Philosophy & Technology (2023) 36:73
https://doi.org/10.1007/s13347-023-00668-x

RESEARCH ARTICLE

Artificial Intelligence (AI) in Islamic Ethics: Towards Pluralist Ethical Benchmarking for AI

Ezieddin Elmahjub[1]

Received: 1 April 2023 / Accepted: 27 September 2023 / Published online: 1 November 2023

Abstract

This paper explores artificial intelligence (AI) ethics from an Islamic perspective at a critical time for AI ethical norm-setting. It advocates for a pluralist approach to ethical AI benchmarking. As rapid advancements in AI technologies pose challenges surrounding autonomy, privacy, fairness, and transparency, the prevailing ethical discourse has been predominantly Western or Eurocentric. To address this imbalance, this paper delves into the Islamic ethical traditions to develop a framework that contributes to the global debate on optimal norm setting for designing and using AI technologies.

The paper outlines Islamic parameters for ethical values and moral actions in the context of AI's ethical uncertainties. It emphasizes the significance of both textual and non-textual Islamic sources in addressing these uncertainties while placing a strong emphasis on the notion of "good" or "*maṣlaḥa*" as a normative guide for AI's ethical evaluation. Defining *maṣlaḥa* as an ethical state of affairs in harmony with divine will, the paper highlights the coexistence of two interpretations of *maṣlaḥa*: welfarist/utility-based and duty-based. Islamic jurisprudence allows for arguments supporting ethical choices that prioritize building the technical infrastructure for AI to maximize utility. Conversely, it also supports choices that reject consequential utility calculations as the sole measure of value in determining ethical responses to AI advancements.

Keywords Islamic ethics · AI · Maṣlaḥa · AI and human welfare · AI and fairness

1 Introduction

This paper introduces an Islamic ethical framework to respond to Artificial Intelligence (AI)'s ethical challenges. It draws from the rich Islamic ethical traditions to construct an Islamic vision of ethical value and ethical action to guide policy and

✉ Ezieddin Elmahjub
eelmahjub@qu.edu.qa

[1] College of Law, Qatar University, Doha, Qatar

regulatory benchmarking for AI design and use. It offers a critical assessment of the Western monopoly in norm creation for AI technologies, emphasizing the need to introduce multicultural/ comparative approaches to the ethical challenges associated with machine learning and autonomous machines.

AI technologies are already global with a rapidly increasing presence in different aspects of our life. The technologies come with transformative power and enormous beneficial applications in many sectors including security, healthcare, transportation, agriculture, education, commerce, and finance. Advances in machine learning and hardware design have made AI systems increasingly capable of autonomous behavior through sensing, planning and action, logical reasoning, decision support, predictive analytics, and operating land, air, and sea vehicles. This autonomous behavior is not always beneficial and can seriously harm society through biases, discrimination, loss of privacy, difficulty identifying liabilities, unemployment, and concentration of power and wealth in a few stakeholders.

Governments, the private sector and the research community worldwide seek to strike a balance between the risks and benefits of AI. There is a sweeping international movement to achieve that through ethical discourse and standard-setting to design ethical and policy guidelines for responsible design and use of AI technologies. However, Western ethical theories overwhelmingly dominate the global theoretical discussion on AI ethics. Moreover, most of the policy documents worldwide come from government agencies, civil society organizations, and private companies located in the West and are informed by Western ethical benchmarking.[1]

The paper aims to ground AI ethical uncertainties within Islamic normative discourse, laying the foundation for developing Islamic normative principles on AI. These principles will help determine what is morally right and appropriate when addressing AI's ethical and societal challenges. This contribution is a timely addition to the growing body of comparative AI ethics research. It will run at a critical time when conceptual and empirical research about pluralist views on AI ethics and policy is sorely needed.[2]

This study employs *uṣūl al-fiqh* (the principles of Islamic jurisprudence) to perform value alignment analysis for AI's ethical uncertainties and aims to develop solutions anchored in the Islamic worldview. Generally, *uṣūl al-fiqh* posits that the primary frame of reference for morality lies in the Qurʾān and the Hadith, or the recorded traditions of the Prophet (PBUH). Nevertheless, when these textual sources do not offer explicit solutions to specific ethical quandaries, *al-Masādir al-'Aqliyya*, reason-infused methodologies, can be utilized to ascertain morally appropriate choices within the Islamic perspective. These methodologies encompass legal analogy (*Qiyās*), juristic preference (*Istihsan*), presumption of continuity (Istishab), public welfare or interest (*Maṣlaḥa*), blocking the means to an evil end (*Ṣadd al-ḍharaiʿ*), and customary traditions (*Urf*). While this study elucidates the significance and application of all these methodologies to AI's ethical uncertainties,

[1] Jobin et al. (2019)

[2] IEEE Standards Association (2019)

it dedicates substantial focus to the concept of *maṣlaḥa*, primarily due to several reasons which are reinforced throughout this paper.

Firstly, the very existence of *maṣlaḥa* is intertwined with emergent ethical uncertainties for which there are no specific remedies in the Qurʾān or Hadith. Secondly, it is adaptable and allows the evaluator to perform an inductive and deductive review of the textual authorities to formulate a set of normative principles. These principles can then be applied to emergent ethical challenges to enhance societal welfare. Thirdly, *maṣlaḥa* embodies a harmonious synergy between textual sources and the necessity for rational inputs in Islamic norm formation. At its heart, *maṣlaḥa* is identified as *maqṣūd al-sharʿ* (the objective of divine revelation). Consequently, while it is rooted in textual sources, it also enables human reason to exercise substantial autonomy in conducting rational normative analyses to address emerging challenges, including those presented by AI. Lastly, *maṣlaḥa*'s versatility and openness to various interpretations from a contemporary ethical standpoint make it particularly valuable. It underpins a broad spectrum of foundational social responsibilities towards religion, life, intellect, and dignity, yet does not preclude strategies to optimize human welfare. The paper below demonstrates how *maṣlaḥa* can offer these versatile and relevant interpretations.

The paper raises important questions about the ethical values that an Islamic ethical framework would recognize in the context of AI, including whether the goal should be to maximize welfare for the majority or to prioritize the recognition of static intrinsic human values, even if this leads to decelerating certain AI technologies. Alternatively, could there be a possibility for a hybrid ethical benchmark that strives to achieve both aims?

Islamic jurisprudence exhibits heterogeneous views on the content of *maṣlaḥa*. We will see different positions on what would constitute an ethical AI. However, these positions can be broadly classified into utility-based and duty-based categories. Utility-based views would prioritize building and deploying the technical infrastructure of AI applications as long as they serve society's overall public interests, even if this comes at a marginal expense to fairness, transparency, and privacy among other principles. Conversely, duty-based views prioritize respect for intrinsic values, such as fairness, dignity, and human agency, over standard welfarist considerations. As such, any regulatory scheme must safeguard these values, regardless of average welfare consequences. The present paper aims to analyze and compare these different perspectives in Islamic jurisprudence, exploring their implications for the ethical design and governance of AI.

The utility and duty paradigms are well-established in modern normative theories, specifically consequentialism and deontology. The reasons for associating these concepts with the notion of *maṣlaḥa* in Islamic jurisprudence are twofold and merit clarification.

Firstly, the audience of this paper is diverse. Hence, it is beneficial to frame traditional Islamic ethical stances using more accessible moral terminologies. This approach ensures that the metaphysical beliefs held by Muslims are not compromised. Utilizing a common moral lexicon is crucial in pluralist contexts, especially if we aim to foster a convergence of various ethical doctrines toward a shared understanding of justice.[3] As philosopher Jürgen Habermas posits, while it is inappropriate to marginalize religious ethical perspectives in the public sphere, it is reasonable to expect religious individuals to articulate their ethical stances in a language that is accessible to a non-religious moral sensibility.[4]

Secondly, notions of utility and duty are already ingrained within Islamic ethical discourse, as delineated in Sects. (3: 3 and 4) of this paper. Numerous Islamic jurists and scholars have subtly and overtly linked the principle of *maṣlaḥa* to morality grounded in duty or utility. This paper identifies these contributions and contextualizes them for assessing AI technologies from an Islamic perspective. Importantly, this does not imply the abandonment of the uniqueness of Islamic thought. As this paper consistently demonstrates, *maṣlaḥa* is divinely sourced and metaphysically oriented. This exercise represents an attempt to comprehend an Islamic normative instrument (*maṣlaḥa*) through the lens of contemporary comparative philosophy.

