studies in Islamic Ethics, vol. 3. Brill
Kamali, M. H. (2022). Between separation and unity: The interplay of law and morality in islamic jurisprudence. In *Sharia law In The Twenty-First Century* (pp. 21-46).
Katibah, L. (2024). The genocide will be automated—Israel, AI and the future of war, *Middle East Report* 312.
Khan, S. (2024a). The futures of intelligence: On ethics, inclusive AI, decolonising AI, indigenous AI, emotional AI. *Suhairk Substack*. https://suhairk.substack.com
Khan, S. (2024b). Why inclusive AI matters for the creative industry. D&AD Annual. https://www.dandad.org/annual/2024/editorial/voices-why-inclusive-ai-matters-for-the-creative-industry
Messeri & Crockett 2024, Artificial intelligence and illusions of understanding in scientific research. https://www.nature.com/articles/s41586-024-07146-0
Mohadi, M., & Tarshany, Y. (2023). "مقاصد الشريعة وأخلاقيات الذكاء الاصطناعي: تحديات معاصرة" (Maqasid Al-Shari'ah and the Ethics of Artificial Intelligence: Contemporary Challenges). *Journal of Contemporary Maqasid Studies, 2*(2), 79–102.
Morozov, E. (2013). *To Save Everything*. The Folly of Technological Solutionism. PublicAffairs.
Munn, L. (2023). The uselessness of AI ethics. *AI Ethics, 3*, 869–877.
Mutahhari, M. (2002). *Understanding Islamic Sciences*. Islamic College for Advanced Studies Publishing.
Mutahhari, M. (2003). Man and Universe. Ansariyan Publications.
Nawi, M., et al. (2022). Development of islamic research ethics framework: *Towards a global code of ethics for artificial intelligence study within shariah parameters*. Project Report.
Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak. "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025. https://hai.stanford.edu/ai-index/2025-ai-index-report
Pattison, G. (2007). Thinking About God in an Age of Technology, *Oxford University Press*.

Rachmawati, F. et al. (2024). The ethical implications of AI in expressing religious beliefs online: A restatement of the concept of religion. In *Proceedings International Conference on Religion, Science and Education* (Vol. 3, pp. 71–79).
Raquib, A., Channa, B., Zubair, T., et al. (2022). Islamic Virtue-Based Ethics for Artificial Intelligence. *Discover Artificial Intelligence, 2*, 11.
Shabaz Patel, Hassan Kane, Rayhan Patel, 2023, Building Domain-Specific LLMs Faithful To The Islamic Worldview: Mirage or Technical Possibility, https://arxiv.org/abs/2312.06652
Taha, A. (1994). العمل الديني وتجديد العقل (Religious action and the renewal of reason). المركز الثقافي العربي,.
Taha, A. (2000).سؤال الأخلاق: مساهمة في النقد الأخلاقي للحداثة الغربية (The Question of Ethics: Contributions to the Ethical Critique of Western Modernity). المركز الثقافي العربي,.
Taha, A. (2001). La religion et la morale: *Contribution à une critique de la rationalité morale contemporaine*. Presses Universitaires du Maroc.
Taha, A. (2005). الحق الإسلامي في الاختلاف الفكري (The islamic right to intellectual diversity). المركز الثقافي العربي,.
Taha, A. (2006). روح الحداثة: المدخل إلى تأسيس الحداثة الإسلامية (The spirit of modernity: *An islamic critique of modernity)*. المركز الثقافي العربي,.
Taha, A. (2020a). المفاهيم الأخلاقية بين الائتمانية والعَلْمانية: 2-المفاهيم العلمانية (*Ethical concepts between theism and secularism:* 2-secular concepts). المركز الثقافي العربي,.
Taha, A. (2020b). المفاهيم الأخلاقية بين الائتمانية والعَلْمانية: 1- المفاهيم الائتمانية (Ethical concepts between *Trusteeship and Secularism: 1-Trusteeship Concepts*). المركز الثقافي العربي,.
Taha, A. (2022). التأسيس الائتماني لعلم المقاصد (The Trusteeship Foundation of the Science of Objectives). المركز الثقافي العربي,.
Taha, A. (2023). سؤال السيرة الفلسفية: بحث في حقيقة التفلسف الائتمانية (The question of philosophical biography: *An Investigation into the Reality of Trusteeship Philosophy*). المركز الثقافي العربي,.
Waters, B. (2006). From Human to Posthuman: *Christian Theology and Technology in a Postmodern World.* Ashgate Publishing Limited.
Zaim, Mahmoud, and Abdel Rahman Maidi. (2023). "قواعد الشريعة التي تحكم تقنيات الذكاء الاصطناعي" (The Sharia Rules Governing Artificial Intelligence Technologies). In *Proceedings of the Conference "Artificial Intelligence and Its Applications in Islamic Sciences",* edited by Ibrahim Rahmani, Miloud Laifa, and Tijani Ad.
Zubair, Talat, and Dr. Amana Raquib. (2020). "Islamic Perspective on Social Media Technology, Addiction, and Human Values". *Journal of Islamic Thought and Civilization 10* (2).

Authors and Affiliations

Fatima Ali[1] · Karim Bouzoubaa[2,3] · Frank Gelli[4] · Boumediene Hamzi[5,6] · Suhair Khan[4]

✉ Boumediene Hamzi
bhamzi@turing.ac.uk

[1] University College London, London, UK

[2] Mohammadia School of Engineers, Mohammed Vth University, Rabat, Morocco

[3] Department of Computing, University of Roehampton, London, UK

[4] Central Saint Martins, University of the Arts London, London, UK

[5] Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA

[6] The Alan Turing Institute, London, UK