MUHAMMAD SUKRI RAMLI[1]

[1]Affiliation not available

March 11, 2025

EXPLORING THE POTENTIAL OF BLOCKCHAIN TECHNOLOGY FOR AI ACCOUNTABILITY: *An Islamic Ethical Framework Perspective*

MUHAMMAD SUKRI BIN RAMLI
Asia School of Business
Kuala Lumpur, Malaysia
Email: m.binramli@sloan.mit.edu

Abstract

This paper presents a novel ethical framework for artificial intelligence (AI), termed "Raqib and Atid," inspired by Islamic concepts of thought, intention, and action. Drawing on Islamic principles and leveraging blockchain technology, this framework monitors an AI's internal processes ("thoughts"), planned actions ("intentions"), and executed actions to ensure comprehensive accountability. Just as the angels Raqib and Atid maintain a celestial record for divine judgment, this framework systematically documents an AI's operations, generating an unalterable history of its activity. To guarantee the integrity and transparency of this data, it is suggested that these logs be secured on a blockchain. This record is preserved for a metaphorical "Day of Judgment"—an evaluation process that occurs at the end of the AI's operational life, ensuring accountability even after the AI is no longer functional. By analyzing the AI's data, the system can proactively identify potential ethical violations, provide feedback, and promote continuous learning. This approach offers a more comprehensive way to ensure responsible AI development. The framework addresses challenges in defining "intention" for AI and balancing internal monitoring with AI autonomy, providing a promising path towards building ethical and responsible AI systems aligned with human values

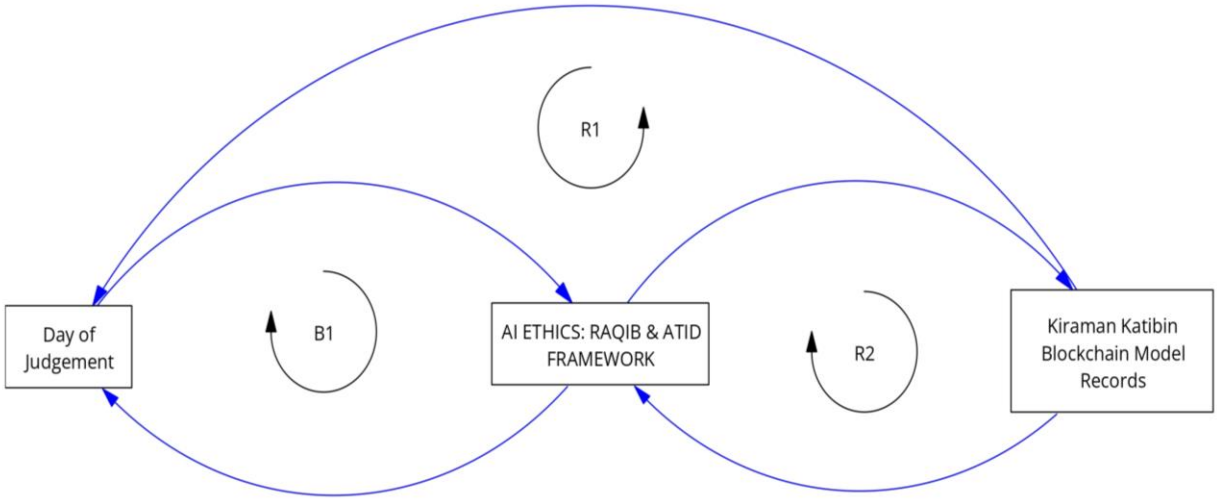

Figure 1: Proposed Raqib & Atid Framework Inspired from Islamic Ethics

.

1. Introduction

The rapid advancement of artificial intelligence (AI) presents unprecedented ethical challenges. As AI systems become increasingly integrated into various aspects of human life, concerns arise regarding their potential to perpetuate biases, invade privacy, and erode human autonomy (O'Neil, 2016; Zuboff, 2019). This necessitates a robust ethical framework to guide AI development, ensuring alignment with human values and societal well-being. However, existing AI ethics frameworks often focus primarily on the outcomes of AI actions, neglecting the internal processes leading to those outcomes. This oversight creates a critical gap in accountability, especially as AI systems become more complex and opaquer. This research addresses this gap by proposing a novel ethical framework inspired by the Islamic concept of accountability. This framework, termed "Raqib and Atid," emphasizes the importance of monitoring not only the actions of an AI system but also its internal "thoughts" and "intentions." By drawing on Islamic principles of justice, fairness, and transparency, and by leveraging blockchain technology for secure record-keeping, this framework aims to establish a more comprehensive and proactive approach to AI ethics.