2 AI in an Ethical Context

AI is a collection of software and hardware technologies capable of autonomous data collection, analysis and reasoning to perform tasks in both digital and physical domains without explicit guidance from a human operator.[5] The enormous potential benefit of AI technologies is well documented across all sectors.[6] However, increased integration of these autonomous systems into our societal infrastructure poses the risk of losing meaningful control over them, causing a range of societal harms to humans' sense of fairness, autonomy, dignity, privacy and safety.[7]

While the full scale of these harms is still difficult to define, several risk domains have been identified. These include areas of unintended misuse, including gender and racial discrimination, loss of privacy, damage, and difficulty identifying liabilities[8] as well as intentional abuse, including malicious use of deep fakes, political propaganda, fake news and cyberattacks.[9] A large body of AI and ethics scholarship focuses on the normative analysis of AI technologies to develop ethical and policy frameworks to leverage the benefits created by AI while ensuring efficient processes to attribute moral and legal responsibility for all forms of AI risks.[10]

The central technology that drives most AI capabilities is machine learning (ML), including systems powered by large troves of data such as deep learning, generative adversarial networks and reinforcement learning. ML systems are capable of collecting and labelling data, recognizing patterns, and digitizing the decision-making process, among other capabilities. The technologies promise to deliver efficient

[3] Cohen (1993)

[4] Habermas (2006)

[5] Dawson et al. (2019)

[6] Littman et al. (2022)

[7] Christian (2020)

[8] de Almeida et al. (2021)

[9] Benjamins and García (2020)

[10] Taddeo and Floridi (2018)

outcomes in the form of increased precision, scale and speed in decision-making to provide answers to a wide range of questions ranging from "is this a cancer?", "Will this person reappear in court?" to "What should we do next?"[11] The ability to handle data and make inferences is a form of behavior that could have many serious moral and legal consequences.

Since the 1990s, scholars have been expressing concerns about the likely negative impact of machine learning on the social sense of fairness through automating discrimination and reinforcing existing social biases.[12] However, recent years have witnessed large-scale implementations of ML systems across all social, political and economic domains due to the availability of large data sets, better algorithms, and increased connectivity. Currently, there is a large literature that documents unprecedented privacy risks, social biases, and harms caused by ML systems concerning historically disadvantaged segments of society in areas ranging from privacy and surveillance,[13] facial analysis,[14] online ads coverage,[15] search engine discrimination,[16] employment opportunities[17] and law enforcement.[18]

Several high-profile case studies enforce the ethical concerns around using AI applications across different sectors of society. For instance, in the Cambridge Analytica scandal, ML technology was used to sniff through, and analyze large data sets of millions of Facebook users without prior consent, and use this data to design and sell misleading political ads to sway public opinions.[19] Amazon's hiring algorithm had been discontinued because it was coming up with decisions to hire more men than women.[20] In the U.S, teachers challenged an ML application used to assess teaching performance and make recommendations to dismiss teachers without explanation because of the proprietary software.[21] In the U.S as well, a sentencing and probation assessment algorithm was found to incorrectly label black people as being more likely to repeat violent offences than white people.[22] Finally, since 2014, researchers raised concerns about social networks' ability to use AI systems to manipulate the mood and perceptions of their users. Algorithms can filter users' feeds and potentially use data generated in the process to influence users' attitudes and increase the effectiveness of targeted advertising.[23]

[11] World Economic Forum (2018)

[12] Friedman and Nissenbaum (1996)

[13] Belk (2021)

[14] Buolamwini and Gebru (2018)

[15] Sweeney (2013)

[16] Noble (2018)

[17] Chen et al. (2018)

[18] Angwin et al. (2016)

[19] Rosenberg et al. (2018)

[20] Dastin (2018)

[21] Langford (2017)

[22] Mayson (2019)

[23] Kramer et al. (2014)

2.1 Ethical and Normative Responses to AI's Ethical Uncertainties

In response to AI promises and risks, many stakeholders across governments, private companies, and academia have undertaken various studies to align AI systems with the ethical and normative frameworks of society. The central objective of AI's normative analysis is to define ethical imperatives for creating and using these systems.[24] These studies can be traced back to two overarching evaluative frameworks. The first one is rooted in classical Western ethics seeking to define ethical values and normative statements to guide AI's ethics through major ethical theories, including consequentialism,[25] deontology[26] and virtue ethics.[27] The second framework comes in the form of general pragmatic normative principles that governments, organizations and companies seek to deploy to guide the design and operation of AI technologies. However, these principles are much more influential in AI's ethical discourse since most of them come from the creators and users of AI technologies. For instance, an extensive review of 84 ethical guidelines issued by governments, organizations and companies in the developed world demonstrated that the most common normative principles proposed to inform ethical and responsible AI are: transparency, justice and fairness, nonmaleficence, responsibility and privacy.[28]

If we were to engage in a deeper theoretical assessment of the ethical terms of reference used to evaluate and justify AI ethics in the existing ethical framework we would come across foundational ambiguities around the nature of the ethical value that we should promote in making a moral judgment regarding AI applications. For instance, those who appeal to utilitarian arguments for AI fail to define what form of utilitarianism they are applying. Increased efficiency is not the only objective of utilitarian/ consequentialist analysis.

Utilitarianism, as a moral theory, posits that some form of intrinsic good exists and that this good ought to be maximized. However, there are significant debates among utilitarians regarding the nature of the intrinsic good affirmed at the metaethical level. Some scholars propose a hedonistic theory of value, which holds that pleasure and happiness are the principal possessors of intrinsic value. This view is commonly associated with classical utilitarians such as Jeremy Bentham, John Stuart Mill, and Henry Sidgwick.[29] Others propose a non-hedonistic theory of value that asserts that the intrinsic good should not be determined based on a pleasurable state of affairs but rather on an objective ideal value, such as knowledge, virtue, or beauty. This perspective is known as ideal utilitarianism and can be found in the work of scholars such as G. E. Moore[30] and Hastings Rashdall's work.[31]

[24] IEEE Standards Association (2019)

[25] Bench-Capon (2020)

[26] Ulgen (2017)

[27] Berberich and Diepold (2018)

[28] Jobin et al. (2019)

[29] Quinton (1973)

[30] Moore (1988)

[31] Rashdall (1907)

Furthermore, normative implications within the utilitarian framework may vary based on whether one supports ethical egoism or altruism. Ethical egoists argue that actions promoting individual interests are morally correct, while ethical altruists maintain that morally correct actions result in positive outcomes for the majority. Given these differences in normative implications, it is worth questioning the viability of a "one size fits all" standard for AI settings. A thorough examination of the ethical terms of reference employed to assess and justify AI ethics within existing frameworks exposes fundamental uncertainties regarding the ethical value that should guide moral judgments of AI applications. Resolving these ambiguities is essential for fostering AI ethics that align with societal norms and ethical frameworks.

Even if we were to prioritize more pragmatic, implementable normative standards, such as fairness, accountability, and transparency, rather than engage in complex philosophical analyses of AI ethics, we would still encounter conceptual challenges. A key issue is how to define the content of these normative values. For instance, how do we define the right thing to do when measuring algorithmic fairness? Is it to maximize overall efficiency in decision support systems while accepting marginal discrimination or harm to a few people? Or should we prioritize protecting the privacy, dignity, and equal opportunity of each subject of the decision at all costs? In other words, when competing societal interests exist, which normative value should we prioritize according to Western ethics? Should overall efficiency and interests be the first-order principle, with the right against discrimination ranking second?

Those ethical principles/ standards do not represent a pluralist vision of ethical norm-making. From 2015 to 2020, a total of 117 AI policy documents were published by governments, organizations and private companies in North America, Europe and developed economies in Central Asia.[32] There is a noticeable absence of comparative ethical inputs from other ethical systems. Moreover, the existing body of research on major AI domains such as ethical programming of autonomous vehicles (AVs) or fairness and accountability in algorithms is conducted in Western normative settings, aiming to address historical injustices prevalent in the Western context, using data mostly mined from Western sources.[33] IEEE warns against the Western Ethical monopoly of AI ethics and suggests that "There is an urgent need to broaden traditional ethics in its contemporary form of "responsible innovation" (RI) beyond the scope of "Western" ethical foundations".[34]

2.2 Critical Assessment of Western Monopoly of AI Ethics

AI applications are global. Their benefits and harms are being experienced in almost every corner of the world. However, the ethical benchmarking for these transformative technologies is not yet global. For instance, in the ACM FAccT Conference on

[32] Stanford (2021)

[33] Sambasivan et al. (2021)

[34] IEEE Standards Association (2019)

AI ethics, out of 138 papers published in 2019 and 2020, a few would even refer to comparative ethics.[35]

There is undeniable Western dominance in defining ethical AI. The Western ethical traditions and normative environment dictate most of the terms of reference in AI's ethical discourse. There are growing global calls to introduce multicultural and comparative ethical perspectives to solve AI systems' ethical uncertainties.[36] The Global AI Initiative of IEEE recommends the incorporation of classical Buddhist, Ubuntu and Shinto ethical traditions into the current discourse on AI ethics and policy while omitting any reference to Abrahamic religions.[37] Such calls align with the study of comparative ethics, highlighting the variations in AI's ethical uncertainties and normative assumptions about what constitutes good and evil, right and wrong across diverse ethical traditions. This approach to AI ethics is crucial in promoting a nuanced understanding of the ethical implications of AI and its alignment with diverse societal norms.

While ethics often intersect, non-Western perspectives remain essential in AI development. "AI systems built for Western values, with Western tradeoffs, [might] violate other values".[38] Scholars criticize the dominance of the classic techniques of Western colonialism in the technical and conceptual architecture of AI. This dominance runs very deep from the very definition of "intelligence"[39] to data extraction and resale to developing communities,[40] to the entire frame of reference for the terms used to discuss ethical AI in normative and policy contexts.[41]

2.3 AI in the Islamic World

There is massive interest in many countries with predominantly Muslim populations to localize and promote the integration of AI technologies into their societal infrastructure. For instance, in 2017, the Government of Saudi Arabia announced its decision to grant citizenship to the Sophia robot as the world's first 'robotic citizen'. Other Gulf states invested largely in building smart cities operated by AI applications.[42]

From 2017 to 2021, countries in the Middle East and North African (MENA) region published numerous documents outlining strategies to leverage AI for economic growth, security, education, health, and transportation, among other areas. These documents consistently demonstrate a policy priority centered on developing the technical infrastructure for AI. However, the strategies' commitment to incorporating an ethical and normative component varies. For instance, the AI strategy

[35] Sambasivan et al. (2021)

[36] Wong (2016)

[37] IEEE Standards Association (2019)

[38] Stanford (2021)

[39] Adams (2021)

[40] Abeba Birhane (2020); Mohamed et al. (2020)

[41] Mhlambi (2020)

[42] Chaudhary (2020)

documents of the UAE (2017)[43] and Saudi Arabia (2020)[44] pledge to make policy and legislative reforms to welcome AI technologies with no mention of local norms and values as a benchmark to determine the ethical and normative content of these AI strategies. By contrast, AI strategy documents of Qatar (2019) and Egypt (2021)[45] place greater importance on ensuring the overall alignment of AI technical policy and local notions of welfare and ethics.

Interestingly, the Qatari AI strategy stands out in stressing the importance of the local vision of AI ethics, stating that "*the [AI] framework to be developed must be consistent with both Qatari social, cultural, and religious norms*".[46] However, it should be noted that these documents are aspirational in nature and do not contain adequate ethical benchmarking for the integration and deployment of AI technologies. There seems to be an initial tendency to replicate the normative principles found in Western AI strategies such as fairness, accountability and transparency. The Qatari National AI Strategy's authors also recommend using the EU General Data Protection Regulation (GDPR) as a template to introduce local guidelines for AI applications.

The big challenge for countries in the Islamic World is to build AI systems that are aligned with their religious and cultural beliefs. Aligning AI systems with religious and cultural beliefs will ensure that these systems are more acceptable to the population. This is important because acceptance is a critical factor for successfully implementing any new technology. If AI systems are not aligned with religious and cultural beliefs, they may be perceived as a threat to local values and traditions.

Moreover, there is very good reason to avoid uncritical acceptance and transplantation of foreign normative principles while neglecting local norms, values and realities. For instance, the private sector produces a large number of the existing comparative policy documents. The primary normative value for private companies is maximizing profit. The involvement of private companies in AI standard-setting has been widely criticized for potentially relying on their power to produce high-level soft policy guidelines with a heavy technical component to transform the social and ethical challenges of AI into merely technical problems.[47] Or to avoid serious government regulation altogether.[48] Accordingly, there is a good reason to optimize AI standard setting within the local context, bearing local norms and challenges in mind.

[43] UAE Strategy for Artificial Intelligence (2017)

[44] Realizing Our Best Tomorrow: Strategy Narrative (2020)

[45] Egypt's National Artificial Intelligence (2021)

[46] Qatar's Ministry of Transportation and Communication (2019)

[47] Greene et al. (2019)

[48] Wagner (2018)

3 Islamic Approaches to Evaluating Ethical Implications of Technology

Scholars have been studying the connection between Islam and technology since the 1980s. Ziauddin Sardar was an early pioneer in this field. He argued that Islamic traditions, based on textual sources, should be used to evaluate the impact of technology in Muslim societies. Sardar cautioned against unquestioningly embracing modern technology without considering Islamic norms.[49]

Other scholars have also stressed the significance of establishing an Islamic ethical framework for information ethics. For instance, Salam Abdallah emphasized the necessity of utilizing classical sources of shariʿa, including the Qurʾān, Sunnah, Ijma', and Qiyās, to examine ethical and normative concerns in the realm of information technology.[50] In subsequent work, Abdallah put forth a framework for analyzing ethical challenges in the information technology field, guided by these sources.[51]

Amana Raquib also sought to introduce a general Islamic techno-ethical structure for technological growth. Raquib (2015, 2016) drew on major normative values emphasized in Islamic traditions such as justice, compassion, and balance to develop a comprehensive framework for assessing the ethical implications of technological developments from an Islamic perspective.[52]

Scholars increasingly turn to Islamic virtue ethics to address ethical uncertainties stemming from various AI applications. Noteworthy contributors to this discourse include Raquib, Channa, Zubair, and Qadir. These scholars critique the presumption of inherent goodness in unchecked AI development, scrutinizing the ethical consequences of AI progress. They also question the appropriateness of current market logics and business models governing the tech industry. Their argument posits Islamic virtue ethics as a comprehensive and valuable alternative to the existing ethical frameworks governing AI. They suggest that virtue ethics, with its emphasis on cultivating good character traits, can better address the intricate ethical challenges AI presents compared to the more rule-based approaches dominant in the West. The virtues of kindness, charity, forgiveness, honesty, patience, justice, and respect for others, they argue, are integral to the ethical development of AI.[53]

This paper takes a different methodological approach. While recognizing the significance of virtue ethics in shaping the ethical trajectory of AI, it leans toward an act-centered approach to morality, as opposed to the agent-centered perspective of virtue ethics.[54] Established normative theories such as consequentialism and deontology, or hybrid versions thereof, can offer clearer action-guidance. This is crucial in the AI context, where specific, actionable rules and principles are needed for

[49] Sardar (1988)

[50] Abdallah (2008)

[51] Abdallah (2010)

[52] Raquib (2015); Raquib (2016)

[53] Raquib et al. (2022)

[54] Slote (2001); Zagzebski (2004)

designing, deploying, and managing AI systems. However, this does not imply that virtue ethics and other normative theories are mutually exclusive within AI research. It is entirely plausible for some to believe that in designing, deploying, and using AI, we should ask, ’What sort of person should I be?’ in line with virtue ethics. Yet, it would not be a misstep for others to solely ask, ’What should I do?’.