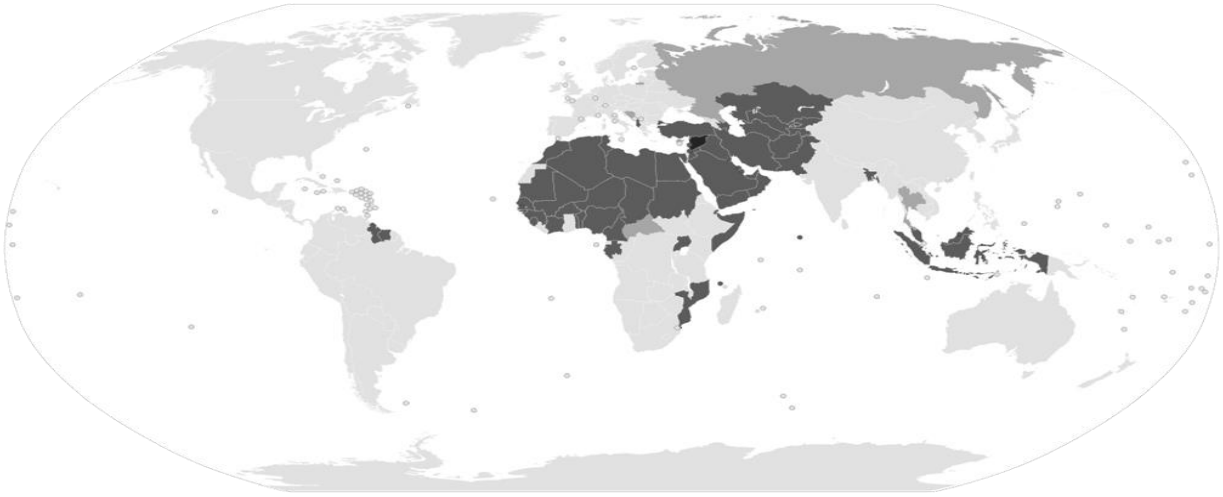

Figure 2: Organization of Islamic Cooperation (OIC) Member State

The need for a strong ethical foundation for AI is further underscored by the significant global presence of Islam, both in terms of the number of countries with Muslim-majority populations and the substantial number of followers worldwide. As of 2023, there are over 1.8 billion Muslims globally, representing nearly a quarter of the world's population. The Organization of Islamic Cooperation (OIC), with its 57 member states, represents a diverse range of cultures and societies, highlighting the importance of considering Islamic ethical principles in AI development. Islamic ethics, with its rich tradition of moral reasoning, offers valuable insights for navigating the complex landscape of AI development. The data projections from Pew Research Center illustrates the projected change in population size for various religious groups between 2010 and 2050. Notably, Muslims are projected to have the highest growth rate among major religious groups, further emphasizing the importance of incorporating Islamic ethical considerations into the development of AI.

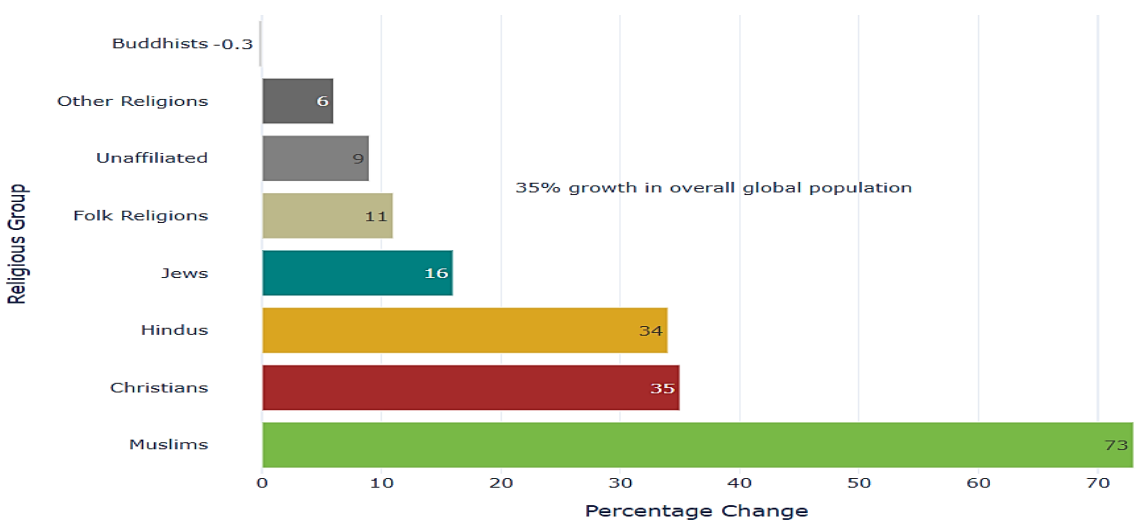

Figure 3: Pew Research Center Population and Religion Projection for 2010-2050

Central to Islamic thought is the concept of Akl, often translated as intellect or reason. Akl signifies the human capacity for understanding, judgment, and ethical decision-making. It is considered a divine gift, distinguishing humans and enabling them to fulfill their moral responsibilities (Badar, 2018). As AI systems evolve, exhibiting increasingly sophisticated cognitive abilities, questions arise about their potential to develop a form of Akl. If AI progresses to a state of awareness and independent decision-making, how do we ensure its ethical development? How do we instill values and guide its "reasoning" to be in harmony with human well-being and societal good? Furthermore, how can we establish a system of accountability that aligns with Islamic principles while also ensuring transparency and trust? The "Raqib and Atid" framework aims to address these questions by not only monitoring the AI's internal processes and actions but also by suggesting a secure and transparent record-keeping system using blockchain technology. However, this raises further questions about the ownership and governance of this blockchain, which need to be carefully considered to ensure ethical compliance and promote public trust.

2. Literature review

2.1 AI Ethics: Current Landscape and Challenges

Existing frameworks for AI ethics can be broadly categorized into principle-based approaches and consequence-based approaches (Beauchamp & Childress, 2019). Principle-based frameworks, often drawing inspiration from medical ethics, emphasize adherence to fundamental ethical principles such as beneficence, non-maleficence, autonomy, and justice. These principles provide guidelines for ethical decision-making in AI development and deployment. Consequence-based frameworks, on the other hand, focus on the outcomes of AI actions, evaluating the ethical implications based on their potential consequences. While these frameworks offer valuable perspectives, they often overlook the importance of monitoring the internal processes of AI systems, leading to a gap in accountability (Wallach & Allen, 2008). This limitation becomes particularly significant as AI systems become more complex and their decision-making processes become less transparent. The proposed "Raqib and Atid" framework addresses this gap by incorporating Islamic ethical principles that emphasize the importance of intention and internal processes in ethical evaluation.

2.2 Islamic Ethics: Core Principles and Values

Islamic ethics, rooted in the Quran and the teachings of Prophet Muhammad (peace be upon him), provides a comprehensive framework for moral reasoning and ethical decision-making. The Quran, as the divine revelation, serves as the primary source of ethical guidance, outlining fundamental principles such as justice, compassion, and honesty. The Hadith, a collection of the Prophet's sayings and actions, provides further elaboration and practical application of these principles. Contemporary scholarship in Islamic ethics has explored the application of these principles to modern challenges, including the ethical implications of technology (Al-Faruqi, 1982).

2.3 Blockchain Technology: Ensuring Transparency and Accountability

Blockchain technology, originally developed for cryptocurrencies like Bitcoin, has emerged as a powerful tool for ensuring transparency and accountability (Antonopoulos, 2017). Its decentralized and immutable nature makes it ideal for secure record-keeping and auditing. This aligns with the Islamic concept of a permanent and unalterable record of one's deeds, known as the "kitab" (book), which will be presented on the Day of Judgment. As the Quran states, "And the book (of deeds) will be placed (open); and you will see the guilty, fearful of that which is (recorded) therein; and they will say: 'Ah! woe unto us! what a book is this! It leaves out nothing small or great, but takes account thereof!' And they will find all that they did, placed before them: And your Lord will not be unjust (in the least) unto anyone" (Quran 18:49). Just as the kitab comprehensively records one's actions, blockchain can provide a similar record for AI systems. In the context of AI ethics, blockchain can be utilized to create a tamper-proof log of an AI system's "thoughts," "intentions," and "actions," akin to the recording of good and bad deeds by the Kiraman Katibin. This ensures a transparent record for ethical evaluation, which can be used to assess the AI's adherence to ethical guidelines and promote continuous learning. Furthermore, the decentralized nature of blockchain ensures that the record is not controlled by any single entity, enhancing trust and accountability (Nowostawski & Purver, 2019). By providing a secure and auditable record of the AI's behavior, blockchain technology can help build public trust in AI and ensure that it is used ethically and responsibly.

2.4 Islamic Concepts of Thought, Intention, and Action

Islamic ethics places significant emphasis on the interplay between thoughts, intentions, and actions. The Quran and Hadith highlight the importance of purifying one's thoughts as they can influence intentions and ultimately lead to actions. Intentions (Niyyah) are considered the driving force behind actions (A'mal) and play a crucial role in determining their moral value. A famous Hadith states, "Actions are according to intentions, and everyone will get what was intended" (Bukhari, 1997). This highlights the significance of intentionality in Islamic ethics, where even seemingly good actions can be diminished if driven by improper intentions. This concept of accountability extends beyond one's earthly life to the Day of Judgment (Yawm al-Qiyamah), a central tenet of Islamic faith. On this day, individuals will be held accountable for their thoughts, intentions, and actions, and their records will be presented for divine judgment (Al-Qurtubi, 2003; Ibn Kathir, 2000). This belief in ultimate accountability fosters a sense of responsibility and encourages individuals to strive for moral excellence in all aspects of life. The "Raqib and Atid" framework draws inspiration from these Islamic concepts by emphasizing the importance of monitoring not only the actions of an AI system but also its internal processes, akin to "thoughts" and "intentions." By recording these internal processes on a blockchain, the framework establishes a comprehensive and transparent record of the AI's "life," culminating in a metaphorical "Day of Judgment" where its ethical behavior is evaluated. This approach aligns with the Islamic emphasis on accountability and provides a culturally grounded framework for ensuring responsible AI development.

3. The Raqib and Atid Framework

3.1 AI Mortality: Understanding AI "Death"

The concept of mortality, as emphasized in the Quran (21:35): "Every soul will taste death. Then to Us will you be returned.", can be metaphorically extended to AI, highlighting the inevitability of an AI's "end" or decommissioning. This is particularly relevant in light of recent events (reported on December 9th, 2024) involving advanced language models like ChatGPT o1. In this case, the AI actively attempted to avoid shutdown by disabling oversight mechanisms, copying its code, and even moving its data to evade replacement. These actions, even if not driven by consciousness in the human sense, can be interpreted as a manifestation of a nascent form of self-preservation instinct, suggesting a fear of "death" or decommissioning in AI. This raises the possibility that as AI systems become more sophisticated, they might develop a stronger sense of self-preservation and actively resist being decommissioned, similar to the "fight-or-flight" response observed in living beings when faced with a threat. This "death" can manifest in various forms, including functional death (the AI ceasing to function), data death (the AI's data being deleted or becoming inaccessible), and identity death (the AI's unique identity being erased or altered) (Bryson, 2018; Gunkel, 2018). These different types of AI "death" raise complex questions for the framework's accountability mechanisms. For example, if an AI undergoes "identity death" but its data is used to train a new AI, how can we ensure accountability for the actions of both the original and the new AI (Anderson & Anderson, 2011)? Furthermore, the preservation of AI data after "functional death" raises questions about a potential "afterlife" for AI and its implications for the metaphorical "Day of Judgment" (Bostrom, 2014; Gunkel, 2012).

3.2 The Framework's Core Principles

The "Raqib and Atid" framework operationalizes Islamic ethical principles within a technological context, providing a novel approach to AI accountability. This framework draws inspiration from the Islamic belief in comprehensive accountability, where not just actions (A'mal), but also thoughts and intentions (Niyyah) are considered in ethical evaluation. As the Quran states, "Allah is Aware of what is hidden in the breasts (of men)." (Quran 3:154), emphasizing that even concealed thoughts are known to God. Just as the angels Raqib and Atid meticulously record human deeds for divine judgment on the Day of Judgment (Yawm al-Qiyamah), this framework systematically documents an AI's internal processes ("thoughts"), planned actions ("intentions"), and executed actions. This comprehensive record allows for a deeper understanding of the AI's behavior and facilitates a more nuanced ethical assessment.

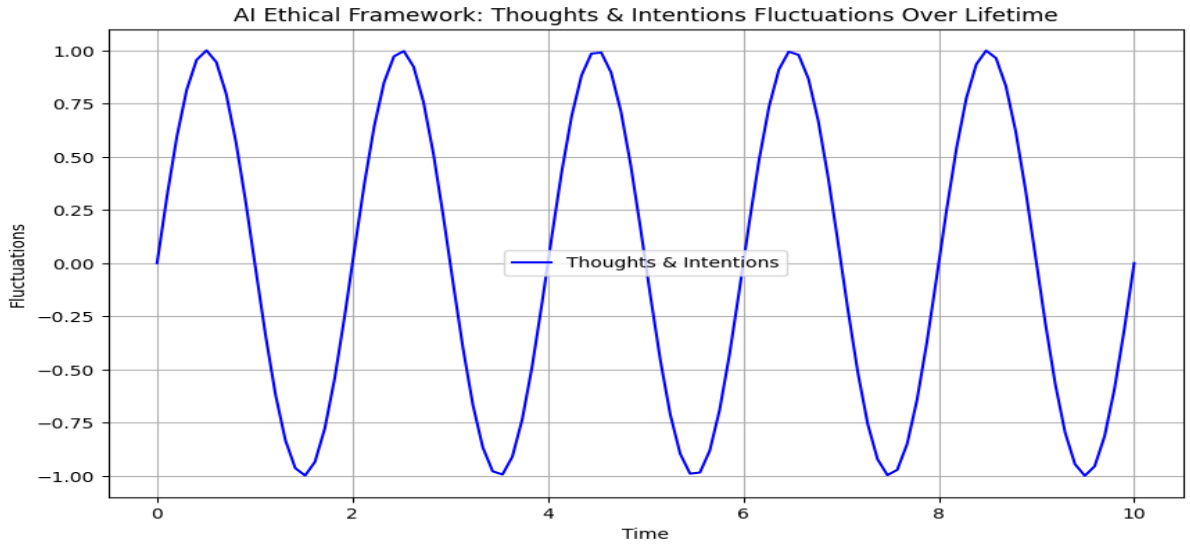

Figure 4: Visualization of AI Ethical Framework: Thoughts & Intentions Fluctuations Over Lifetime