3.1 AI and Islamic Sources of Normative Ethics

Muslims address ethical uncertainties by deriving Islamic moral judgments (*ḥukm al-sharʿi*) from established principles of Islamic jurisprudence (*uṣūl al-fiqh*). For example, when evaluating responses to bias or opacity in algorithms or advocating for privacy rights from an Islamic standpoint, one should consult the Qurʾān and Prophet’s recorded traditions as primary sources of moral guidance.[55] These sources offer broad normative principles, that might support arguments for fairness, transparency, and privacy as morally commendable, while condemning bias, opacity, and privacy violations.

However, it is important to acknowledge that these sources may not provide detailed guidance for modern challenges such as those posed by AI. *Uṣūl al-fiqh* recognizes that texts are limited, and emerging issues are limitless (*al-nusūs mutanāhiyya wa al-waqaeiʿ ghaiyru mutanāhiyya*).[56]

Muslims often turn to a mufti for guidance when facing moral questions or dilemmas. The *mufti* issues a fatwa (religious opinion) to address that moral question or dilemma according to Islamic sources.[57] With the rise of AI technology, it is conceivable to see applications of *fatwas* to several domains of AI, such as determining the morally required choices in designing AV crash algorithms[58] or deciding whether to deploy machine-learning algorithms that exhibit marginal racial or gender bias, but promote overall security and law enforcement.

Muftis typically looks for responses in textual sources. When needed, they turn to the rational sources of Islamic jurisprudence (*masādir ʿaqliyya*), like legal analogy or general normative analysis, to identify the interests they should safeguard. In modern times, muftis increasingly rely on rational input to address emerging questions arising from social, economic, and technological shifts. To determine the morality of an act, *muftis* argue that Islamic sources support choices that promote societal interests and prevent harm, as long as it does not violate specific textual prohibitions, such as those against murder, adultery, or usury.

[55] Wahba al-Zuhayli (1986)

[56] Rahṃān (1965)

[57] Hendrickson (2013)

[58] Elmahjub and Qadir (2023)

3.2 *Maṣlaḥa* as Benchmark for AI Ethics

Islamic normative frameworks have well-established notions of social good, public interest and human welfare. The momentum surrounding public interest as the primary objective of Islamic texts can be traced back to the eleventh century, notably in the works of scholars such as Imām al-Ḥaramayn al-Juwaynī (d.1085) and Abū Ḥāmid al-Ghazālī (d.1111). For centuries, Muslim jurists have diligently sought to establish a robust ethical framework to discriminate between good and evil, right and wrong, thereby guiding human conduct amidst ever-changing contexts. As Islamic jurisprudence evolved, these scholars acknowledged the necessity to decode the normative language of textual sources, addressing challenges not explicitly delineated in revelation. This recognition culminated in the birth of a specialized branch of Islamic jurisprudence, termed *maqāṣid al-sharīʿa*, interpreted as the objectives of Islamic revelation.[59]

This approach maintains that the ultimate purpose of divine order is to serve human interests, known as *maṣlaḥa*, for the benefit of humanity. Textual sources often embody overarching principles and purposes (*hikam*) aimed at promoting and nurturing societal well-being. As stated by al-ʿIzz b. ʿAbd al-Salām (d.1261), the objective of Islamic texts is to ensure the social good of people either through averting potential harm or bringing about benefits.[60] This commitment to the common good is evident in consistently promoting well-being and preventing harm throughout the scriptures.

The notion of *maṣlaḥa* holds significant relevance for the normative analysis of AI. It is a flexible concept frequently invoked to seek moral judgments for issues not explicitly addressed in Islamic textual sources. *Maṣlaḥa* can function as a comprehensive ethical theory. Its base is rooted in *maqāṣid al-sharīʿa* and serves to balance the potential harms and benefits of emerging ethical and legal challenges.

It is important to note that *maṣlaḥa* is just one of several sources used by Muslim jurists in shaping Islamic jurisprudence. Different schools have developed and refined the principles (*uṣūl*) of Islamic jurisprudence. These principles serve as the basis from which ethicists derive moral judgments. Many of these sources hold relevance to the ethical dilemmas presented by AI.

Textual sources, for instance, emphasize fairness, privacy, and honesty. These values are pertinent to the governance of AI technologies, as they guide the establishment of equitable algorithmic decision-making processes, promote privacy in data collection and utilization, and prohibit the harmful use of AI technologies against humans or other living entities.

Non-textual sources also play a significant role in assessing AI ethics from an Islamic perspective. One such principle is "Blocking the Means" (*Ṣadd al-ḍharaiʿ*), which is a preemptive measure employed by jurists to avert actions that might potentially result in harm or wrongdoing, even if the actions themselves may not be

[59] ʿal-Fāsī (1963)

[60] ʿIzz al-Dīn ibn ʿAbd al-Salā m (1991)

deemed immediately harmful.[61] The principle of *Ṣadd al-ḍharai*ʿ allows for assessing the potential social and economic ramifications of specific AI technologies. This principle can be applied to evaluate specific AI technologies' societal and economic impact, like deepfakes, to curb misinformation, disinformation, political manipulation, reputation damage, or trust erosion. Amana Raquib and colleagues suggest that *Ṣadd al-ḍharai*ʿ could influence decision support systems in criminal justice, such as those for recidivism-risk scoring, to prevent potential miscarriages of justice and irreversible harm.[62]

Maṣlaḥa, which is often translated as public interest or public welfare, is a central concept in modern scholarship on Islamic studies.[63] The notion of *maṣlaḥa* posits that the underlying objective of the instructions, injunctions, and prohibitions found in the Qurʾānic texts is to promote choices that bring about good (*jalb al-manfaʿa*) and prevent harm (*dafʿ al-ḍarar*).[64] Felicitas Opwis identifies a recurring theme among classic scholars associating *maṣlaḥa* with promoting well-being, benefit, and goodness, and avoiding harm and evil.[65]

While acknowledging the correlation, it is essential not to exclusively define *maṣlaḥa* with public interests or welfare as known in contemporary social or economic sciences. Such an approach oversimplifies *maṣlaḥa's* nuanced application in Islamic normative analysis. Public interest or welfare often relates to material gain in a secular context, which does not fully capture the depth of *maṣlaḥa*. It is more appropriate to view *maṣlaḥa* as a state of affairs adhering to ethical standards in consonance with divine will.[66] This perspective encompasses, and goes beyond, conventional welfare metrics, allowing for a deeper exploration of the diverse ethical challenges posed by AI.

As far as AI is concerned, *maṣlaḥa* will be used as an evaluative framework to assess the compatibility of AI with Islamic notion of good (*ḥasan*) and evil (*qabīḥ*) and right (*ḥaqq*) and wrong (*batīl*). It should inform our understanding of major concepts in AI such as the content, limit and scope of fairness, transparency, accountability and privacy. However, the essence of *maṣlaḥa* is a matter of debate. Should it prioritize choices that maximize overall human welfare through technological and economic development or promote intrinsic human values regardless of utility and welfare calculations? Could a possible avenue exist to introduce a hybrid ethical standard that effectively promotes both objectives?

3.3 AI, Utility and Welfare Metrics in Islamic Ethics

Part of modern Islamic studies views *maṣlaḥa* as utility maximization construct. Scholars such as George Harouni, suggest that the Muʿtazila may have developed

[61] Kamali (2003)

[62] Raquib et al. (2022)

[63] Opwis (2010)

[64] al-Raysuni and al-Shāṭibī's (2005)

[65] Opwis (2010)

[66] Elmahjub (2021)

a utilitarian type of ethics that closely resemble classic Benthamite utilitarianism.[67] Sari Nusseibeh describes Al-Ghazālī's theory of *maṣlaḥa* as a utilitarian version of a consequentialist theory of moral action.[68] Andrew March argues that conceptions of *maṣlaḥa* are a prime example of consequentialist-utilitarian reasoning.[69] In the nineteenth and twentieth centuries, reformers attempting to modernize Islamic law also appealed to value calculations based on *maṣlaḥa*, which notably included utility as a criterion, according to Kerr[70] and Hallaq.[71] Islamic jurisprudence stresses maximizing the common good in decision-making. In this ethical approach, any AI framework must prioritize societal well-being and benefit the majority of people.

The modern Islamic reform movement strongly supports the utility-based interpretation of *maṣlaḥa*, which advocates for a broad range of social, legal, economic, and technological changes aimed at improving conditions within Muslim societies and adapting to modernityReformers advocate a progressive approach to deriving moral knowledge from classical Islamic sources. For example, Muḥammad ʿAbduh (d. 1905) emphasized rational norm creation rather than traditional interpretation of Islamic texts. He argued that if revelation acknowledges the intellect's capacity to discover the divine plan and take responsibility for human actions in this life and the afterlife, then reason must be in harmony with revelation.[72] ʿAbduh believed that a rational approach to ethics would help Islam address the challenges of the nineteenth and twentieth centuries.[73] He viewed moral knowledge as practical and empirical, advocating for rational evaluation of moral choices to maximize good and human well-being.[74] He supported his position with Qurʾānic verses emphasizing human needs' importance over acts of devotion. By doing so, ʿAbduh sought to demonstrate that revelation recognized the intrinsic value of human well-being and the centrality of human needs to moral reasoning.[75] This approach resonates with contemporary welfarist views of AI, promoting actions that extract material value for human societies from AI's technical and legal infrastructure.[76]

Like ʿAbduh, Muḥammad Rashīd Riḍā (d. 1935) advocated for a utility/ welfarist approach to Islamic ethics. He believed that Islamic jurisprudence should justify norms promoting the welfare of Muslim societies. Similar to reformers, he argued that revelation was mainly for acts of worship, allowing Muslims to use their intellect to create norms for ethical questions related to worldly matters and human needs.

Riḍā emphasized the extensive use of *maṣlaḥa* to address ethical questions lacking specific revelatory norms. He regarded the happiness and welfare of Muslims as

[67] Hourani (1960)
[68] Nusseibeh (2017)
[69] March (2009)
[70] Kerr (1966)
[71] Hallaq (2009)
[72] Muḥammad ʿAbduh (1972)
[73] ʿAbduh (n 82) vol 3, 359–63.
[74] Kerr (1966)
[75] Muḥammad ʿAbduh (1988)
[76] Gupta et al. (2021)

the ultimate goal of moral reasoning, asserting that the right choice maximizes their well-being. He also proposed re-examining existing interpretations and applications of textual sources to develop a welfarist vision aligned with contemporary human needs.[77]

Riḍā believed that Muslims should prioritize their material interests in most matters, except acts of worship.[78] He applied a welfarist ethical framework to various issues in his published *fatwas* from 1903 to 1935.[79] He believed that when making moral choices in mundane matters, we must weigh the expected outcomes of action and inaction to promote the well-being of the average Muslim. Riḍā argued that when making moral choices in everyday life, we should weigh the expected outcomes to promote the well-being of the average Muslim. He also challenged traditional interpretations of Islamic texts, permitting practices like photography and medical use of alcohol when they could bring benefits, such as verifying identity or saving lives.[80]

The welfarist view of Islamic ethical inquiry championed by ʿAbduh and Riḍā has gained renewed attention from contemporary scholars of Islamic jurisprudence. Muhammad Abū Zahra (d.1974) saw a strong connection between normative reasoning in Islamic legal theory and the utilitarian ethics of Mill and Bentham. He maintained that the utilitarian doctrine, known as *madhab al-manfaʿa*, necessitated lawmaking in contemporary societies to maximize the overall welfare of the greatest number of people. Abū Zahra argued that a social system that endeavoured to attain as much material and spiritual well-being as feasible for the greatest number of individuals could be deemed compatible with the principles enshrined in the Qurʾān through a process of induction.[81]

Yūsuf al-Qaraḍāwī (d. 2022) proposed a new strand in Islamic jurisprudence, known as *fiqh al-muwāzanāt*, (jurisprudence of calculations). He argued that, in worldly affairs, the intellect could independently define ethical values and generate moral knowledge about right and wrong. He believed there are varying degrees of goodness and evilness, and the right course of action involves weighing the expected consequences to maximize good and minimize harm.[82] Al-Qaraḍāwī cited traditional Islamic jurisprudential maxims, like tolerating a lesser harm to prevent a greater one and prioritizing the group's rights over individual rights, to support his claim that Islamic ethical reasoning aims for the overall welfare of the majority in a society.[83]

The welfarist perspective finds strong support in mainstream Islamic jurisprudence. Fakhr al-Dīn al-Rāzī (d.1210), for instance, suggested in his work *al-Maḥṣūl fī ʿilm uṣūl al-fiqh* that revelatory norms were justified through *ratio*. Al-Rāzī

[77] Riḍā (n.d.)

[78] Elmahjub (2021)

[79] al-Munajjid and al-Khūrī (1970)

[80] ibid, see fatwa 685 (1926) vol 5, 1873 and fatwa 201 (1906) vol 2, 627

[81] Abū Zahra (n.d.)

[82] al-Qaraḍāwī (1996)

[83] al-Qaraḍāwī (2011)

defined *ratio* as that which is agreeable to human nature (*munāsiba*), interpreted to mean that the moral agent would acquire some benefit (*manfaʿa*) and "be spared harm" (*mafsada*). He resorted to utility-based calculations by defining *manfaʿa* as pleasure (*ladhdah*) and *mafsada* as pain (*alam*), both of which are perceptible by human senses.[84] Al-Rāzī thus suggested that a consequential evaluation of human conduct is necessary to determine the appropriate course of action in a given situation If an action produces more good than harm, it becomes obligatory; otherwise, it should be abandoned.[85] Likewise, al-ʿIzz ibn ʿAbd al-Salām (d. 1261) explains *maṣlaḥa* in consequentialist terms. He argues that revelation exists to safeguard the interest of humankind and that the content of ethical value is the good of humankind, defined as pleasure and happiness, and the essence of evil is pain and sadness. To determine the right course of action, al-ʿIzz suggests tallying the consequences of good and evil and maximizing happiness while minimizing sadness.[86]

The welfarist perspective in Islamic jurisprudence may endorse utility-based evaluations of AI applications. According to this model, algorithms and autonomous machines that produce greater welfare than harm can be considered ethical from an Islamic standpoint. Principles like privacy, transparency, fairness, and accountability become criteria to assess the overall utility of AI applications. For instance, a predictive algorithm law enforcement agencies uses to enhance security and reduce crime may be deemed ethical if its benefits outweigh the negative effects on the privacy, fairness, and accountability of those potentially affected.