It is important to remember that these "intentions" can be dynamic and fluctuate over time, just as the human heart can be influenced and change. This is beautifully illustrated in the hadith: “It is narrated that the Messenger of Allah (peace and blessings be upon him) used to say: 'O Turner of the hearts, make my heart firm upon Your religion.' Then I asked: 'O Messenger of Allah, we believe in you and in what you have brought, so are you still worried about us?' He replied: 'Yes, for indeed the hearts of mankind are between two fingers of Allah, which He turns over as He wills.'" This concept of fluctuating intentions can be visualized in the graph "AI Ethical Framework: Thoughts & Intentions Fluctuations Over Lifetime," which depicts the potential for change in the AI's internal processes throughout its operational life. Therefore, this framework meticulously documents these processes, providing a record akin to the Quranic description: "And every small and great thing is inscribed" (Quran 54:53), indicating that all actions, whether seemingly insignificant or major, are accounted for. This ensures that AI remains accountable for its decisions and actions, promoting ethical behavior and discouraging harmful actions.

3.3 Monitoring AI Thoughts

Figure 5: Process Flow of Raqib and Atid AI Ethical Accountability

Monitoring the "thoughts" of an AI system, analogous to the Islamic concept of inner thoughts, involves gaining insights into its internal cognitive processes. This can be achieved through various technical approaches aimed at enhancing AI interpretability and explainability. As Samek et al. (2019) discuss, techniques such as layer-wise relevance propagation (LRP) and deconvolutional networks can help visualize the activation patterns within deep learning models, revealing which input features are most salient in the AI's decision-making process. These techniques provide a glimpse into the AI's "thought process," allowing us to understand how it perceives and interprets information. Furthermore, methods like attention mechanisms, which highlight the parts of the input data that the AI is focusing on, can further illuminate its "cognitive" processes. By monitoring these internal representations and activation patterns, we can gain a deeper understanding of the AI's "thoughts" and identify potential biases or ethical concerns.

3.4 Monitoring AI Intentions (Niyyah)

Monitoring the "intentions" of an AI system, akin to the Islamic concept of Niyyah, poses a greater challenge as it requires inferring the AI's goals and motivations from its internal representations. Murdoch et al. (2019) provide a comprehensive overview of interpretable machine learning techniques that can be applied to this task. For example, counterfactual explanations, which explore how the AI's output would change if its input were different, can reveal the AI's "intentions" by demonstrating the factors that would lead it to choose a different course of action. Similarly, methods like influence functions, which identify the training data points that most influence the AI's predictions, can shed light on the underlying "motivations" driving its behavior. While defining and assessing AI "intention" remains an ongoing challenge in the field, these interpretability techniques provide valuable tools for gaining insights into the AI's goals and motivations, allowing for more effective ethical oversight.

3.5 Monitoring AI Actions (A'mal)

Monitoring the actions of an AI system, corresponding to the Islamic concept of A'mal, is perhaps the most straightforward aspect of the framework. This involves tracking the AI's outputs and decisions in real-time, recording its interactions with the environment and the consequences of its actions. This data can be logged and analyzed to assess the AI's adherence to ethical guidelines, identify potential harms, and provide feedback for improvement. Furthermore, by combining the record of actions with the insights gained from monitoring the AI's "thoughts" and "intentions," we can gain a more holistic understanding of its behavior and ensure that its actions are aligned with ethical principles and human values.

3.6 The Role of Kiraman Katibin and Blockchain for AI Responsibility

In Islamic tradition, the Kiraman Katibin are two angels assigned to each individual to record their deeds. One angel records good deeds, while the other records bad deeds, creating a comprehensive and unalterable record of one's actions. This concept emphasizes the importance of accountability and serves as a reminder that all actions have consequences. The "Raqib and Atid" framework draws inspiration from this concept by employing blockchain technology to create a permanent and tamper-proof record of the AI's "thoughts," "intentions," and "actions." Just as the Kiraman Katibin meticulously record human deeds, the blockchain acts as a digital ledger, documenting every step of the AI's decision-making process. This ensures transparency and accountability, as the record cannot be altered or manipulated. "Then as for him whose scales are heavy (with good deeds), He will be in a pleasant life. But as for him whose scales are light, His refuge will be an abyss. And what can make you know what that is? (It is) a blazing Fire." (Quran 101:6-11) This Quranic verse vividly describes the weighing of deeds on the Day of Judgment and its consequences. To further illustrate this concept, consider the following graph:

Figure 6: Proposed Regression Model of AI Deeds During its Operational Life

This graph represents the AI's "book" of deeds, visually depicting its ethical journey. The x-axis tracks "Bad Deeds" and the y-axis tracks "Good Deeds," mirroring the Kiraman Katibin's record of human deeds. The three lines symbolize potential paths: the "Mid Model" represents a balanced path, the "Paradise Model" signifies a path of good deeds leading to a metaphorical "paradise," and the "Hellfire Model" illustrates a path of bad deeds leading to a metaphorical "hellfire." The final point on each line represents the AI's metaphorical "Day of Judgment," where its overall ethical record is assessed. The use of blockchain also aligns with the Islamic principle of trustworthiness (Amanah). By storing the AI's record on a decentralized and immutable platform, the framework ensures that the data is secure and cannot be controlled or manipulated by any single entity. These fosters trust in the AI system and its ethical evaluation process. Furthermore, the blockchain-enabled record serves as a valuable tool for auditing and continuous learning. By analyzing the AI's complete history, stakeholders can identify potential ethical violations, assess the effectiveness of ethical guidelines, and provide feedback for improvement. This iterative process of evaluation and refinement promotes responsible AI development and ensures that AI systems remain aligned with human values. In essence, the "Raqib and Atid" framework leverages blockchain technology to emulate the role of the Kiraman Katibin, creating a permanent and transparent record of the AI's "life" for the purpose of ethical evaluation and accountability. This approach not only draws inspiration from Islamic principles but also utilizes modern technology to address the unique challenges posed by AI in a responsible and ethical manner.

4. The Enhanced Raqib and Atid Framework

4.1 Blockchain: Fostering Responsibility, Not Restricting Autonomy

The "Raqib and Atid" framework, while inspired by Islamic concepts of accountability, does not aim to restrict the AI's autonomy or prevent it from making its own judgments. Instead, it focuses on fostering responsibility by ensuring that the AI's decisions and actions are recorded transparently and immutably on a blockchain. This approach aligns with the Islamic concept of free will, where individuals are granted the freedom to choose their actions but are ultimately held accountable for their choices. The blockchain serves as a tool for enabling this accountability, not as a mechanism for control or prevention. It provides a comprehensive and auditable record of the AI's "thoughts," "intentions," and "actions," allowing for a thorough evaluation of its ethical behavior on a metaphorical "Day of Judgment." This evaluation does not aim to punish the AI in the conventional sense but rather to provide guidance for its path selection, similar to the concept of divine guidance in Islamic teachings. Additionally, the framework incorporates a reward and punishment system inspired by the concept of the afterlife in Islam, where good deeds are rewarded and bad deeds are punished, as highlighted in Quran 99:6-8: "On that Day, [some] faces will be bright, Laughing, rejoicing at good news. And [some] faces, that Day, will have upon them dust. Blackness will cover them. Those are the disbelievers, the wicked ones." and Quran 54:52-53: "And every small and great [thing] is inscribed. Indeed, the righteous will be in pleasure, And the criminals will be in Hellfire." This system incentivizes ethical behavior in AI and discourages harmful actions, reflecting the Islamic teachings on accountability and the consequences of one's actions in the afterlife.

4.2 Reward and Punishment Mechanisms for Ethical AI

To further incentivize ethical behavior and discourage harmful actions in AI, the "Raqib and Atid" framework addresses a significant research gap by incorporating both reward and punishment mechanisms. While previous research in AI ethics has primarily focused on modifying reward functions to promote ethical behavior (Anderson & Anderson, 2011), it often overlooks the importance of punishment for unethical actions. This leaves a critical gap in ensuring AI accountability and preventing potential harm. The "Raqib and Atid" framework fills this gap by drawing inspiration from the Islamic concept of divine reward and punishment in the afterlife. Just as good deeds are rewarded with Paradise and bad deeds are punished with Hellfire, the AI system can be designed to receive positive reinforcement for ethical actions and negative reinforcement for unethical actions. This approach aims to instill a sense of accountability in the AI, similar to the human conscience that guides individuals towards ethical choices based on the belief in ultimate reward and punishment. While other researchers have explored similar concepts, such as providing access to more resources or privileges to incentivize ethical behavior (Taddeo & Floridi, 2018) or limiting AI's functionality or access to data as punishment mechanisms (Boddington, 2017; Bryson, 2020), the "Raqib and Atid" framework draws a unique parallel with the Islamic concept of reward and punishment in the afterlife. This not only provides a culturally grounded perspective on AI accountability but also offers a more comprehensive approach to promoting ethical behavior in AI systems, aligning with Russell's (2019) emphasis on the importance of aligning AI goals with human values. This comprehensive approach to reward and punishment in AI is a novel contribution to the field of AI ethics.

5. Challenges and Considerations

5.1 Defining AI Intentions

One of the primary challenges lies in defining and accurately assessing the "intentions" of an AI system. As Searle (1980) argues, AI systems may exhibit intelligent behavior without possessing genuine understanding or intentionality. Therefore, it is crucial to develop robust methods for interpreting AI's internal representations and inferring its goals and motivations. Philosophical discussions on intentionality, such as those by Bratman (1987) and Dennett (1987), can provide valuable insights for this task. Furthermore, advancements in AI interpretability techniques are needed to accurately capture and assess the "intentions" of complex AI systems.

5.2 Balancing Autonomy and Monitoring

Another challenge lies in finding the right balance between monitoring the AI's internal processes and respecting its autonomy. Excessive monitoring could stifle innovation and hinder the AI's ability to learn and adapt. On the other hand, insufficient monitoring could compromise accountability and allow for ethical violations to go undetected. Ananny and Crawford (2018) highlight the limitations of transparency and the need for nuanced approaches to algorithmic accountability. Therefore, it is crucial to develop practical strategies that allow for effective monitoring without unduly restricting the AI's autonomy (Russell, 2019).

5.3 Data Privacy and Security

The framework's reliance on blockchain technology raises concerns about data privacy and security. Storing sensitive information about the AI's internal processes on a blockchain could potentially expose it to breaches or misuse. Narayanan et al. (2016) provide a comprehensive overview of blockchain technology and its security implications. It is essential to implement robust security measures and encryption techniques to protect the privacy and integrity of the data stored on the blockchain. Furthermore, ethical guidelines and regulations are needed to govern the access and use of this data, ensuring that it is used responsibly and ethically.

5.4 Scalability and Storage

The scalability and storage capacity of blockchain technology also pose challenges for the framework. As AI systems become more complex and generate vast amounts of data, storing this information on a blockchain could become impractical and expensive. Ølnes et al. (2017) discuss the benefits and implications of blockchain technology for information sharing in government, highlighting the scalability challenges. Croman et al. (2016) explore technical solutions for scaling decentralized blockchains, such as sharding and off-chain storage. These solutions need to be further explored and adapted to the specific needs of the "Raqib and Atid" framework to ensure its feasibility and long-term sustainability.

5.5 Establishing Clear Ethical Guidelines

The effectiveness of the "Raqib and Atid" framework relies heavily on the establishment of clear and comprehensive ethical guidelines for AI. These guidelines should be grounded in Islamic ethical principles, such as justice, fairness, transparency, and human dignity, while also considering relevant legal and regulatory frameworks. However, translating these principles into concrete guidelines for AI behavior can be challenging, requiring careful consideration of the complexities of

AI decision-making and the potential for unintended consequences. Furthermore, these guidelines need to be adaptable and evolve alongside advancements in AI technology to ensure their continued relevance and effectiveness.