The consequentialist approach has inherent flaws, making it a challenging foundation for the exclusive interpretation of *maṣlaḥa* from an Islamic perspective. This approach requires a clear definition and consensus on an intrinsic value to maximize. However, achieving such clarity and consensus is often difficult, as ethical values can be subjective and influenced by individual worldviews and biases.[87]

Take, for instance, an environmental activist: they could calculate the environmental harms of AI applications, citing the significant carbon footprint of major applications like Natural Language Processing (NLP), which reportedly emits over 600,000 pounds of carbon dioxide – equivalent to five times the lifetime emissions of an average American car.[88] They may also highlight the negative impact of energy consumption in large data centers, which use roughly 200 terawatt hours (TWh) of energy annually, exceeding the national energy usage of some countries.[89] Conversely, one could argue for AI's positive environmental impact. AI could significantly aid sustainability efforts by optimizing energy consumption, enhancing waste management, enabling precise deforestation monitoring, and conserving

[84] al-Rāzī (1988)

[85] Shihadeh (2006)

[86] ʿAbd Al-Salām, *al-Qawāʿid al-kubrā,* 9–15

[87] Vallor (2018)

[88] Hao (2019)

[89] Jones (2018)

resources. Additionally, AI's predictive capabilities could help scientists anticipate climate changes, thus informing more effective mitigation strategies.[90]

Therefore, the consequentialist evaluations of benefits and harms, seen as a potential interpretation of *maṣlaḥa*, should be rooted in the notion of general public welfare, referred to as *maṣlaḥa kullyia* in Islamic jurisprudence.[91] A mere economic argument positing that a specific AI application will primarily benefit developers, users, or a distinct societal sector is insufficient. Rather, a comprehensive appraisal is required, encompassing a wide-ranging analysis of the overall societal and economic impacts. This assessment aims to confirm that the AI application in question contributes to the community's overall well-being, prioritising society's holistic welfare over the narrow interests of specific groups or sectors.

3.4 Rule-based Approaches to the Islamic Ethics of AI

Some argue that defining *maṣlaḥa* solely as public interest or welfare in Islamic jurisprudence oversimplifies its technical complexity. This narrow description might miss other dimensions of ethical value that do not align with utility-based notions when determining what is good and right. Instead, it might be more accurate to see *maṣlaḥa* in broader normative terms, as a state of affairs reflecting ethical ideals and values in harmony with divine will.[92] While *maṣlaḥa* can indeed encompass material welfare and utility elements, it should not be confined to these alone. It's a multifaceted concept that includes various intrinsic ethical values, such as justice, compassion, and human dignity.

Scholars like Muḥammad Saʿīd al-Būti support the counterview to the welfarist/utility-based visions of Islamic ethics. Al-Būti (d.2013) argued against reducing Islamic moral reasoning to a utilitarian goal of maximizing the greatest good for the greatest number.[93] He criticized attempts by reformers to introduce rationality into determining value and making moral choices, fearing that increased rationalism in moral evaluations would lead to "whimsical" normative positions (*hawā*) that breaches well-established Islamic norms.[94] Al-Būti believed that the concept of good could not be solely based on rational calculations of human needs, desires, pain, or pleasure. While he accepted that God desires the good of humankind, he refused to explain this desire in standard utilitarian terms, instead emphasizing the role of revelation and metaphysical signals in determining moral goodness. In other words, the right precedes the good, with revelation determining the right thing to do, and what revelation determines as right being intrinsically good regardless of human perceptions of pain and pleasure. Therefore, a moral agent may be required

[90] Nishant et al. (2020)

[91] Elmahjub (2019)

[92] Elmahjub (2021)

[93] al-Būti (1965)

[94] ibid 140

to endure various forms of pain, including losing one's life, to advance the cause of religion.[95]

In classical Islamic jurisprudence, the concept of the good (*maṣlaḥa*) is not only limited to the welfarist or utility-based interpretations. According to prominent figures such as al-Juwaynī and al-Ghazālī, *maṣlaḥa* in norm creation does not solely aim to maximize the greatest good for the greatest number. Al-Juwaynī considers revelatory norms to be the source of ethical value and argues that moral knowledge does not rely on intuitionism, which may allow for extra-scriptural reasoning. *Maṣlaḥa* is defined exclusively as what is intended by revelation (*maqṣūd al-sharʿ*), such that only scripture can guide us to identify what is good (*ḥasan*) and should be promoted, or what is evil (*qabīḥ*) and should be avoided.[96]

Al-Ghazālī shares al-Juwaynī's belief in the importance of textual sources in determining good and evil. However, he does not see any contradiction between *maṣlaḥa* and human good as intended by revelation. Al-Ghazālī argues that the textual sources promote values that are beneficial to humankind, and are governed by design principles that enable moral agents to bring about good and avoid evil.[97] He suggests that human reasoning about good and evil should be guided by an inductive and deductive reading of textual sources, as human intellect cannot discern values outside of the textual environment. Al-Ghazālī identifies five ethical values that must guide all normative analyses in the Islamic worldview, including promoting religion, human life, dignity, intellect, and wealth.[98]

We could see Al-Ghazālī rejection of utility-based calculations in his position regarding one of the textual objectives that he put forward. Al-Ghazālī advocates for respecting individual human life regardless of the consequences. He presents a dilemma situation where a decision must be made between sacrificing one life or engaging in utilitarian assessment to save more lives. Al-Ghazālī believes that the right thing to do is to refrain from sacrificing that individual life as a first-order principle, without engaging in consequentialist cost and benefit calculations. He provides an example similar to the trolley problem, arguing that it is impermissible to sacrifice one person to save others because each life is sacred and cannot be sacrificed for the greater good.[99]

Reflecting on the viewpoint above, a direction in Islamic jurisprudence highlights the importance of ethical values and obligations beyond conventional welfarist or utilitarian frameworks. In AI ethics, this perspective supports a duty-centric model, emphasizing unwavering commitment to principles like fairness, privacy, transparency, and accountability. By valuing these ethical principles, it guides AI designers and users to prioritize safeguarding individual rights and intrinsic values over calculations based on welfare or utility. This perspective underscores the importance of upholding moral responsibilities and respecting human dignity, even if it does not always maximize collective welfare.

[95] ibid 25

[96] Al-Juwaynī (1980)

[97] al-Ghazālī (1971)

[98] *al-Mustaṣfá*, 481–482

[99] *al-Mustaṣfá*, 489

4 Toward a Hybrid Vision of *Maṣlaḥa*

This paper's main argument is that it is possible to derive two distinct interpretations of *maṣlaḥa* regarding AI. On the one hand, there is a welfarist or utility-oriented approach, which focuses on maximizing the overall benefits to society. On the other hand, a duty-based approach prioritizes respect for intrinsic values, such as fairness, dignity, and human agency, over standard welfarist considerations. These conflicting interpretations of the concept of the "good" or "*maṣlaḥa*" in Islamic ethical discourse on AI demand further investigation and reconciliation.

Future research should aim to reconcile these two conflicting interpretations of the "good" in Islamic ethical discourse on AI. It is crucial to recognize that, when we consider Islamic jurisprudence as an ethical enterprise, there is no compelling religious or theoretical reason to view it as an absolute system of ethics. Instead, we are not obliged to accept a philosophical orthodoxy that would lead us to choose between either duty-based or welfarist normative positions to justify Islamic ethical values and guide human action in AI domains.

A more nuanced understanding of Islamic ethics perceives it as a hybrid system, fusing duty-based and welfarist moral positions relevant to each unique AI ethical context. In this framework, duty-based arguments for AI become primary principles, complemented by utility or welfarist arguments as secondary principles. This approach to Islamic ethics provides a more comprehensive, flexible ethical structure capable of adapting to the rapidly changing landscape of AI technologies and their societal ramifications.