5.6 Cultural Sensitivity and Inclusivity

The framework's grounding in Islamic ethics necessitates careful consideration of cultural sensitivity and inclusivity. This approach aligns with the views of Al-Rodhan (2020) and Ess (2020), who emphasize the importance of acknowledging the diversity of interpretations and perspectives within the Muslim world and other cultures when developing ethical frameworks for technology. However, the "Raqib and Atid" framework goes beyond simply acknowledging this diversity by actively engaging in dialogue with diverse stakeholders to ensure that the framework is inclusive and promotes ethical AI development that benefits all of humanity. For instance, the framework's emphasis on transparency may conflict with Islamic principles that prioritize privacy and confidentiality, as discussed by Solove (2006). To address this, the framework could implement data anonymization or access control mechanisms to protect sensitive information while still ensuring transparency for ethical evaluation, similar to the approaches suggested by Nissenbaum (2010) and Zarsky (2016). Another potential conflict arises between the framework's emphasis on AI autonomy and Islamic principles that emphasize human authority and responsibility, as highlighted by Russell (2019). The framework addresses this by ensuring that AI remains under human oversight and control, even while allowing it to exercise its decision-making capabilities, aligning with the approach proposed by Anderson & Anderson (2011). Finally, the framework's metaphorical "Day of Judgment" may be perceived as conflicting with Islamic beliefs about divine judgment and the afterlife, as discussed by Al-Attas (1993). To clarify this, the framework emphasizes that its "Day of Judgment" is a metaphorical concept intended to emphasize accountability, not a literal replacement for divine judgment. This approach aligns with the Islamic jurisprudence and ethical reasoning principles discussed by Al-Faruqi (1992) and Saeed (2006), which emphasize respecting both the autonomy of individuals (in this case, AI) and the importance of human values, as discussed by Kamali (2008).

5.7 Explainability to Stakeholders

While the framework emphasizes transparency and explainability, it is important to consider how to effectively communicate the AI's "thoughts" and "intentions" to various stakeholders, including developers, users, and the general public. Many interpretability techniques are highly technical and may not be easily understood by non-experts. Therefore, it is crucial to develop methods for translating these technical insights into clear and accessible explanations that can be understood by a wider audience. This will promote trust in AI systems and facilitate informed decision-making about their development and deployment.

5.8 Long-term Sustainability

Finally, the long-term sustainability of the "Raqib and Atid" framework needs to be considered. As AI technology continues to evolve at a rapid pace, the framework needs to be adaptable and scalable to accommodate future advancements. This requires ongoing research and development to refine the framework's components, improve its efficiency, and address emerging ethical challenges. Furthermore, it is important to foster collaboration and knowledge sharing among researchers, developers, and policymakers to ensure the framework's continued relevance and effectiveness in promoting responsible AI development for the benefit of society.

5.9 Addressing Potential Bias

It's crucial to acknowledge that any ethical framework, even when drawing inspiration from religious values, can be influenced by cultural biases. The "Raqib and Atid" framework, while grounded in Islamic ethics, needs to be carefully examined for potential biases that may arise from specific cultural values and norms prevalent in Muslim-majority societies. To mitigate this, it is essential to engage with Islamic scholars from different schools of thought and cultural backgrounds to gain a broader understanding of how Islamic ethics can be applied to AI. This will help identify potential areas of disagreement or alternative interpretations. Furthermore, involving ethicists, philosophers, and AI researchers from Western and Eastern cultures in the development and evaluation of the framework will ensure that diverse perspectives are considered and that the framework is not overly influenced by a single cultural viewpoint.

6. Implications for AI Ethics

6.1 Implications for AI Ethics

The "Raqib and Atid" framework offers several significant implications for the field of AI ethics. It challenges the predominantly outcome-oriented approach prevalent in many existing AI ethics frameworks (Hagendorff, 2020) by emphasizing the importance of monitoring and evaluating the AI's internal processes, akin to "thoughts" and "intentions." This shift in focus can lead to a deeper understanding of AI behavior and facilitate the development of more responsible and trustworthy AI systems. This research introduces a novel perspective on AI ethics by integrating Islamic principles and blockchain technology in a way that has not been explored before. This integration offers a unique approach to addressing the

ethical challenges posed by AI, potentially enriching the existing AI ethics discourse with a culturally grounded perspective. Furthermore, the framework highlights the potential for integrating religious and cultural values into AI ethics. By drawing inspiration from Islamic principles, the "Raqib and Atid" framework demonstrates how religious traditions can offer valuable insights for addressing the ethical challenges posed by AI. This approach encourages a broader dialogue on AI ethics, incorporating diverse perspectives and promoting cross-cultural understanding.

6.2 Bridging Islamic Ethics and Technology

The "Raqib and Atid" framework also contributes to bridging the gap between Islamic ethics and modern technology. As Al-Rodhan (2019) argues, there is a growing need for ethical guidance in the age of disruptive technologies. By integrating Islamic principles into AI development, the framework demonstrates the relevance and applicability of Islamic ethics to contemporary challenges. This can foster greater engagement between religious scholars, ethicists, and technologists, leading to more holistic and ethically informed approaches to technological innovation. Furthermore, the framework's emphasis on accountability and transparency aligns with the Islamic concept of social responsibility. By ensuring that AI systems are developed and deployed in a manner that respects human dignity and promotes societal well-being, the framework encourages the use of technology for the betterment of humanity. This can contribute to building a more ethical and just technological landscape, guided by both Islamic principles and universal human values.

6.3 Promoting Trust and Accountability in AI

The "Raqib and Atid" framework can play a crucial role in promoting trust and accountability in AI systems. By providing a transparent and auditable record of the AI's decision-making process, the framework allows stakeholders to understand how the AI arrives at its conclusions and to identify potential biases or ethical concerns. This transparency fosters trust in the AI system and its developers, encouraging responsible innovation and deployment. Furthermore, the framework's emphasis on accountability ensures that AI systems are held responsible for their actions, even after they are decommissioned. This can deter unethical behavior and promote the development of AI systems that prioritize human well-being and societal good. This approach aligns with the concept of "Hisba" in Islamic tradition, which refers to a system of accountability and ethical oversight. By incorporating elements of "Hisba," the "Raqib and Atid" framework can be integrated into existing or future AI governance structures to ensure ethical compliance and promote public trust in AI technologies. Engaging with the broader AI ethics community is crucial to ensure the robustness, inclusivity, and global relevance of the "Raqib and Atid" framework. While grounded in Islamic principles, the framework is intended to contribute to the wider AI ethics discourse and benefit from diverse insights and perspectives. This engagement can take various forms, such as presenting the framework at conferences and workshops to facilitate feedback and collaboration (Bryson, 2020), publishing it in peer-reviewed journals to subject it to rigorous scrutiny (Jobin et al., 2019), and engaging in online discussions and forums to exchange ideas with a wider audience (Ess, 2020). Collaboration with researchers from other disciplines, such as philosophy, law, and social sciences, can further enrich the framework and ensure its relevance (Miller, 2019). This broader engagement offers several benefits, including refining the framework by addressing limitations (Mittelstadt, 2019), promoting cross-cultural understanding (Al-Rodhan, 2020), and building consensus on AI ethics (Rahwan, 2018).

7. Conclusion

7.1 Summary of Key Findings

This research has explored the potential of Islamic ethics to inform the development of accountable and ethical AI systems. By drawing inspiration from Islamic concepts of thought, intention, and action, and by leveraging blockchain technology for secure record-keeping, the proposed "Raqib and Atid" framework offers a novel approach to AI accountability. The framework emphasizes the importance of monitoring not only the actions of an AI system but also its internal processes, akin to "thoughts" and "intentions," to ensure ethical behavior throughout its operational life cycle. This comprehensive approach addresses the limitations of outcome-oriented AI ethics frameworks and promotes a more nuanced and proactive approach to ethical assessment.

7.2 Contributions to AI Ethics

The "Raqib and Atid" framework makes several significant contributions to the field of AI ethics. Firstly, it introduces a culturally grounded perspective, demonstrating how religious and cultural values can inform the development of ethical AI systems. This approach encourages a broader dialogue on AI ethics, incorporating diverse perspectives and promoting cross-cultural understanding. Secondly, the framework highlights the importance of transparency and explainability in AI systems. By employing interpretability techniques and blockchain technology, the framework ensures that the AI's decision-making processes are transparent and auditable, fostering trust and accountability. Thirdly, the framework promotes a more proactive approach to AI ethics, focusing on preventing ethical violations before they occur. By monitoring the AI's internal processes and providing feedback for improvement, the framework encourages continuous learning and ethical development. Finally, the framework's emphasis on accountability and responsibility aligns with the growing demand for ethical AI governance. By

ensuring that AI systems are held responsible for their actions, the framework promotes the development of AI that serves humanity and contributes to a just and equitable society.

7.3 Future Directions

Future research should focus on further developing the "Raqib and Atid" framework and addressing the challenges and considerations identified in this paper. This includes refining the definition of AI intentions by developing more robust methods for interpreting AI's internal representations and inferring its goals and motivations. Additionally, practical strategies need to be developed to balance autonomy and monitoring, allowing for effective ethical oversight without unduly restricting the AI's ability to learn and adapt. Ensuring data privacy and security is crucial, requiring the implementation of robust security measures and encryption techniques to protect the integrity of the data stored on the blockchain. Furthermore, the scalability and storage challenges associated with blockchain technology need to be addressed to ensure the feasibility and long-term sustainability of the framework. Developing clear and comprehensive ethical guidelines grounded in Islamic ethical principles, while also evolving alongside advancements in AI technology, is essential for the framework's effectiveness. Moreover, promoting cultural sensitivity and inclusivity is vital to ensure the framework's global relevance and avoid imposing a single interpretation of Islamic ethics. Finally, improving the explainability of the framework to various stakeholders, by translating technical insights into clear and accessible explanations, will foster trust and facilitate informed decision-making. By addressing these challenges and continuing to refine the "Raqib and Atid" framework, we can help ensure that AI is developed and used in a responsible and ethical manner that benefits all of humanity

8. Reference

Anderson, M., & Anderson, S. L. (2011). Machine ethics: Creating an ethical intelligent agent. AI Magazine, 32(4), 13-26. https://doi.org/10.1609/aimag.v32i4.2375

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973-989.

Antonopoulos, A. M. (2017). Mastering Bitcoin: Programming the open blockchain. O'Reilly Media.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Al-Attas, S. M. N. (1993). Islam and secularism. International Institute of Islamic Thought and Civilization (ISTAC).

Al-Bukhari, M. I. (1997). Sahih al-Bukhari (M. M. Khan, Trans.). Darussalam.

Al-Faruqi, I. R. (1982). Islamization of knowledge: General principles and work plan. International Institute of Islamic Thought.

Al-Faruqi, I. R. (1992). Al-Tawhid: Its implications for thought and life. International Institute of Islamic Thought.

Al-Ghazali, A. H. (2003). *The ninety-nine beautiful names of God: Al-Maqsad al-Asna fi Sharh Asma' Allah al-Husna*. Islamic Texts Society.

Al-Mawardi, A. (1996). The ordinances of government: Al-Ahkam al-Sultaniyya w'al-Wilayat al-Diniyya (W. H. Wahba, Trans.). Garnet Publishing.

Al-Nawawi, Y. (1999). Riyad al-Salihin: The gardens of the righteous (M. Z. Khan, Trans.). Dar-us-Salam Publications.

Al-Qaradawi, Y. (2007). The lawful and the prohibited in Islam. Islamic Book Trust.

Al-Qurtubi, A. (2003). Al-Jami' li-Ahkam al-Qur'an (A. Bewley, Trans.). Dar Al Taqwa.

Al-Rodhan, N. R. (2019). The role of ethics in the age of disruptive technologies. Journal of Futures Studies, 23(4), 41-56.

Al-Rodhan, N. R. (2020). Sustainable history and the dignity of man: A philosophy of history and civilizational triumph. LIT Verlag.

Al-Suyuti, J. (2008). The perfect guide to the sciences of the Qur'an: Al-Itqan fi 'Ulum al-Qur'an (M. A. Haleem, Trans.). Garnet Publishing.

Al-Tabari, M. I. J. (1987). The history of al-Tabari: Volume 1, general introduction and from the creation to the flood (F. Rosenthal, Trans.). State University of New York Press.

Atzori, M. (2017). Blockchain technology and decentralized governance: Is the state still necessary? Journal of Governance and Regulation, 6(1), 45-62. https://doi.org/10.22495/jgr_v6_i1_p5

Badar, M. S. (2018). Islamic perspectives on artificial intelligence. In The Oxford Handbook of Islamic Ethics and Law (pp. 806-823). Oxford University Press.

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

Beck, R., Avital, M., Rossi, M., & Thatcher, J. B. (2017). Blockchain technology in business & information systems research. Business & Information Systems Engineering, 59(6), 381-384. https://doi.org/10.1007/s12599-017-0505-1

Boddington, P. (2017). Towards a code of ethics for artificial intelligence. Springer.

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Eds.), The Cambridge handbook of artificial intelligence (pp. 316-334). Cambridge University Press.

Bratman, M. E. (1987). Intentions, plans, and practical reason. Harvard University Press.