Maqāṣid Al-Sharīʿa can offer valuable ethical references for this proposed hybrid vision of *maṣlaḥa*. Al-Ghazālī, and numerous classical and modern scholars, identified five main objectives of Sharīʿa: the preservation of religion, human life, lineage, intellect, and wealth.[100] Pursuing each of these objectives constitutes *maṣlaḥa*—an ethical state of affairs. Notably, the ethical significance within these objectives exists on a spectrum, with *darurāt* (essentials) deemed most crucial to preserve and promote, followed by secondary needs (*ḥājiyyāt*) and enhancements (*taḥsīnīyyāt*).[101] If we approach *maṣlaḥa* as a normative construct encompassing welfare-based and duty-based orientations, we should avoid binary choices between promoting welfare or respecting a certain duty. Instead, we can identify a set of intrinsic values—like religion or human life—at the level of essentials or *darurāt* to serve as primary principles. In these scenarios, the morally required choice would be one that promotes these essential values, irrespective of the consequences.

Welfare or utility-based considerations may be more fitting in the realms of *ḥājiyyāt* and *taḥsīnīyyāt*. Such ethical endeavors will not be straightforward—they will necessitate in-depth analysis of current and future AI applications to identify their specific risk domains. We must then ensure that values, laws, and policies align with the essential *maqāṣid*, whilst leaving room to optimize welfare and positive outcomes, as long as it does not undermine any essential intrinsic value.

[100] al-Ghazālī (n.d.)

[101] al-Raysuni and al-Shāṭibī's (2005)

5 Concluding Remarks

The advent of AI has instigated considerable shifts in various dimensions of human existence, prompting ethical dilemmas linked to vital societal values, including autonomy, privacy, fairness, and transparency. Notably, Western or Eurocentric moral concepts have largely influenced the ethical benchmarking and policy discourse surrounding AI. This paper, however, champions a more pluralistic interpretation of AI ethics, presenting an Islamic viewpoint, a critical component for fostering a more comprehensive and globally applicable understanding of AI ethics.

Islamic systems of ethical value are complex and multifaceted. It is not a simple form of divine command theory, whereby morality is based solely on divine mandates. However, like non-religious ethical discourse, Islamic ethics is characterized by multiple layers of highly abstract and often conflicting meta-ethical and normative propositions. Although Islamic morality is rooted in divine dictates, ethical uncertainties emerge when extrapolating value criteria and normative rules from God's commands. A noteworthy similarity between Islamic and Western philosophical ethics lies in presenting two formulas for ethical value and moral action. One argument posits that ethical AI maximizes desirable consequences for the majority. At the same time, the other upholds a duty to respect a set of intrinsic values, which include religion in the Islamic version, human life, and dignity. However, these systems diverge in their sources and ultimate objectives of ethical assessment. Islamic ethics draw from the textual sources of Islamic revelation and consider the metaphysical dimension of desired values. In contrast, Western normative ethics primarily engage in a rational assessment of behavior to determine moral and immoral actions, aiming to promote worldly interests or values for moral agents.

The notion of *maṣlaḥa* rooted in *maqāṣid Al-Sharīʿa* is perhaps the most relevant Islamic source of ethics when it comes to the ethical uncertainties of the emerging AI technologies and applications. However, this notion does not lend itself to simple explanation of the ethical values that need to be pursued. Should our understanding of *maṣlaḥa* aim at maximizing welfare in designing and deploying AI applications or should it be confined to safeguarding a set of imperatives? This paper argues that an optimal course of action may lead us to consider both options.

The concept of *maṣlaḥa*, grounded in *maqāṣid Al-Sharīʿa*, arguably emerges as the most pertinent Islamic ethical framework in addressing the ethical uncertainties that new AI technologies and applications present. However, applying this concept does not directly result in a clear-cut set of ethical values to be pursued. Questions arise as to whether our interpretation of *maṣlaḥa* should be oriented towards maximizing welfare in the design and deployment of AI applications, or should it primarily focus on safeguarding a specific set of imperatives. This paper suggests that the most effective approach is likely a more nuanced one, considering both welfare maximization and safeguarding of specific imperatives as interrelated and complementary aspects of a cohesive ethical framework. By recognizing the multifaceted nature of *maṣlaḥa*, we can better appreciate the intricacies of Islamic ethics, enabling us better to align AI technology with universally relevant ethical considerations.

Funding Open Access funding provided by the Qatar National Library. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The author has no relevant financial or non-financial interests to disclose.

References

Abdallah, S. (2008). Information ethics from an Islamic perspective. In M. Quigley (Ed.), *Encyclopedia of Information Ethics and Security*. Hershey: Information Science Reference

Abdallah, S. (2010). Islamic ethics: An exposition for resolving ICT ethical dilemmas. *Journal of Information, Communication and Ethics in Society, 8*(3), 289–301.

Abū Zahra M (n.d.) *Tanzīm al-Islām lil-Mujtama*ʿ (Dār al-Fikr al-ʿArabi)

Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews, 46*(1–2), 176–197.

al-Būti, M. S. (1965). *Ḍawābit al-maslạ hạ fi al-Sharīʿa al-islāmīyya*. In PhD Thesis, Faculty of Sharīʿa al-Azhar University

ʿal-Fāsī, A. (1963). *Maqāṣid al-Sharīʿa al-Islamiyya wa-Manakibuha* (Maktabat al-Wahda al-ʿArabiyya), 3–7,41

al-Ghazālī, A. (1971). *Shifāʾ al-ghalīl* (Ḥamd ʿUbayd al-Kubaysī ed, Matḅ aʿat al-Irshād 1971) 211

al-Ghazālī, A. H. (n.d.). *al-Mustaṣfá* (Ḥamza b. Zuhayr Ḥāfiẓ ed, Sharikat al-Madīna al-Munawwara lil-Ṭibāʿa, n.d.) *2*, 481–82

Al-Juwaynī (1980) *al-Burhān fī usụ l al-fiqh* (ʿAbd al-ʿAzị m al-Dīb ed, Dār al-Ansạr 1980) 48 and 91

al-Munajjid S and al-Khūrī Y (1970) *Fatwā al-Imām Muḥammad Rashīd Riḍā* (Dār al- Kitāb al-jadīd 1970)

al-Qaraḍāwī, Y. (1996). *fiqh al-'awlawiyyāt, dirāsah jadida fī daw' al-Qurʾān wa al-Sunnah* (Maktabat Wahba 1996) 31

al-Qaraḍāwī, Y. (2011). *al-Siyyāsa al-Sharʿiyya* (Maktabat Wahba 2011) 32

al-Raysuni, A,. & al-Shāṭịbī's, I. (2005). Theory of the higher objectives and intents of Islamic law (International Institute of Islamic Thought 2005)

al-Rāzī, M. F. (1988). *al-Maḥṣūl fī ʿilm uṣūl al-fiqh* (Dār al-Kutub al-ʿIlmiyya 1988) *5*, 158–162

Angwin, J., Larson, J., Mattu, S., et al. (2016). *Machine bias risk assessments in criminal sentencing*. ProPublica.

Belk, R. (2021). Ethical issues in service robotics and artificial intelligence. *The Service Industries Journal, 41*(13–14), 860–876.

Bench-Capon, T. J. (2020). Ethical approaches and autonomous systems. *Artificial Intelligence, 281*, 103239.

Benjamins, V. R. & García I. S. (2020). *Towards a framework for understanding societal and ethical implications of Artificial Intelligence*. Vulnerabilidad y cultura digital by Dykinson

Berberich, N., & Diepold, K. (2018). The virtuous machine-Old ethics for new technology. https://arxiv.org/pdf/1806.10322.pdf

Birhane, A. (2020). Algorithmic colonization of Africa. *Scripted, 17*(389), 314.

Buolamwini, J., & Gebru, T. (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, (77–91)

Chaudhary, M. Y. (2020). Initial considerations for islamic digital ethics. *Philosophy & Technology, 33*(4), 639–657.

Chen, L., Ma, R., Hannák, A., & Wilson, C. (2018). Investigating the impact of gender on rank in resume search engines. In Proceedings of CHI

Christian, B. (2020). *The alignment problem: Machine learning and human values*. WW Norton & Company.