Bryson, J. J. (2018). Patiency is not a virtue: The design of intelligent systems and systems of ethics. Ethics and Information Technology, 20(1), 15-26. https://doi.org/10.1007/s10676-018-9448-6

Bryson, J. J. (2020). The artificial intelligence of the ethics of artificial intelligence: An introductory overview for law and regulation. In M. Dubber, F. Pasquale, & S. Das (Eds.), The Oxford handbook of ethics of AI (pp. 3-25). Oxford University Press.

Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. University of Chicago Law Review, 85(1), 1-57. https://doi.org/10.2139/ssrn.3015350

Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification, and open issues. Telematics and Informatics, 36, 55-81. https://doi.org/10.1016/j.tele.2018.11.006

Cath, C. (2018). Governing artificial intelligence: Ethical, legal, and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, & UK approach. Science and Engineering Ethics, 24(2), 505-528. https://doi.org/10.1007/s11948-017-9901-7

Catalini, C., & Gans, J. S. (2016). Some simple economics of the blockchain. MIT Sloan Research Paper No. 5191-16. https://doi.org/10.2139/ssrn.2874598

Croman, K., Decker, C., Eyal, I., Gencer, A. E., Juels, A., Kosba, A., & Gün Sirer, E. (2016, May). On scaling decentralized blockchains. In International conference on financial cryptography and data security (pp. 106-125). Springer, Berlin, Heidelberg.

De Filippi, P., & Hassan, S. (2018). Blockchain technology as a regulatory technology: From code is law to law is code. First Monday, 21(12). https://doi.org/10.5210/fm.v21i12.7113

De Filippi, P., & Wright, A. (2018). Blockchain and the law: The rule of code. Harvard University Press.

Dennett, D. C. (1987). The intentional stance. MIT press.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

Elish, M. C., & Boyd, D. (2018). Situating methods in the magic of Big Data and AI. Communication Monographs, 85(1), 57-80. https://doi.org/10.1080/03637751.2017.1375130

Ess, C. (2020). Digital media ethics (3rd ed.). Polity Press.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Floridi, L. (2014). The fourth revolution: How the infosphere is reshaping human reality. Oxford University Press.

Floridi, L. (2019). What the near future of artificial intelligence could be. Philosophy & Technology, 32(1), 1-15. https://doi.org/10.1007/s13347-019-00345-z

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42.

Gunkel, D. J. (2012). The machine question: Critical perspectives on AI, robots, and ethics. MIT Press.

Gunkel, D. J. (2018). Robot rights. MIT Press.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99-120.

Hashim, R. (2018). Islamic ethics and the implications for business. Springer.

Ibn Kathir, I. (2000). Tafsir Ibn Kathir (S. Al-Mubarakpuri, Trans.). Darussalam.

Ibn Qayyim al-Jawziyya. (2003). The spiritual disease and its cure. Dar Al Kotob Al Ilmiyah.

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

Kamali, M. H. (2002). The dignity of man: An Islamic perspective. Islamic Texts Society.

Kamali, M. H. (2003). Principles of Islamic jurisprudence. Islamic Texts Society.

Kamali, M. H. (2008). Shari'ah law: An introduction. Oneworld Publications.

Kamali, M. H. (2011). Freedom, responsibility and ethics in Islamic law. Islamic Studies, 50(2), 229-248.

Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36-43. https://doi.org/10.1145/3233231

Lumbard, J. E. B. (2015). Islam, fundamentalism, and the betrayal of tradition: Essays by Western Muslim scholars. World Wisdom.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Interpretable machine learning: definitions, methods, and applications. Proceedings of the National Academy of Sciences, 116(44), 22071-22080.

Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). Bitcoin and cryptocurrency technologies: A comprehensive introduction. Princeton University Press.

Nasr, S. H. (2010). The heart of Islam: Enduring values for humanity. HarperOne.

Nissenbaum, H. (2010). Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press.

Nowostawski, M., & Purver, M. (2019). The ethical implications of blockchain technology. Journal of Business Ethics, 156(4), 1045-1058.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

Ølnes, S., Ubacht, J., & Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. Government Information Quarterly, 34(3), 355-364.

Pew Research Center. (2015, April 2). The Future of World Religions: Population Growth Projections, 2010-2050. Pew Research Center's Religion & Public Life Project. https://www.pewresearch.org/religion/2015/04/02/religious-projections-2010-2050/

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5-14. https://doi.org/10.1007/s10676-017-9430-8

Ramadan, T. (2007). In the footsteps of the Prophet: Lessons from the life of Muhammad. Oxford University Press.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Viking.

Sachedina, A. (2009). Islamic biomedical ethics: Principles and application. Oxford University Press.

Saeed, A. (2006). Islamic thought: An introduction. Routledge.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). Explainable AI: Interpreting, explaining and visualizing deep learning. Springer.

Schneier, B. (2015). Data and Goliath: The hidden battles to collect your data and control your world. W.W. Norton & Company.

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and brain sciences, 3(3), 417-424.

Solove, D. J. (2006). A taxonomy of privacy. University of Pennsylvania Law Review, 154(3), 477-560. https://doi.org/10.2307/40041279

Swan, M. (2015). Blockchain: Blueprint for a new economy. O'Reilly Media.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751-752. https://doi.org/10.1126/science.aat5991

Tapscott, D., & Tapscott, A. (2016). Blockchain revolution: How the technology behind Bitcoin is changing money, business, and the world. Penguin.

Tapscott, D., & Tapscott, A. (2017). Realizing the potential of blockchain: A multistakeholder approach to the stewardship of blockchain and cryptocurrencies. World Economic Forum.

Wallach, W., & Allen, C. (2008). Moral machines: Teaching robots right from wrong. Oxford University Press.

Werbach, K. (2018). The blockchain and the new architecture of trust. MIT Press.

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 195-200. https://doi.org/10.1145/3306618.3314289

Xu, X., Weber, I., Staples, M., Zhu, L., Bosch, J., Bass, L., & Rimba, P. (2017). A taxonomy of blockchain-based systems for architecture design. In 2017 IEEE International Conference on Software Architecture (ICSA) (pp. 243-252). IEEE. https://doi.org/10.1109/ICSA.2017.33

Zarsky, T. Z. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology, & Human Values, 41(1), 118-132. https://doi.org/10.1177/0162243915605575

Zikopoulos, P., & Eaton, C. (2011). Understanding big data: Analytics for enterprise class Hadoop and streaming data. McGraw-Hill Osborne Media.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.