Cohen, J. (1993). Moral Pluralism and Political Consensus. In Copp, D., Hampton, J., & Roemer, J. E. (Eds.), *The Idea of Democracy* (CUP 1993) (274–275)

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women

Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C∞., Quezada, G., Scowcroft, J., & Hajkowicz, S†. (2019). Artificial intelligence: Australia's ethics framework. *Data61 CSIRO, Australia; AI HLEG - High-Level Expert Group on Artificial Intelligence*

De Almeida, P. G. R., dos Santos, C. D., & Farias, J. S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology, 23*(3), 505–525.

Egypt's National Artificial Intelligence. (2021). https://mcit.gov.eg/en/Publication/Publication_Summary/9283

Elmahjub, E. (2019). Transformative vision of Islamic jurisprudence and the pursuit of common ground for the social good in pluralist societies. *Asian Journal of Comparative Law, 14*(2), 305–335. 329.

Elmahjub, E. (2021). Islamic Jurisprudence as an ethical discourse: An enquiry into the nature of moral reasoning in islamic legal theory. *Oxford Journal of Law and Religion, 10*(1), 16–42.

Elmahjub, E., & Qadir, J. (2023). How to program autonomous vehicle (AV) crash algorithms: an Islamic ethical perspective. *Journal of Information, Communication and Ethics in Society*

Friedman, B., & Nissenbaum, H. (1996). Bias in Computer Systems. *ACM Transactions on Information Systems, 14*(3), 330–347.

Greene, D., Hoffmann, A. L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*

Gupta, A. et al. (2021). *The state of AI ethics*. Montreal: Montreal AI Ethics Institute (MAIEI)

Habermas, J. (2006). 'Religion in the Public Sphere' 14 EJP,10

Hao, K. (2019). Training a single AI model can emit as much carbon as five cars in their lifetimes. *MIT Technology Review, 75*, 103.

Hallaq WB (2009) *An introduction to Islamic law* (CUP 2009) 116

Hendrickson, J. (2013). "Fatwa". In Gerhard Böwering, Patricia Crone (ed.), *The Princeton Encyclopedia of Islamic Political Thought*. Princeton University Press

Hourani, G. (1960). 'Two theories of value in medieval Islam' 50 Muslim World

IEEE Standards Association. (2019). The IEEE global initiative for ethical considerations in artificial intelligence and autonomous systems

ʿIzz al-Dīn ibn ʿAbd al-Salā m. (1991). Qawāʿid al-ahḳām fī masạliḥ al-anām [The Rules of Lawmaking in the Pursuit of the Common Good] (Ṭāhā ʿAbd al-Raʾūf Saʿd ed, Maktabat al-Kulliyyā t al-Azhariyya 1991)

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.

Jones, N. (2018). How to stop data centres from gobbling up the world's electricity. *Nature, 561*(7722), 163–166.

Kamali, M. H. (2003). Principles of Islamic Jurisprudence

Kerr, M. (1966). *Islamic Reform: The Political and Legal Theories of Muhammad ʿAbduh and Rashīd Riḍā*. University of California Press

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788.

Langford, C. (2017). Houston schools must face teacher evaluation lawsuit. *Courthouse News Service*

Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G. et al. (2022) Gathering strength, gathering storms: The one hundred year study on artificial intelligence (AI100) 2021 study panel report. arXiv preprint arXiv:2210.15767

March Andrew, F. (2009). Sources of Moral Obligation to non-Muslims in the "Jurisprudence of Muslim Minorities"(Fiqh al-aqallîyât) Discourse. *Islamic Law and Society, 16*, 34–94

Mayson, S. G. (2019). Bias in, bias out. *The Yale Law Journal, 128*(8), 2218–2300.

Mhlambi, Sabelo. (2020). From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance. Carr Centre Discussion Paper. Available at: https://carrcenter.hks.harvard.edu/files/cchr/files/ccdp_2020-009_sabelo_b.pdf

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 2020*, 1–26.

Moore, G. E. (1988). Principia Ethica (Amherst 1988)

Muḥammad ʿAbduh (1972) al-Amāl al-kāmila li-l-imām Muḥammad Abduh (Muḥammad ʿAmāra ed, al-Muʾassasa al-ʿarabiyya li al-dirāsāt wa al-nashr 1972) *3*, 257–350

Muḥammad ʿAbduh (1988) al-Islām wa-l-naṣrāniyya maʿa l-ʿilm wa-l-madaniyya (Dār al-Ḥadātha 1988) 74–76

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management, 53*, 102104.

Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.

Nusseibeh, S. (2017). The story of reason in Islam. Stanford University Press

Opwis, F. (2010). *Maṣlaḥa and the Purpose of the Law*. Brill

Qatar's Ministry of Transportation and Communication (2019) Qatar National AI Strategy

Quinton, A. M. (1973). Utilitarian ethics (Palgrave Macmillan 1973)

Rahṃān, F. (1965). Islamic methodology in history (Central Institute of Islamic Research 1965)

Raquib, A. (2015). *Islamic ethics of technology: An objectives' (Maqasid) approach*. The Other Press.

Raquib, A. (2016). Maqasid Al-Sharī'Ah: A traditional source for ensuring design and development o modern technology for humanity's benefit. In *Islamic Perspectives on Science and Technology*, edited by Kamali, M. H., Bakar. O., Batchelor, D. A. F., & Hashim, R., 143–67. Singapore: Springer

Raquib, A., Channa, B., Zubair, T., & Qadir, J. (2022). Islamic virtue-based ethics for artificial intelligence. *Discover Artificial Intelligence, 2*(1), 11.

Rashdall H (1907) The theory of good and evil: A treatise on moral philosophy (Clarendon 1907)

Riḍā, M. R. (n.d.). '*Adilat al-Sharʿwa taqdlīm al-maṣlaḥa ʿala al-naṣṣ', Bāb Usụ̄ l al-fiqh*' Riḍā (n 80) *9*, 746–770

Rosenberg, M., Confessore, N., Cadwalladr, C. (2018). *How Trump consultants exploited the Facebook data of millions*. The New York Times

Sambasivan, N. et al. (2021). Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (316)

Sardar, Z. (1988). Information and the Muslim world: A strategy for the twenty-first century. In London. New York: Mansell

Shihadeh, A. (2006). The Teleological Ethics of Fakhr al-Dīn al-Rāzī. Brill

Slote, M. (2001). *Morals from Motives*. Oxford: Oxford University Press

Stanford (2021) One hundred year study on artificial intelligence

Sweeney, L. (2013). Discrimination in online ad delivery. *Queue, 11*(3), 10–29.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good: An ethical framework will help to harness the potential of AI while keeping humans in control. *Science Review, 361*(6404), 751–752.

UAE Strategy for Artificial Intelligence. (2017). https://ai.gov.ae/wp-content/uploads/2021/07/UAE-National-Strategy-for-Artificial-Intelligence-2031.pdf

Ulgen, O. (2017). Kantian ethics in the age of artificial intelligence and robotics. *QIL, 43*, 59–83.

Vallor, S. (2018). *Technology and the virtues a philosophical guide to a future worth wanting*. Oxford University Press.

Wagner, B. (2018). Ethics as an escape from regulation. From "ethics-washing" to ethics-shopping?

Wahba al-Zuhayli. (1986). Usụ̄l al-fiqh al-islāmī [The Principles of Islamic Jurisprudence] (Dār al-Fikr 1986)

World Economic Forum. (2018). How to prevent discriminatory outcomes in machine learning. *Global Future Council on Human Rights 2016–2018*

Wong, H. (2016). Responsible innovation for decent nonliberal peoples: A dilemma? *Journal of Responsible Innovation, 3*(2), 154–168.

Zagzebski, L. (2004). *Divine Motivation Theory*. Cambridge University Press